東海大學統計研究所

碩士論文

指導教授:沈葆聖博士

雙變數重複發生事件分配的逆加權估計

The Inverse Probability Weighted Estimator of the Bivariate Recurrence Time Distribution

研究生:劉昱宏

中華民國一百零四年七月

謝誌

　硕班的學習，最感謝的是我的指導教授一沈葆聖博士，在學業上不辭辛苦的教導我，在論文上盡心盡力指導我的論文，除了學業方面，在待人處事上也給了我不少指點，當在研究室討論時，教授展現的是一種對人生的態度，對生活的態度，得知我考認證失利時，也是第一時間找我，鼓勵我人生就是如此偶而失常，下次才會更細心。感謝口試委員戴政博士和張玉媚博士百忙之中抽空來擔當我論文的口委，給了需多指點以及改進的地方，可以讓我更加成長。感謝系上的所有的老師以及助教因為在東海統計系讀了大學四年和碩士班兩年，從朦朧無知的 18 歲到即將踏出社會的 26 歲，經過系上老師，助教諄諄教誨，早已有深厚的情誼，再來感謝，在求學旅途中的朋友在大學期間或是在碩士期間的幫忙，拉出陷入低潮以及對人生規劃有迷惘時能夠一起討論，分享。

　我感謝我的家人，從小的栽培我，在長達 22 年求學的旅途中，無怨悔的支持我，也會尊重我的想法，灌輸我，家是家中每個成員一同經營的，所以家人要同心協力。

　最後希望我所認識的人，都能在未來的旅途中，都能夠平安喜樂。

昱宏 07/07/2015

Abstract

Recurrent event data frequently arise in longitudinal studies. In many applications, subjects may experience two different types of events alternatively over time or a pair of subjects may experience recurrent events of the same type. In this article, using the inverse-probability weighted (IPW) approach, we propose nonparametric estimators for the joint distribution functions of bivariate recurrence times. The asymptotic properties of the IPW are established under independent censoring. A simulation study is conducted to investigate the performance of the proposed estimators.

Key Words: Inverse-probability-weighted; recurrent events; informative censoring.

Contents

Chapter 1

1. Introduction

Recurrent event data are frequently encountered in longitudinal studies. For the analysis of recurrent event data, two different time scales are employed in the literature: the times since entering the study and the times since last event. For the first type of time scale, many statistical methods have been developed, such as Prentice et al. (1981), Andersen and Gill (1982), Pepe and Cai (1993), Huang and Louis (1998), Wang and Wells (1998), Lin et al. (1999, 2000), and Wang et al. (2001). The methodology considered by these authors are based on formulation of either the intensity function or the occurrence rate function of the underlying event process.

When the time between consecutive events (gap time) is used for study, the stochastic ordering structure of recurrent events poses problems for statistical analysis, such as induced censoring and sampling biases. Recently, various statistical methods have been proposed for analysis of recurrent gap time data. Under the identically independent distributed (i.i.d.) assumptions of gap times, Pena et al. (2000) showed that the generalized Kaplan-Meier (1958) estimator is the nonparametric maximum likelihood estimator (NPMLE) of the survival function of the univariate recurrence time. Using the extended risk methods, Wang and Chang (1999) proposed an estimator in the case where within-unit interoccurrence times are correlated.

In many applications, bivariate recurrent event data can arise. Consider the following applications.

Case 1

The subjects experience two different types of events alternately over time. For example, in psychiatric study, a patient with schizophrenia could be repeatedly admitted into and discharged from a hospital. In a reliability study, a mechanical system can alternate between periods of use and repair.

Case 2

A pair of study subjects can experience repeated occurrences of certain diseases. For example, in a genetic study, each twin (or a parent/child) may experience repeated occurrences of certain diseases.

The analysis of bivariate recurrent event data plays an important role in estimating the association within bivariate recurrent events and provide a basis of model building. In literature, many methods have been developed to estimate multivariate distribution when events are of different types, e.g. Visser (1996), Huang and Louis (1998), Wang and Wells (1998) and Lin et al. (1999). Although these methods can be applied to bivariate recurrent event data, they are inefficient since only the first pair of recurrence times are used.

For data from case 1, Huang and Wang (2005) proposed a class of nonparametric estimators for bivariate distribution of recurrence times by combing techniques for univariate recurrent event data and techniques for bivariate gap times data. In Section 2, using the inverse-probability weighted (IPW) approach, we propose an alternative nonparametric estimator for data from case 1. We also propose an IPW estimator for data from case 2. The asymptotic properties of the proposed estimator are established. In Section 3, a simulation study is conducted to investigate the performance of the proposed estimators.

Chapter 2

2. The Proposed Estimators

2.1. Method for Case 1

Suppose that a sequence of bivariate event is to be observed in a follow-up study. Let (T_{ij1}, T_{ij2}) denote the bivariate recurrence time from the $(j-1)^{th}$ to j^{th} event for the i^{th} pair. Let (C_i) denote the pair of censoring times, i.e. the times between the initial event to the end of follow-up. The bivariate recurrent event process $N_i = \{(T_{ij1}, T_{ij2}) : j = 1, 2, \dots\}$ is subject to right censoring. Assume that subjects are sampled independently and

$(A.1)$: There exists a latent variable Z_i such that conditional on Z_i, the bivariate random vectors $(T_{ij1}, T_{ij2}), j = 1, 2 \dots$, are i.i.d.;

$(A.2)$: The censoring time C_i is independent of (Z_i, N_i).

Notice that the recurrence times T_{ij1} and T_{ij2} are allowed to be correlated conditional on Z_i. Let $F(t_1, t_2)$ denote the joint distribution of T_{ij1} and T_{ij2}, i.e.

$$F(t_1, t_2) = \int P(T_{ij1} \leq t_1, T_{ij2} \leq t_2 | Z_i = z) dP_Z(z),$$

where $P_Z(z)$ is the probability distribution function of Z. Let m_i be the index of censored bivariate recurrence times for the i^{th} individual such that

$$\sum_{j=1}^{m_i-1} (T_{ij1} + T_{ij2}) \leq C_i \text{ and } \sum_{j=1}^{m_i} (T_{ij1} + T_{ij2}) > C_i.$$

Hence, m_i is a random variable and T_{im_i1} may or may not be completely observed while T_{im_i2} is always censored. Notice that $F(t_1, t_2)$ is identifiable only for $t_1 + t_2 \leq \tau_c$, where τ_c is the right support of C_i.

2.1.1. The approach of Huang and Wang

Now, we briefly review the method proposed by Huang and Wang (2005). Denote $X_{ij} = T_{ij1} + T_{ij2}$ and $Y_{ij} = (T_{ij1}, T_{ij2})$. Let $D_{ij} = C_i - \sum_{l=1}^{j-1} X_{il}$ and denote $\tilde{X}_{ij} = \min(X_{ij}, D_{ij})$, $\tilde{Y}_{ij} = Y_{ij}\delta_{ij}$ where $\delta_{ij} = I_{[\tilde{X}_{ij} \leq D_{ij}]}$. We further define the functions $F_a(t, y) = E[a_i I_{[\tilde{X}_{i1} \leq t, \tilde{Y}_{i1} \leq y, \delta_{i1}=1]}]$ and $R_a(t) = E[a_i I_{[\tilde{X}_{i1} \geq t]}]$, where $y = (y_1, y_2)$ is a vector of real numbers and $a_i = a(C_i)$ is a nonnegative function of C_i with $E[a_i^2] < \infty$. Under assumptions $(A.1)$ and $(A.2)$, $F_a(ds, y)/R_a(s) = F_{X,Y}(ds, y)/S_X(s-)$, where $F_{X,Y}(s, y) = P(X_{i1} \leq s, Y_{i1} \leq y)$ is the joint distribution function of X_{i1} and Y_{i1} and $S_X(s) = P(X_{i1} > s)$ is the marginal survival function of X_{i1} and it follows that

$$F_{X,Y}(t, y) = \int_0^t S_X(s-) F_a(ds, y)/R_a(s).$$

Let $m_i^* = m_i - 1$ for $m_i \geq 2$ and $m_i^* = 1$ for $m_i = 1$. Let

$$\hat{F}_a(t, y) = n^{-1} \sum_{i=1}^{n} \frac{a_i I_{[m_i^* \geq 1]}}{m_i^*} \sum_{j=1}^{m_i^*} I_{[\tilde{X}_{ij} \leq t, \tilde{Y}_{ij} \leq y]}, \quad \hat{R}_a(y) = n^{-1} \sum_{i=1}^{n} \frac{a_i}{m_i^*} \sum_{j=1}^{m_i^*} I_{[\tilde{X}_{ij} \geq t]}.$$

Thus, $F_{X,Y}(t, y)$ can be estimated by

$$\hat{F}_{X,Y}(t, y) = \int_0^t \hat{S}_X(s-) \hat{F}_a(ds, y) / \hat{R}_a(s),$$

where

$$\hat{S}_X(s) = \prod_{u \leq s} \left(1 - \frac{\hat{F}_a(du, \infty)}{\hat{R}_a(u)} \right).$$

Through the identity $F(t_1, t_2) = F_{X,Y}(t_1 + t_2, (t_1, t_2))$, it follows that $F(t_1, t_2)$ can be estimated by

$$\hat{F}_n(t_1, t_2) = \hat{F}_{X,Y}(t_1 + t_2, (t_1, t_2)) = \int_0^{t_1 + t_2} \hat{S}_X(s-) \hat{F}_a(ds, (t_1, t_2)) / \hat{R}_a(s).$$

Huang and Wang (2005) showed that $n^{1/2}(\hat{F}_n(t_1, t_2) - F(t_1, t_2))$ converges to a mean zero Gaussian process with variance-covariance function $E[\varphi_1(t_1, t_2)\varphi_1(t_1', t_2')]$, where

$$\varphi_1(t_1, t_2) = \int_0^{t_1} F_{X,Y}(s, t_2) \omega_1(ds, \infty) + \int_0^{t_1} S_X(s) \omega_1(ds, t_2) - F_{X,Y}(t_1, t_2) \omega_1(t_1, \infty),$$

where

$$\omega_i(t_1, t_2) = \frac{a_i I_{[m_i^* \geq 1]}}{m_i^*} \sum_{j=1}^{m_i^*} \frac{I_{[\tilde{X}_{ij} \leq t_1, \tilde{Y}_{ij} \leq t_2]}}{R_a(\tilde{X}_{ij})} - \int_0^{t_1} \frac{a_i}{m_i^*} \sum_{j=1}^{m_i^*} \frac{I_{[\tilde{X}_{ij} \geq s]}}{R_a(s)^2} F_a(ds, t_2).$$

The variance-covariance function $E[\varphi_1(t_1, t_2)\varphi_1(t_1', t_2')]$ can be consistently estimated by

$$n^{-1} \sum_{i=1}^{n} \hat{\varphi}_i(t_1, t_2) \hat{\varphi}_i(t_1', t_2'),$$

where

$$\hat{\varphi}_i(t_1, t_2) = \int_0^{t_1} \hat{F}_{X,Y}(s, t_2) \hat{\omega}_i(ds, \infty) + \int_0^{t_1} \hat{S}_X(s) \hat{\omega}_i(ds, t_2) - \hat{F}_{X,Y}(t_1, t_2) \hat{\omega}_i(t_1, \infty),$$

and

$$\hat{\omega}_i(t_1, t_2) = \frac{a_i I_{[m_i^* \geq 1]}}{m_i^*} \sum_{j=1}^{m_i^*} \frac{I_{[\tilde{X}_{ij} \leq t_1, \tilde{Y}_{ij} \leq t_2]}}{\hat{R}_a(\tilde{X}_{ij})} - \int_0^{t_1} \frac{a_i}{m_i^*} \sum_{j=1}^{m_i^*} \frac{I_{[\tilde{X}_{ij} \geq s]}}{\hat{R}_a(s)^2} \hat{F}_a(ds, t_2).$$

2.1.2. The IPW approach

Next, using the IPW approach, we propose an alternative estimator. Consider the conditional distribution function

$$\tilde{F}(t_1, t_2) = P(T_{i11} \le t_1, T_{i12} \le t_2 | m_i^* \ge 1)$$
$$= P(T_{i11} \le t_1, T_{i12} \le t_2, T_{i11} + T_{i12} \le C_i)$$
$$= \int_0^{t_2} \int_0^{t_1} S_C((x+y)-) F(dx, dy),$$

where $S_C(t) = P(C_i > t)$ is the survival function of C_i. Note that $S_C(t)$ can be estimated by $\hat{S}_C(t) = n^{-1} \sum_{i=1}^n I_{[C_i > t]}$. Furthermore, under assumptions $(A1)$ and $(A2)$, conditional on $m_i^* \ge 1$ and Z_i, $(T_{i11}, T_{i12}), (T_{i21}, T_{i22}), \ldots, (T_{im_i^*1}, T_{im_i^*2})$ are identically distributed. Let $n^* = \sum_{i=1}^n I_{[m_i^* \ge 1]}$. Hence, an unbiased estimator of $\tilde{F}(t_1, t_2)$ is given by

$$\tilde{F}_n(t_1, t_2) = \frac{1}{n^*} \sum_{i=1}^n \frac{I_{[m_i^* \ge 1]}}{m_i^*} \sum_{j=1}^{m_i^*} I_{[T_{ij1} \le t_1, T_{ij2} \le t_2]},$$

since

$$E[\tilde{F}_n(t_1, t_2)] = E\left[\frac{1}{n^*} \sum_{i=1}^n \frac{I_{[m_i^* \ge 1]}}{m_i^*} \sum_{j=1}^{m_i^*} I_{[T_{ij1} \le t_1, T_{ij2} \le t_2]}\right]$$
$$= E_1\left[\frac{1}{n^*} \sum_{i=1}^n I_{[m_i^* \ge 1]} E_{2|1}\left[\frac{1}{m_i^*} \sum_{j=1}^{m_i^*} I_{[T_{ij1} \le t_1, T_{ij2} \le t_2]} \middle| m_i^* \ge 1\right]\right]$$
$$= E_1\left[\frac{1}{n^*} \sum_{i=1}^n I_{[m_i^* \ge 1]} \int_0^{t_2} \int_0^{t_1} S_C((x+y)-) F(dx, dy)\right],$$
$$= \int_0^{t_2} \int_0^{t_1} S_C((x+y)-) F(dx, dy) = \tilde{F}(t_1, t_2),$$

where $E_{2|1}$ and E_1 denote the conditional expectation given m_i^* and expectation of m_i^*, respectively. Thus, given $\hat{S}_C$, $F(t_1, t_2)$ can be estimated by

$$\hat{F}_W(t_1, t_2) = \int_0^{t_2} \int_0^{t_1} \frac{1}{\hat{S}_C((x+y)-)} \tilde{F}_n(dx, dy) = \frac{1}{n^*} \sum_{i=1}^n \frac{I_{[m_i^* \ge 1]}}{m_i^*} \sum_{j=1}^{m_i^*} \frac{I_{[T_{ij1} \le t_1, T_{ij2} \le t_2]}}{\hat{S}_C((T_{ij1} + T_{ij2})-)}.$$

The asymptotic results of $\hat{F}_W$ are given in Theorem 1.

Theorem 1

Under assumptions $(A.1)$, $(A.2)$ and assuming that n^*/n converges in probability to p, then $\sqrt{n}(\hat{F}_W(t_1,t_2) - F(t_1,t_2))$ has an asymptotically i.i.d. representation

$$\sqrt{n}(\hat{F}_W(t_1,t_2) - F(t_1,t_2)) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi_i(t_1,t_2) + o_p(1),$$

which converges weakly to a mean zero Gaussian process with the variance-covariance function $E[\psi_i(t_1,t_2)\psi_i(t_1^{'},t_2^{'})]$, where $\psi_i(t_1,t_2) = \eta_i(t_1,t_2) + \zeta_i(t_1,t_2)$,

$$\eta_i(t_1,t_2) = p^{-1}\frac{I_{[m_i^*\geq 1]}}{m_i^*}\sum_{j=1}^{m_i^*}\frac{I_{[T_{ij1}\leq t_1, T_{ij2}\leq t_2]}}{S_C((T_{ij1}+T_{ij2})-)} - F(t_1,t_2),$$

and

$$\zeta_i(t_1,t_2) = -\int_0^{t_2}\int_0^{t_1}\frac{1}{S_C^2((x+y)-)}[I_{[C_i\geq(x+y)]} - S_C((x+y)-)]\tilde{F}(dx,dy).$$

Proof: The proof is technical and omitted here.

2.2 Method for Case 2

For data from case 2, there exist two censoring times, denoted by C_{i1} and C_{i2}. Assumption $(A.2)$ is modified to

$(B.2)$: The censoring times C_{i1} and C_{i2} are independent of (Z_i, N_i).

For $k = 1,2$, let m_{ik} be the index of censored bivariate recurrence times for the i^{th} individual such that

$$\sum_{j=1}^{m_{ik}-1}T_{ijk} \leq C_{ik} \text{ and } \sum_{j=1}^{m_{ik}}T_{ijk} > C_{ik}.$$

Clearly, m_{ik} is a random variable and the last recurrence event time for the k^{th} event of the i^{th} pair is subject to right censoring. Notice that m_{i1} may not be equal to m_{i2} and either $T_{i1m_{i1}}$ or T_{i2m_i} is subject to right censoring.

For data from case 2, consider the function

$$\begin{aligned}\tilde{H}(t_1,t_2) &= P(T_{i11}\le t_1, T_{i12}\le t_2 | m_{i1}\ge 2, m_{i2}\ge 2)\\ &= P(T_{i11}\le t_1, T_{i12}\le t_2, T_{i11}\le C_{i1}, T_{i12}\le C_{i2})\\ &= \int_0^{t_2}\int_0^{t_1} S_Q(x-,y-)F(dx,dy),\end{aligned}$$

where $S_Q(x,y) = P(C_{i1} > x, C_{i2} > y)$ denotes the joint survival function of C_{i1} and C_{i2}. Note that $S_Q(x,y)$ can be consistently estimated by $\hat{S}_Q(x,y) = n^{-1}\sum_{i=1}^n I_{[C_{i1}>x, C_{i2}>y]}$. Let $K_i = \min(m_{i1}-1, m_{i2}-1)$. Furthermore, under assumptions $(A1)$ and $(B2)$, conditional on (K_i, Z_i, C_i), $(T_{i11}, T_{i12}), (T_{i21}, T_{i22}), \ldots, (T_{iK_i1}, T_{iK_i2})$ are identically distributed. Let $n_d = \sum_{i=1}^n I_{[K_i\ge 1]}$. Thus, an unbiased estimator of $\tilde{H}(t_1,t_2)$ is given by

$$\tilde{H}_n(t_1,t_2) = n_d^{-1}\sum_{i=1}^n \frac{I_{[K_i\ge 1]}}{K_i}\sum_{j=1}^{K_i} I_{[T_{ij1}\le t_1, T_{ij2}\le t_2]}.$$

Hence, given $\hat{S}_Q$, $F(t_1,t_2)$ can be estimated by

$$\hat{F}_W(t_1,t_2) = \int_0^{t_2}\int_0^{t_1}\frac{1}{\hat{S}_Q((x-,y-)}\tilde{H}_n(dx,dy) = n_d^{-1}\sum_{i=1}^n\frac{I_{[K_i\ge 1]}}{K_i}\sum_{j=1}^{K_i}\frac{I_{[T_{ij1}\le t_1, T_{ij2}\le t_2]}}{\hat{S}_Q(T_{ij1}-, T_{ij2}-)}.$$

The asymptotic results of $\hat{F}_W$ are given in Theorem 2.

Theorem 2

Under assumptions $(A.1)$, $(B.2)$ and n_d/n converges in probability to p_d, then $\sqrt{n}(\hat{F}_W(t_1,t_2) - F(t_1,t_2))$ has an asymptotically i.i.d. representation

$$\sqrt{n}(\hat{F}_W(t_1,t_2) - F(t_1,t_2)) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \xi_i(t_1,t_2) + o_p(1),$$

which converges weakly to a mean zero Gaussian process with the variance-covariance function $E[\xi_i(t_1,t_2)\xi_i(t_1',t_2')]$, where $\xi_i(t_1,t_2) = \varsigma_i(t_1,t_2) + \upsilon_i(t_1,t_2)$,

$$\varsigma_i(t_1,t_2) = \frac{I_{[K_i\ge 1]}}{K_i}\sum_{j=1}^{K_i}\frac{I_{[T_{ij1}\le t_1, T_{ij2}\le t_2]}}{S_Q(T_{ij1}-, T_{ij2}-)} - F(t_1,t_2)$$

and

$$\upsilon_i(t_1,t_2) = -\int_0^{t_2}\int_0^{t_1}\frac{1}{S_Q^2(x-,y-)}[I_{[C_{i1}\ge x, C_{i2}\ge y)]} - S_Q(x-,y-)]\tilde{F}(dx,dy).$$

Proof: The proof is similar to that of Theorem 1 and is omitted.

Chapter 3

3. Simulation study

3.1 Data of Case 1

To evaluate the performance of the propose estimator $\hat{F}_W$, we conduct numerical simulation studies. We consider the setup used by Huang and Wang (2005). The latent variable Z_i is generated from an exponential distribution with mean 1. Given $Z_i = z$, the i.i.d. bivariate recurrence times are generated from Clayton's bivariate failure time distribution (Clayton (1978)) with joint survival function

$$S(t_1, t_2|z) = (S_1(t_1|z)^{1-\theta} + S_2(t_2|z)^{1-\theta} - 1)^{\frac{1}{1-\theta}}, \quad (3.1)$$

where $\theta \geq 1$, $S_1(t_1|z) = P(T_{ij1} > t_1) = \exp(-e^z t_1^2)$ and $S_2(t_2|z) = P(T_{ij2} > t_2) = \exp(-e^{-z} t_2^{1.5})$. The values of θ are chosen as 3 and 9 so that the corresponding Kendall's tau are 0.5 and 0.8, respectively. The distribution of C_i is uniform$(0, \tau_c)$, with $\tau_c = 8, 15$ such that the proportion of subjects having at least one pair of recurrence times is approximately 68% for $\tau_c = 8$ and 81% for $\tau_c = 15$. The sample size is chosen as $n = 200$ and the replication is 1000 times. Table 1 shows the biases, standard deviations (std) and root mean squared errors (rmse) of the estimator $\hat{F}_W$ at grid points based on the combination of t_1 and t_2 with $t_1 = 0.5, 1, 2$ and $t_2 = 1, 2, 4$. For purpose of comparison, we also report the results of Huang and Wang's estimator with wight function $a_i = C_i$, which perform best according to Tables 1 to 4 of Huang and Wang (2005).

3.2 Data of Case 2

The distribution of T_{ij1} and T_{ij2} are the same as that used in case 1. The C_{1i} and C_{2i} are independently generated from uniform$(0, \tau_c)$, with $\tau_c = 8, 15$. The proportion of subjects having at least one pair of recurrence times is approximately 68% for $\tau_c = 8$ and 80% for $\tau_c = 15$. The sample size is chosen as $n = 100, 200$ and the replication is 1000 times. Table 3 reports the simulation results.

Based on Tables 1 and 2, we have conclusions as follows.

(1) For case 1, Table 1 indicates that the biases of both estimators are very small. For most of cases considered, the standard deviations of the IPW estimator $\hat{F}_W$ are very close to that of $\hat{F}_n$, the estimator of Huang and Wang (2005). Given (t_1, t_2), the standard deviations of both estimators decrease as τ_c increases, i.e. the proportion of subjects having at least one pair of recurrence times increases.

(2) For case 2, Table 2 indicates that the IPW estimator $\hat{F}_W$ works reasonably well. The biases are small for most of the cases considered and the standard deviations decreases as τ_c (or c) increases.

Table 1. Simulation results for $\hat{F}_n$ and $\hat{F}_W$ (Case 1)

					$\hat{F}_n$			$\hat{F}_W$	
θ	τ_c	t_1	t_2	bias	std	rmse	bias	std	rmse
3	8	0.5	1	0.002	0.022	0.022	-0.003	0.024	0.024
3	8	0.5	2	0.003	0.028	0.028	-0.002	0.030	0.030
3	8	0.5	4	0.003	0.032	0.032	-0.003	0.034	0.034
3	8	1.0	1	0.006	0.029	0.029	-0.004	0.030	0.030
3	8	1.0	2	0.000	0.035	0.035	-0.001	0.037	0.037
3	8	1.0	4	0.001	0.034	0.034	-0.003	0.036	0.036
3	8	2.0	1	0.006	0.030	0.030	0.004	0.032	0.032
3	8	2.0	2	-0.001	0.036	0.036	-0.003	0.038	0.038
3	8	2.0	4	0.000	0.032	0.032	-0.004	0.033	0.033
3	15	0.5	1	0.002	0.017	0.017	-0.001	0.019	0.019
3	15	0.5	2	0.003	0.021	0.021	-0.003	0.021	0.021
3	15	0.5	4	0.003	0.025	0.025	-0.003	0.025	0.025
3	15	1.0	1	0.007	0.023	0.024	-0.006	0.023	0.023
3	15	1.0	2	-0.000	0.026	0.026	-0.002	0.026	0.026
3	15	1.0	4	0.000	0.026	0.026	-0.001	0.027	0.027
3	15	2.0	1	0.006	0.024	0.025	0.004	0.024	0.024
3	15	2.0	2	-0.002	0.028	0.028	-0.003	0.028	0.028
3	15	2.0	4	-0.001	0.025	0.025	0.000	0.026	0.026
9	8	0.5	1	-0.003	0.024	0.024	-0.002	0.026	0.026
9	8	0.5	2	-0.002	0.029	0.029	-0.002	0.030	0.030
9	8	0.5	4	-0.003	0.033	0.033	-0.004	0.035	0.035
9	8	1.0	1	-0.001	0.029	0.029	-0.003	0.028	0.028
9	8	1.0	2	-0.002	0.034	0.034	-0.004	0.035	0.035
9	8	1.0	4	0.000	0.035	0.035	0.001	0.037	0.037
9	8	2.0	1	0.002	0.029	0.029	0.000	0.028	0.028
9	8	2.0	2	-0.001	0.036	0.036	0.001	0.036	0.036
9	8	2.0	4	0.002	0.032	0.032	0.006	0.033	0.033
9	15	0.5	1	-0.002	0.018	0.018	-0.002	0.021	0.021
9	15	0.5	2	-0.002	0.021	0.021	0.005	0.023	0.023
9	15	0.5	4	-0.004	0.026	0.026	-0.005	0.025	0.025
9	15	1.0	1	-0.001	0.022	0.022	0.004	0.024	0.024
9	15	1.0	2	0.001	0.026	0.026	-0.003	0.027	0.027
9	15	1.0	4	-0.003	0.025	0.025	-0.002	0.026	0.026
9	15	2.0	1	0.003	0.022	0.022	0.001	0.022	0.022
9	15	2.0	2	-0.002	0.028	0.028	-0.003	0.030	0.030
9	15	2.0	4	-0.004	0.024	0.024	-0.004	0.025	0.025

Table 2. Simulation results for $\hat{F}_W$ (Case 2)

				$\theta=3$			$\theta=9$		
τ_c	t_1	t_2	n	bias	std	rmse	bias	std	rmse
8	0.5	1	100	-0.006	0.031	0.031	-0.003	0.029	0.029
8	0.5	1	200	-0.004	0.024	0.024	-0.002	0.023	0.023
8	0.5	2	100	-0.008	0.037	0.038	-0.000	0.031	0.031
8	0.5	2	200	-0.009	0.030	0.031	-0.002	0.024	0.024
8	0.5	4	100	-0.011	0.044	0.045	-0.003	0.034	0.034
8	0.5	4	200	-0.007	0.037	0.038	-0.005	0.024	0.024
8	1.0	1	100	-0.008	0.039	0.040	-0.009	0.032	0.033
8	1.0	1	200	-0.006	0.030	0.031	-0.004	0.030	0.030
8	1.0	2	100	-0.006	0.052	0.052	0.001	0.040	0.040
8	1.0	2	200	-0.007	0.039	0.040	0.002	0.034	0.034
8	1.0	4	100	-0.013	0.053	0.055	0.001	0.038	0.038
8	1.0	4	200	-0.008	0.037	0.038	-0.002	0.032	0.032
8	2.0	1	100	-0.006	0.042	0.043	-0.008	0.034	0.035
8	2.0	1	200	0.000	0.029	0.029	0.001	0.030	0.030
8	2.0	2	100	-0.007	0.055	0.056	-0.008	0.041	0.042
8	2.0	2	200	-0.003	0.039	0.039	0.003	0.034	0.034
8	2.0	4	100	-0.015	0.049	0.051	-0.005	0.040	0.040
8	2.0	4	200	-0.010	0.035	0.036	-0.001	0.033	0.033
15	0.5	1	100	-0.003	0.034	0.034	0.001	0.028	0.028
15	0.5	1	200	-0.002	0.020	0.020	0.000	0.018	0.018
15	0.5	2	100	-0.003	0.038	0.038	0.001	0.035	0.035
15	0.5	2	200	-0.003	0.023	0.024	-0.000	0.025	0.025
15	0.5	4	100	-0.004	0.043	0.043	-0.001	0.041	0.041
15	0.5	4	200	-0.003	0.027	0.027	-0.002	0.025	0.025
15	1.0	1	100	-0.002	0.042	0.042	0.001	0.035	0.035
15	1.0	1	200	-0.003	0.022	0.022	0.000	0.021	0.021
15	1.0	2	100	0.006	0.052	0.052	0.002	0.040	0.040
15	1.0	2	200	0.004	0.025	0.026	0.003	0.026	0.026
15	1.0	4	100	0.002	0.048	0.048	0.004	0.042	0.042
15	1.0	4	200	-0.003	0.026	0.026	0.002	0.024	0.024
15	2.0	1	100	0.002	0.043	0.043	0.001	0.035	0.035
15	2.0	1	200	0.000	0.024	0.024	0.002	0.021	0.021
15	2.0	2	100	0.006	0.055	0.055	0.000	0.041	0.041
15	2.0	2	200	0.003	0.027	0.027	0.000	0.024	0.024
15	2.0	4	100	0.002	0.048	0.048	0.003	0.039	0.039
15	2.0	4	200	0.003	0.023	0.023	-0.002	0.026	0.026

Chapter 4

4. Discussion

The estimation of the bivariate distribution of recurrence times are important in analyzing the association of bivariate recurrent events. In this article, we propose IPW estimators for estimating bivariate recurrent times. For data of case 1, the propose estimator is almost as efficient as the estimator of Huang and Wang (2005). For data of case 2, simulation results indicate that the IPW estimator performs well. For informative censoring, the proposed IPW estimator also performs well and remains consistent if the censoring mechanism is estimated consistently. In some cases, assumption $(A.1)$ may be violated, e.g. there exits trend in the bivariate recurrence times. Further research is required in developing statistical methods to deal with this situation.

References

Anderson P. K. and Gill, R. D. (1982). Cox's regression model for counting processes. *The Annals of Statistics*, **10**, 1100-1120.

Clayton, D. G. (1978). A model for association in bivariate lifetables and its application in epidemiological studies of family tendency in chronic disease incidence. *Biometrika*, **65**, 141-151.

Huang, Y. and Louis, T. A. (1998). Nonparametric estimation of the joint distribution of survival time and mark variable. *Biometrika*, **85**, 785-798.

Huang, Y.-C. and Wang, M.-C. (2005). Nonparametric estimation of bivariate recurrent time distribution. *Biometrics*, **61**, 392-402.

Lin, D. Y., Sun, W., and Ying, Z. (1999). Nonparametric estimation of gap time distributions for serial events with censored data. *Biometrika*, **86**, 59-70.

Lin, D. Y., Wei, L. J., Yang, I., and Ying, Z. (2000). Semiparametric regression for the mean and rate functions of recurrent events. *Journal of the Royal Statistical Society B*, **62**, 711-730.

Pena, E. A., Strawderman, R. L., and Hollander, M. (2001). Nonparametric estimation with recurrent event data. *Journal of the American Statistical Association* **96**, 1299- 1315.

Pepe, M. S. and Cai, J. (1993). Some graphical displays and marginal regression analyses for recurrent failure times and time dependent covariates. *Journal of the American Statistical Association*, **88**, 811-820.

Prentice, R. L., Williams, B. J., and Peterson, A. V. (1981). On the regression analysis of multivariate failure time data. *Biometrika*, **68**, 373-379.

Visser, M. (1996). Nonparametric estimation of the bivariate survival function with application to vertically transmitted AIDS. *Biometrika*, **83**, 507-518.

Wang, M.-C. and Chang, S.-H. (1999). Nonparametric estimation of a recurrent survival function. *Journal of the American Statistical Association*, **94**, 146-153.

Wang, M.-C., Qin, J., and Chang, C.-T. (2001). Analyzing recurrent event data with informative censoring. *Journal of the American Statistical Association*, **96**, 1057-1065.

Wang, W.-J. and Wells, M. T. (1998). Nonparametric estimation of successive duration times under dependent censoring. *Biometrika*, **85**, 561-572.