三種不同資料結構下累積發生函數的無母數估計量研究
Nonparametric Estimation of the Cumulative Incidence Function under Three Types of Data Structures

研究生：陳建豪　　Student：Chien-Hao Chen

指導教授：王維菁　　Advisor：Weijing Wang

國立交通大學
統計學研究所
博士論文

A Thesis
Submitted to Institute of Statistics
College of Science
National Chiao Tung University
in partial Fulfillment of the Requirements
for the Degree of
PHD
in

Statistics

July 2008

Hsinchu, Taiwan, Republic of China

中華民國九十七年七月

三種不同資料結構下累積發生函數的無母數估計量研究

學生：陳建豪　　　　指導教授：王維菁

國立交通大學統計學研究所博士班

摘　　要

在本論文中，我們考慮兩個在生物醫學上常被應用的量：累積發生函數以及長期發生率；針對感興趣的發生原因，我們探討這兩個量的無母數估計。在三種不同的資料結構下(競爭風險和治癒模式在右設限存在下、競爭風險在左截切存在下)，我們分別應用三種不同的想法去得到無母數估計量：分解法，加權法以及補值法。在本文中，我們證明出在每一種資料結構下，使用不用想法所得到的無母數估計量都是相同的。另外，我們也利用數值分析來比較在競爭風險和治癒模式下，何者的無母數估計量更為有效率。

Nonparametric Estimation of the Cumulative Incidence Function under Three Types of Data Structures

student：Chien-Hao Chen

Advisors：Dr. Weijin Wang

Institute of Statistics
National Chiao Tung University

ABSTRACT

In this thesis we consider nonparametric inference of the cumulative incidence function for a particular type of failure and its long-term incidence rate, both of which are useful descriptive measures for biomedical data with multiple endpoints. A unified framework is provided to study different inference techniques under various incomplete data structures. Specifically three approaches, namely decomposition, weighting and imputation, are studied under data settings which include the conventional competing risks data, the framework of a cure model and truncated data. Identity between these methods for each data structure is examined. Numerical examples are provided for comparing the first two data formulations.

Key words: Competing Risks; Cure models; Imputation; Inverse probability weighting; Multi-state model; Nonparametric inference; Sufficient follow-up; Truncation.

誌　　謝

首先，感謝我的指導教授王維菁老師的指導，使我在統計的領域中有更為寬廣的視野，同時並不斷的指正、鼓勵，當我遇到困難時，指引我正確的研究方向，才能順利的完成這篇論文。另外，也要感謝黃信誠老師，陳鄰安老師，徐南蓉老師以及洪慧念老師撥冗前來擔任口試委員，並對本篇論文提出許多寶貴的意見，讓我獲益非淺。

這幾年博士班的生活，衷心感謝統計所所有老師的諄諄教誨，以及用心的指導。在此也感謝熱心的學長姐們，總是適時的給予幫助，使我解決了不少學習上的難題。

最後，感謝家人以及好友們給予我的協助與鼓勵，有你們作後盾，我才能走的那麼遠，並充滿勇氣的面對每一個挑戰。

陳建豪　謹誌于
國立交通大學
統計學研究所
2008　仲夏

Table of Contents

Chapter 1. Introduction

Multiple events data are commonly seen in many biomedical applications. Figure 1 depicts a common competing risks framework, in which there are $J + 1$ states, with state 0 being alive and states $1, 2, \ldots, J$ corresponding to the J distinct types of death. Figure 2 describes another situation in which a subject receiving heart transplantation may experience two different paths (i.e. with/without rejection).

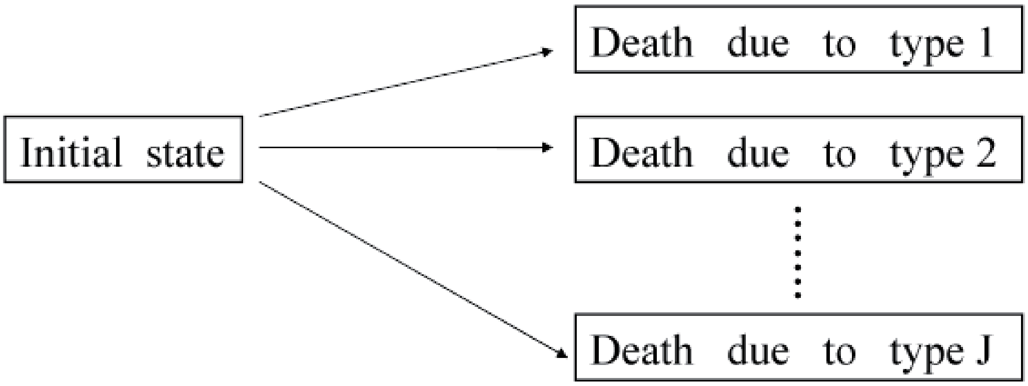

Figure 1: Competing risks of J distinct causes of death

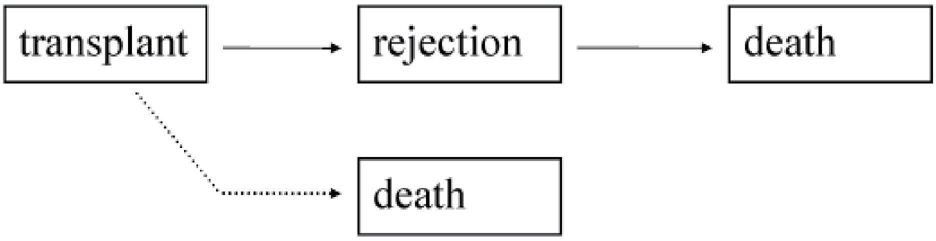

Figure 2: A two-path model for heart transplant

Suppose that a person may experience one of J distinct types of failure or J different paths. Let T be the failure time of the first event and B be the indicator for the corresponding failure type taking values $j = 1, \ldots, J$. The cumulative incidence function (CIF) for the jth cause of failure is defined as $F_j(t) = \Pr(T \le t, B = j)$. In the example of heart transplant, let $(B = 1)$ be the occurrence of rejection and $(B = 2)$ be the death without rejection. In this case, $F_1(t)$ is the cumulative probability of experiencing rejection by time t and $F_2(t)$ is the cumulative probability of death without rejection by time t.

Under the conventional framework of competing risks, the cumulative incidence function(CIF) can be considered as an extension of the Kaplan-Meier estimation with $J = 1$. Properties of the Kaplan-Meier estimator has been thoroughly studied in the literature. Specifically it is a product-limit estimator, satisfies the properties of redistribution-to-the-right and self-consistency, and is the nonparametric MLE. Recently Satten and Datta (2001) showed that the Kaplan-Meier estimator can also be expressed as an inverse-probability-of-censoring weighted average. In this thesis, we discuss how to generalize these inferential techniques to the estimation of CIF.

Most literature on CIF is developed under the competing risks framework. That is, one observes $\{(T_i, B_i)(i = 1, 2, \ldots, n)\}$ or its censored version. Nonparametric estimation of $F_j(t)$ in presence of right censoring has received substantial attention in the literature (Anderson et al., 1993; Kalbfleisch and Prentice, 2002). The martingale expression of the nonparametric MLE(NPMLE) was given in Lin (1997), which is useful for large-sample analysis. Satten and Datta (1999) showed that the NPMLE can also be re-expressed by a product-limit representation with fractional risk sets. For the simple case of $J = 2$, two recent articles by Betensky and Schoenfeld (2001) and Wang (2003) studied the estimation of $F_j(t)$ in the context of a two-state model. In practical applications, it has been founded that the CIF estimate is often misused by practitioners. Golley et al. (1999) and Farley et al. (2001) explained the distinction between the estimate of the cumulative incidence function and the complement of the Kaplan-Meier estimator. We will also address this issue under the framework of cure model.

In this thesis, we apply different inference techniques to estimate $F_j(t)$ and other related quantities(i.e. the long-term incidence rate) under different types of data structures. Chapter 2 reviews some relative papers about the cumulative incidence function. Chapter 3 considers the classical setting of competing risks data. Chapter 4 studies a different data formulation in the context of cure models in which subjects with $B \neq j$ are treated as being cured and hence will never experience the cause of interest, $B = j$, despite of long-term follow-up. In Chapter 5, we extend the results to competing risks data subject to left truncation. Relationships between different inference approaches are examined for each data structure. In Chapter 6 we provide numerical analysis to illustrate the difference between competing risks data and the data formulated based on the a cure model. Concluding remarks are given in Chapter 7.

Chapter 2. Literature Review

In this Chapter, we focus on the competing risks setting. Let T be the failure time of the first event and B be the indicator for the corresponding failure type taking values $j = 1, \ldots, J$. Let $S(t) = Pr(T > t)$, $F(t) = 1 - S(t)$ and

$$\lambda(t) = \lim_{\Delta t \to 0} \frac{\Pr(T \in [t, t + \Delta t) | T \geq t)}{\Delta t}. \tag{1}$$

The cumulative incidence function (CIF) of the jth event define as $F_j(t) = \Pr(T \leq t, B = j)$ which is the cumulative probability of observing the jth event by time t. However in practice, patients may drop out from the study or, at the end of the study, some may still have not developed any type of failure. Let C be the external censoring variable and assume that T and C are independent. Define $\Lambda_j(t) = \int_0^t \lambda_j(u) du$, where $\lambda_j(t)$ is the cause-specific hazard function defined as

$$\lambda_j(t) = \lim_{\Delta t \to 0} \frac{\Pr(T \in [t, t + \Delta t), B = j | T \geq t)}{\Delta t}. \tag{2}$$

In presence of right censoring, observable variables become $X = T \wedge C$ and $\tilde{B} = I(T \leq C) \cdot B$. Let $\{(T_i, C_i, B_i)\ (i = 1, \ldots, n)\}$ be *iid* replications of (T, C, B). Observed data can be expressed as $\{(X_i, \delta_i, \tilde{B}_i)\ (i = 1, \ldots, n)\}$, where $X_i = T_i \wedge C_i$ and $\widetilde{B}_i = \delta_i \cdot B_i = I(T_i \leq C_i) \cdot B_i$.

Consider the special case of estimating $S(t)$. The likelihood can be expressed as

$$\prod_{i=1}^{n} [dF(X_i)]^{\delta_i} S(X_i)^{1-\delta_i}$$

$$= \prod_{i=1}^{n} [d\Lambda(X_i)]^{\delta_i} S(X_i).$$

It has been shown that the NPMLE places mass only at observed failure times. For the general case $J > 1$, the likelihood becomes

$$\prod_{i=1}^{n} \{[\prod_{j=1}^{J} d\Lambda_j(X_i)^{I(\tilde{B}_i=j)}] \prod_{u \leq X_i} \{1 - \sum_{j=1}^{J} d\Lambda_j(u)\}\}.$$

Maximization of the above multinomial likelihood function, gives the MLE

$$d\hat{\Lambda}_j(t) = \frac{\sum_{i=1}^{n} I(X_i = t, \widetilde{B}_i = j)}{\sum_{i=1}^{n} I(X_i \geq t)}.$$

Lin(1997) showed that CIF has the following nice decomposition:

$$\begin{aligned} & \Pr(T \leq t, B = j) \\ = & \int_0^t \Pr(T = u, B = j | T \geq u) \Pr(T \geq u) \\ = & \int_0^t S(u-) d\Lambda_j(u), \end{aligned}$$

where $S(u) = \Pr(T > u)$ and $\Lambda_j(t) = \int_0^t \lambda_j(u) du$. By a simple plug-in approach, $F_j(t)$ can be estimated by

$$\hat{F}_j(t) = \int_0^t \hat{S}(u-) d\hat{\Lambda}_j(u), \tag{3}$$

where $\hat{S}(t)$ is the Kaplan-Meier estimator of $\Pr(T > t)$,

$$\hat{S}(t) = \prod_{u \leq t} \{1 - \frac{\sum_{i=1}^{n} I(X_i = u, \delta_i = 1)}{\sum_{i=1}^{n} I(X_i \geq u)}\}$$

and $d\hat{\Lambda}_j(t)$ is the Nelson-Aalen estimator of $d\Lambda_j(t)$,

$$d\hat{\Lambda}_j(t) = \frac{\sum_{i=1}^{n} I(X_i = t, \widetilde{B}_i = j)}{\sum_{i=1}^{n} I(X_i \geq t)}.$$

Lin (1997) mentioned that (3) is also the NPMLE. The paper also derives the martingale expression of $\sqrt{n}\{\hat{F}_1(t) - F_1(t)\}$ in which weak convergence properties of the process can be established. The results are useful for constructing confidence bands of $F_j(t)$ based on re-sampling techniques.

In survival analysis, there is a well-known relationship between the survival function and the cumulative hazard function. Specifically we have

$$S(t) = exp(-\int_0^t \lambda(u)du).$$

Therefore it seems natural to extend such a relationship to CIF and the cause-specific hazard function. However many researchers, including Golley et al. (1999) and Lin (1997) and Farley et al. (2001), found that

$$1 - F_j(t) \geq exp(-\int_0^t \lambda_j(u)du).$$

We will clarify this issue by introducing another data structure, namely the cure model framework.

Chapter 3. Competing Risks Data

Without external censoring, one observes $\{(T_i, B_i)(i = 1, \ldots, n)\}$. The empirical estimator of $F_j(t) = E[I(T \le t, B = j)]$ is given by

$$\bar{F}_j(t) = \sum_{i=1}^{n} I(T_i \le t, B_i = j)/n = \sum_{i=1}^{n} \int_0^t dI(T_i \le u, B_i = j)/n. \tag{4}$$

Competing risks data in presence of right censoring, can be expressed as $\{(X_i, \delta_i, \tilde{B}_i)$ $(i = 1, \ldots, n)\}$, where $X_i = T_i \wedge C_i$ and $\widetilde{B}_i = \delta_i \cdot B_i$. Notice the indicator of interest, $I(T_i \le t, B_i = j)$ may be missing due to censoring and hence the empirical estimator (4) is not applicable.

3.1. Estimation of Cumulative Incidence Function

As mentioned in Chapter 2, the NPMLE of $F_j(t)$ can be obtained explicitly by

$$\hat{F}_j^D(t) = \int_0^t \hat{S}(u-)d\hat{\Lambda}_j(u). \tag{5}$$

Here we apply two useful principles for handling missing data to estimate $F_j(t)$. One way is treating $I(X \le t, \tilde{B} = j)$ as a biased proxy of $I(T \le t, B = j)$ and then applying the technique of weighting to correct the sampling bias. Notice that

$$E[\frac{I(X \le t, \widetilde{B} = j)}{G(T-)}|T] = I(T \le t, B = j),$$

where $G(u) = \Pr(C > u)$ and $G(u-) = \Pr(C \ge u)$. Hence $F_j(t)$ can be estimated by the following weighted average:

$$\hat{F}_j^W(t) = \frac{1}{n}\sum_{i=1}^{n} \frac{I(X_i \leq t, \tilde{B}_i = j)}{\hat{G}(X_i-)} = \frac{1}{n}\int_{u \leq t} \frac{\sum_{i=1}^{n} dI(X_i \leq u, \tilde{B}_i = j)}{\hat{G}(u-)}, \tag{6}$$

where

$$\hat{G}(t-) = \prod_{u<t}\{1 - \frac{\sum_{i=1}^{n} I(X_i = u, \widetilde{B}_i = 0)}{\sum_{i=1}^{n} I(X_i \geq u)}\}.$$

Imputation is another way to deal with incomplete observations. By writing $F_j(t) = \int_0^t F_j(du)$, it follows that

$$F_j(du) = E[I(T \in [u, u+du), B = j)] = E\left[E[I(T \in [u, u+du), B = j | X, \widetilde{B})]\right]$$

where $E[I(T_i \in [u, u+du), B_i = j | X_i, \widetilde{B}_i)]$ can be written as

$$I(X_i \in [u, u+du), \widetilde{B}_i = j) + I(X_i < u, \widetilde{B}_i = 0)\frac{\Pr(T_i \in [u, u+du), B_i = j)}{S(X_i)}.$$

This approach yields the following self-consistent equation, $\widehat{F}_j^I(t) = \int_0^t \widehat{F}_j^I(\Delta u)$, where

$$\widehat{F}_j^I(\Delta u) = \sum_{i=1}^{n} I(X_i = u, \widetilde{B}_i = j)/n + \sum_{i=1}^{n} I(X_i < u, \widetilde{B}_i = 0)\widehat{F}_j^I(\Delta u)/n\widehat{S}(X_i). \tag{7}$$

Note that the first component of $\widehat{F}_j^I(\Delta u)$ is the original assigned mass at the failure time point u and the second component is the mass re-distributed from previously censored observations. The estimator $\widehat{F}_j^I(t)$ has the following explicit representation,

$$\hat{F}_j^I(t) = \int_{u \leq t} \frac{\sum_{i=1}^{n} dI(X_i \leq u, \tilde{B}_i = j)}{n - \sum_{i=1}^{n} I(X_i < u, \tilde{B}_i = 0)/\hat{S}(X_i)}.$$

Figure 3 provides a simple example to illustrate the steps of mass re-distribution for estimating $F_1(t)$ when there are two types of failure. Let $x_{(1)} < x_{(2)} < ... < x_{(n)}$ be ordered observations of X_i $(i = 1, \ldots, n)$ and $\delta_{(k)}$ and $\widetilde{B}_{(k)}$ be the indicators associated with $x_{(k)}$. At an observed point on the time line, the associated failure type is marked as $\bigcirc$, $\times$, and $\triangle$ for $\tilde{B} = 0, 1, 2$, respectively. In the first step, every point is assigned with mass $1/n = 1/7$. As can be seen in (7), the mass assigned to points with $\tilde{B} = 2$ has no contribution to the calculation of $\widehat{F}_1^I(t)$. Hence $\widehat{F}_1^I(x_{(1)}) = 1/7$ and $\widehat{F}_1^I(x_{(3)}) = 2/7$ since there are no censored observations with $\tilde{B} = 0$ before these two points. The mass assigned to the censored observation, $x_{(4)}$, will be evenly distributed to $x_{(j)}$ $(j = 5, 6, 7)$ and then the mass assigned to the censored observation, $x_{(6)}$, will be distributed to the point on its right, $x_{(7)}$. The redistribution-to-the right algorithm works in the same way as in the special case of $\Pr(B = 1)$ except that $\widehat{F}_j^I(\Delta u)$ does not receive mass from other competing events with $\tilde{B} \neq j$ $(j > 0)$.

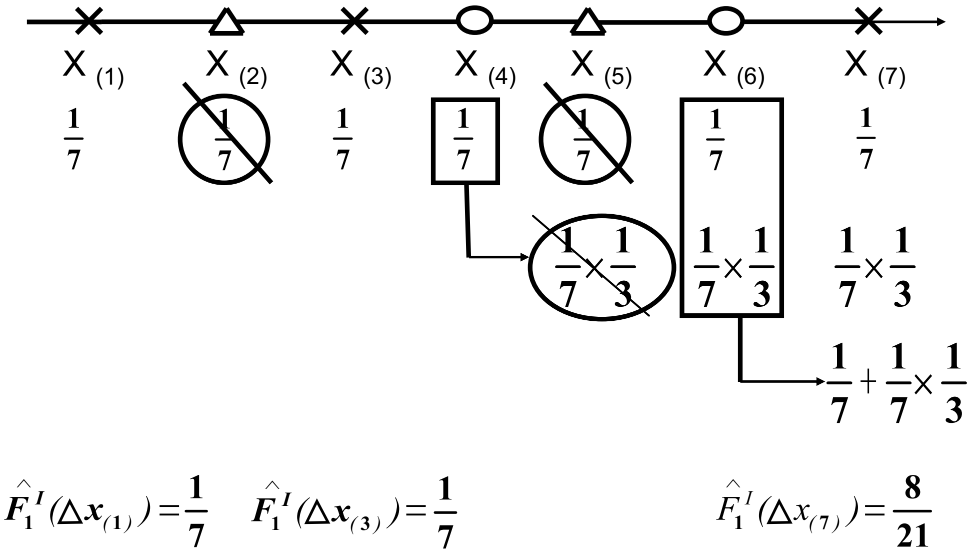
X(1)
X(2)
X(3)
X(4)
X(5)
X(6)
X(7)
1/7
1/7 × 1/3
1/7 + 1/7 × 1/3
$\hat{F}_1^I(\Delta x_{(1)}) = \frac{1}{7}$
$\hat{F}_1^I(\Delta x_{(3)}) = \frac{1}{7}$
$\hat{F}_1^I(\Delta x_{(7)}) = \frac{8}{21}$

Figure 3: Mass calculation for estimation of $\hat{F}_1^I(t)$ based on the first data structure, where ○ : $\tilde{B} = 0$;× : $\tilde{B} = 1$;△ : $\tilde{B} = 2$.

Now, we show that the above three estimators of $F_j(t)$ are equivalent. Specifically the jump sizes of the above estimators at time u are given by

$$\hat{F}_j^D(\Delta u) = \frac{\sum_{i=1}^n I(X_i = u, \tilde{B}_i = j)}{\sum_{i=1}^n I(X_i \geq u)/\hat{S}(u-)},$$

$$\hat{F}_j^W(\Delta u) = \frac{\sum_{i=1}^n I(X_i = u, \tilde{B}_i = j)}{n\hat{G}(u-)},$$

and

$$\hat{F}_j^I(\Delta u) = \frac{\sum_{i=1}^n I(X_i = u, \tilde{B}_i = j)}{n - \sum_{i=1}^n I(X_i < u, \tilde{B}_i = 0)/\hat{S}(X_i)},$$

where $u = x_{(k)}$ is an observed failure point with $\widetilde{B}_{(k)} = j$.

Theorem 1:*Identical relationship of $\widehat{F}_j^D(\Delta u)$, $\widehat{F}_j^W(\Delta u)$ and $\widehat{F}_j^I(\Delta u)$*

Proof:

According to the result of Satten and Datta (2001), we have

$$\hat{S}(u) \;=\; 1 - \frac{1}{n}\sum_{i=1}^{n}\frac{I(X_i \le u, \delta_i = 1)}{\hat{G}(X_i-)}, \tag{8}$$

$$\hat{G}(u-) \;=\; 1 - \frac{1}{n}\sum_{i=1}^{n}\frac{I(X_i < u, \tilde{B}_i = 0)}{\hat{S}(X_i)}. \tag{9}$$

By applying equation (9), it can be shown that the denominators of $\widehat{F}_j^W(\Delta u)$ and $\widehat{F}_j^I(\Delta u)$ are identical. In general, we can write

$$\hat{S}(u-)\hat{G}(u-) = \prod_{s<u}\left\{1 - \frac{\sum_{i=1}^{n} I(X_i = s, \delta_i = 1)}{\sum_{i=1}^{n} I(X_i \ge s)} - \frac{\sum_{i=1}^{n} I(X_i = s, \delta_i = 0)}{\sum_{i=1}^{n} I(X_i \ge s)} + R(s)\right\},$$

where

$$R(s) = \frac{\left(\sum_{i=1}^{n} I(X_i = s, \delta_i = 1)\right)\left(\sum_{i=1}^{n} I(X_i = s, \delta_i = 0)\right)}{\left(\sum_{i=1}^{n} I(X_i \ge s)\right)^2}.$$

If there are no ties between failure and censored observations, $R(s)$ should equal zero for all s and

$$\hat{S}(u-)\hat{G}(u-) = \prod_{s<u}\left\{1 - \frac{\sum_{i=1}^{n} I(X_i = s)}{\sum_{i=1}^{n} I(X_i \ge s)}\right\} = \frac{1}{n}\sum_{i=1}^{n} I(X_i \ge u)$$

and hence $\widehat{F}_j^W(\Delta u) = \widehat{F}_j^D(\Delta u)$. Note that if there are ties between failures and censored observations, we can still make the same conclusion if we impose the conventional assumption that the failures had occurred just before the censored observations.

From now on, we write $\hat{F}_j(t)$ to denote any of the three representations. The equivalence result has already been established for the special case $\Pr(B = 1) = 1$, where

$1 - F_1(t)$ reduces to $S(t)$ and $1 - \hat{F}_1(t)$ reduces to the Kaplan-Meier estimator, $\hat{S}(t)$. Note that Satten and Datta (2001) has shown that

$$\hat{S}(t) = \prod_{u \le t} \{1 - \frac{\sum_{i=1}^{n} I(X_i = u, \delta_i = 1)}{\sum_{i=1}^{n} I(X_i \ge u)}\} = 1 - \frac{1}{n} \sum_{i=1}^{n} \frac{I(X_i \le t, \delta_i = 1)}{\hat{G}(X_i-)},$$

which is a special case of equation (6).

The numerator in $\hat{F}_j(\Delta u)$, $\sum_{i=1}^{n} I(X_i = u, \tilde{B}_i = j)$, counts the number of observing type j failure at time u. The denominator has three equivalent representations which measure the "effective sample size" at time u adjusted for the underlying censoring mechanism. From the expression of $\hat{F}_j^W(\Delta u)$, we see that the effective sample size equals $n\hat{G}(u-)$ which is the (estimated) expected number of subjects that are not right censored at time u. The expression of $\hat{F}_j^I(\Delta u)$ indicates that previously censored observations should be excluded in the calculation of the effective sample size but their influence should be weighted according to their survival probabilities. The expression of $\hat{F}_j^D(\Delta u)$ indicates that the effective sample size is larger than the number at risk since subjects who had died previously should be included in the adjusted sample size which accounts for the possibility of external censoring.

Despite of the equivalence, different representations of the some quantity may provide different applications. The connection with the nonparametric MLE is useful to establish the optimal property. The idea of imputation is related to the re-distribution to the right and self-consistent properties which may be adopted in other inference problems that involve the covariate information or other censoring patterns. The weighted expression of $\hat{F}_j^W(t)$ is very useful for analytical analysis and may be generalized to other censoring mechanisms if the effect of censoring can be formulated explicitly.

3.2. Estimation of Long-Term Incidence Rate

The long-term incidence rate is defined as

$$\pi_j = \Pr(B = j) = \lim_{t \to \infty} \Pr(T \le t, B = j). \tag{10}$$

Estimation of π_j is not just a trivial extension of the previous results since as indicated in (10), π_j involves the tail information which may be missing if the quality of data is the distribution of $F_j(t)$ can not be recovered at the tail region. A crucial condition which measures the quality of data for recovering the tail information is called “sufficient follow-up“. For competing risks data, define the following boundary points, $\tau_T = \sup_t\{t : S(t) > 0\}$,$\tau_C = \sup_t\{t : G(t) > 0\}$, and $\tau_X = \sup_t\{t : \Pr(X > t) > 0\}$. Sufficient follow-up requires $\tau_T \le \tau_C$, which means that the follow-up time is long-enough to recover the whole information of $S(t)$ despite of censoring.

Let $x_{\max}^j$ be the largest failure point for failure type j. When $\tau_T \le \tau_C$, $\tau_X = \tau_T$, $x_{\max}^j$ will be a valid estimator for $\tau_T^j = \sup_t\{t : \Pr(T > t, B = j) > 0\}$ and hence $\hat{\pi}_j^D = \hat{F}_j(x_{\max}^j)$ will be a valid estimator for $\pi_j = F_j(\tau_T^j)$. Now, we show that $\hat{F}_j(x_{\max}^j)$ also has several equivalent representations. Specifically, the principle of weighting yields the estimator,

$$\hat{\pi}_j^W = \sum_{i=1}^{n} \frac{I(\tilde{B}_i = j)}{n\hat{G}(X_i-)}.$$

Applying the idea of imputation, we need to estimate

$$E[I(B_i = j)|\tilde{B}_i, X_i] = I(\tilde{B}_i = j) + I(\tilde{B}_i = 0)w(X_i),$$

where

$$w(X_i) = Pr(B_i = j|X_i, \tilde{B}_i = 0) = \frac{1}{S(X_i)} \int_{X_i}^{\infty} dF_j(u).$$

This approach was considered by Wang (2003) under the framework of a two-path model, where π_j represents the path probability. Using the plug-in principle, $w(X_i)$ can be estimated by

$$\hat{w}(X_i) = \{\int_{X_i}^{\infty} d\hat{F}_j(u)\}/\hat{S}(X_i) \tag{11}$$

and hence π_j can be estimated by

$$\hat{\pi}_j^I = \frac{1}{n} \sum_{i=1}^{n} \{I(\tilde{B}_i = j) + I(\tilde{B}_i = 0)\hat{w}(X_i)\}.$$

Theorem 2: $\hat{\pi}_j^I = \hat{\pi}_j^W = \hat{\pi}_j^D$

Proof:

The estimator $\hat{\pi}_j^I$ can be re-expressed as

$$\begin{aligned} \hat{\pi}_j^I &= \frac{1}{n} \sum_{i=1}^{n} \left\{ I(\widetilde{B}_i = j) + I(\widetilde{B}_i = 0)\hat{w}(X_i) \right\} \\ &= \frac{1}{n} \sum_{i=1}^{n} \left\{ I(\widetilde{B}_i = j) + \frac{I(\widetilde{B}_i = 0)}{\hat{S}(X_i)} \left[\sum_{k=1}^{n} \frac{I(X_k > X_i, \widetilde{B}_k = j)}{n\hat{G}(X_k-)} \right] \right\} \end{aligned}$$

$$= \quad \frac{1}{n}\sum_{i=1}^{n} I(\widetilde{B}_i = j) + \frac{1}{n}\sum_{k=1}^{n} \frac{I(\widetilde{B}_k = j)}{n\hat{G}(X_k-)} \left[\sum_{i=1}^{n} \frac{I(X_k > X_i, \widetilde{B}_i = 0)}{\hat{S}(X_i)} \right].$$

By equation (9),

$$\hat{\pi}_j^I = \frac{1}{n}\sum_{i=1}^{n} I(\widetilde{B}_i = j) + \sum_{k=1}^{n} \frac{I(\widetilde{B}_k = j)}{n\hat{G}(X_k-)} \left\{ 1 - \hat{G}(X_k-) \right\} = \sum_{k=1}^{n} \frac{I(\widetilde{B}_k = j)}{n\hat{G}(X_k-)} = \hat{\pi}_j^W.$$

Furthermore it follows that

$$\hat{\pi}_j = \sum_{i=1}^{n} \frac{I(\widetilde{B}_i = j)}{n\hat{G}(X_i-)} = \sum_{i=1}^{n} \frac{I(X_i \le x_{\max}^j, \widetilde{B}_i = j)}{n\hat{G}(X_i-)} = \hat{F}_j(x_{\max}^j),$$

where $x_{\max}^j$ is the largest observed failure time for type j event.

We see again that the three methods lead to the same estimator of π_j, which will be denoted by $\hat{\pi}_j$.

For competing risks data, insufficient follow-up means that $\tau_T > \tau_C$. There will be large censored observations at the tail so that $\hat{S}(x_{(n)}) > 0$ and $\hat{G}(x_{(n)}) = 0$, where $x_{(n)}$ is the largest value of X. Consequently there will be some mass beyond $x_{\max}^j$ which is not estimable and hence $\hat{\pi}_j$ would underestimate π_j. In fact, nonparametric estimation of any quantity that involves the tail information has the problem of non-identifiability if follow-up is not sufficient. To accommodate the situation of insufficient follow-up, Wang (2003) suggested to impose an artificial assumption, $\Pr(B = j|T > \tau_X) = \pi_j$, which implies that the missing tail area contains the same information about the quantity of interest as in the whole support. The resulting estimator becomes

$$\breve{\pi}_j^I = \frac{\sum_{i=1}^n \{I(\tilde{B}_i = j) + I(\tilde{B}_i = 0)\hat{w}(X_i)\}}{\breve{n}}, \tag{12}$$

where $\breve{n} = n - n_c\hat{S}(x_{(n)})$ and $n_c = \sum_{i=1}^n I(\tilde{B}_i = 0)/\hat{S}(X_i)$. Alternatively, we can apply the weighting approach to estimate π_j by,

$$\breve{\pi}_j^W = \frac{1}{\breve{n}} \sum_{i=1}^n \frac{I(\widetilde{B}_i = j)}{\hat{G}(X_i-)},$$

and

$$\breve{n} = n - n_c\hat{S}(x_{(n)}) = \sum_{i=1}^n \frac{I(\delta_i = 1)}{\hat{G}(X_i-)}.$$

Theorem 3: $\breve{\pi}_j^I(t) = \breve{\pi}_j^W(t)$ *(adjusted for insufficient follow up)*

Proof:

The original expression of $\breve{\pi}_j^I$ is given by

$$\breve{\pi}_j^I = \frac{\sum_{i=1}^n \left\{ I(\widetilde{B}_i = j) + \frac{I(\widetilde{B}_i=0)}{\hat{S}(X_i)} \left[\sum_{k=1}^n \frac{I(X_k>X_i, \widetilde{B}_k=j)}{n\hat{G}(X_k-)} \right] \right\}}{n - n_c\hat{S}(x_{(n)})}.$$

To show $n - n_c\hat{S}(x_{(n)}) = \sum_{i=1}^n \frac{I(\delta_i=1)}{\hat{G}(X_i-)}$, we need to consider the following two cases. Let $\delta_{(n)}$ be the indicator associated with $x_{(n)}$. If $\delta_{(n)} = 1$, which means that $\hat{S}(x_{(n)}) = 0$, by equation (8), we have $\sum_{i=1}^n \frac{I(\delta_i=1)}{\hat{G}(X_i-)} = n$. If $\delta_{(n)} = 0$ which implies that $\hat{G}(x_{(n)}) = 0$ and by equation (9), we have $n_c = \sum_{i=1}^n \frac{I(\widetilde{B}_i=0)}{\hat{S}(X_i)} = n$ and hence

$$n - n_c\hat{S}(x_{(n)}) = n(1 - \hat{S}(x_{(n)})) = \sum_{i=1}^n \frac{I(\delta_i = 1)}{\hat{G}(X_i-)}.$$

Hence in either case, we have that

$$n - n_c\hat{S}(x_{(n)}) = \sum_{i=1}^{n} \frac{I(\delta_i = 1)}{\hat{G}(X_i-)},$$

and it follows that

$$\breve{\pi}_j^I = \frac{\sum_{i=1}^{n} I(\widetilde{B}_i = j)\Big/ \hat{G}(X_i-)}{\sum_{i=1}^{n} I(\delta_i = 1)/\,\hat{G}(X_i-)}.$$

Thus one can view $\breve{n}$ as the sample size that only measures the identifiable region.

When $\hat{S}(x_{(n)}) = 0$, which is an evidence of sufficient follow-up, we have $\breve{n} = n$.

Chapter 4. Cure Model Framework

In classical survival analysis, an implicit but sometimes implausible assumption is that every individual will eventually experience the event of interest. Cure models allow the possibility that a subject may not develop the event of interest despite of long-term follow-up. For more discussions on cure models, one can refer to the book by Maller and Zhou (1996). Data with multiple endpoints can be formulated in the context of cure models if a particular type of event, say type j, is of major interest. Under the framework of a cure model, those who will never experience this type of failure $(B \neq j)$ are treated as being immune (cured, or non-susceptible) for the event of type j. One can define T_j as the hypothetical failure time such that

$$T_j = T \cdot I(B = j) + \infty \cdot I(B \neq j).$$

For $t < \infty$, the cumulative incidence function can be written as

$$F_j(t) = \Pr(T \leq t, B = j) = \Pr(T_j \leq t)$$

and hence its complement becomes

$$S_j(t) = \Pr(T_j > t) = \Pr(T > t, B = j) + \Pr(B \neq j).$$

When there is more than one type of failure, $1 - \pi_j = \lim_{t\to\infty} S_j(t) > 0$. In such a case, T_j is not a proper random variable.

It is useful to compare the failure time variables discussed in the paper via their hazard functions. The hazard functions for T and T_j are given by

$$\lambda(t) = \lim_{\Delta t \to 0} \frac{\Pr(T \in [t, t+\Delta t) | T \geq t)}{\Delta t} \tag{13}$$

and

$$\tilde{\lambda}_j(t) = \lim_{\Delta t \to 0} \frac{\Pr(T_j \in [t, t+\Delta t) | T_j \geq t)}{\Delta t}, \tag{14}$$

respectively. The relationship between a hazard function and its survival function holds for T and T_j such that

$$S(t) = \Pr(T > t) = \exp(-\int_0^t \lambda(u) du),$$

and

$$S_j(t) = \Pr(T_j > t) = \exp\{-\int_0^t \tilde{\lambda}_j(u) du\}.$$

In presence of competing risks, the cause-specific hazard for type j failure is defined as

$$\lambda_j(t) = \lim_{\Delta t \to 0} \frac{\Pr(T \in [t, t+\Delta t), B = j | T \geq t)}{\Delta t}. \tag{15}$$

Note that that $\lambda(t) = \sum_{j=1}^{J} \lambda_j(t)$. It is easy to see that $\tilde{\lambda}_j(t) \leq \lambda_j(t)$ since for $t < \infty$

$$\Pr(T_j \in [t, t+\Delta t)) = \Pr(T \in [t, t+\Delta t), B = j)$$

but $\Pr(T_j \geq t) \geq \Pr(T \geq t)$. The latter implies that the risk set at time t for T_j is a larger set which includes not only those with $T \geq t$ but also those with $T < t, B = k$ $(k \neq j)$.

Consequently

$$S_j(t) = \exp\{-\int_0^t \tilde{\lambda}_j(u)du\} \geq \exp\{-\int_0^t \lambda_j(u)du\}. \tag{16}$$

In summary, we provide a new explanation of (16) which has been discussed in Lin (1997), and Gooley et al. (1997).

4.1. Estimation of Cumulative Incidence Function

In the context of cure models, censored data based on T_j can be formulated as $\{(\check{X}_i^j, \check{\delta}_i^j)\ (i = 1, \ldots, n)\}$, where $\check{X}_i^j$ and $\check{\delta}_i^j$ are *iid* replications of $\check{X}^j = T_j \wedge C$ and $\check{\delta}^j = I(T_j \leq C)$ respectively. It is important to note that $\check{\delta}_i^j = 0$ corresponds to two possible situations of $\tilde{B}_i$, namely $\tilde{B}_i = 0$ and $\tilde{B}_i = k \neq j\ (k > 0)$, provided by the competing risks data structure. However for the cure model considered here, when $\check{\delta}_i^j = 0$, only $\check{X}_i^j = C_i$ is recorded and the information of other competing risks is ignored. Figure 4 illustrates the relationship between the two data structures. The first data structure can be converted to the second one if the failure time associated with $\tilde{B} = 2$ is replaced by the time to the external censoring event, which requires additional information.

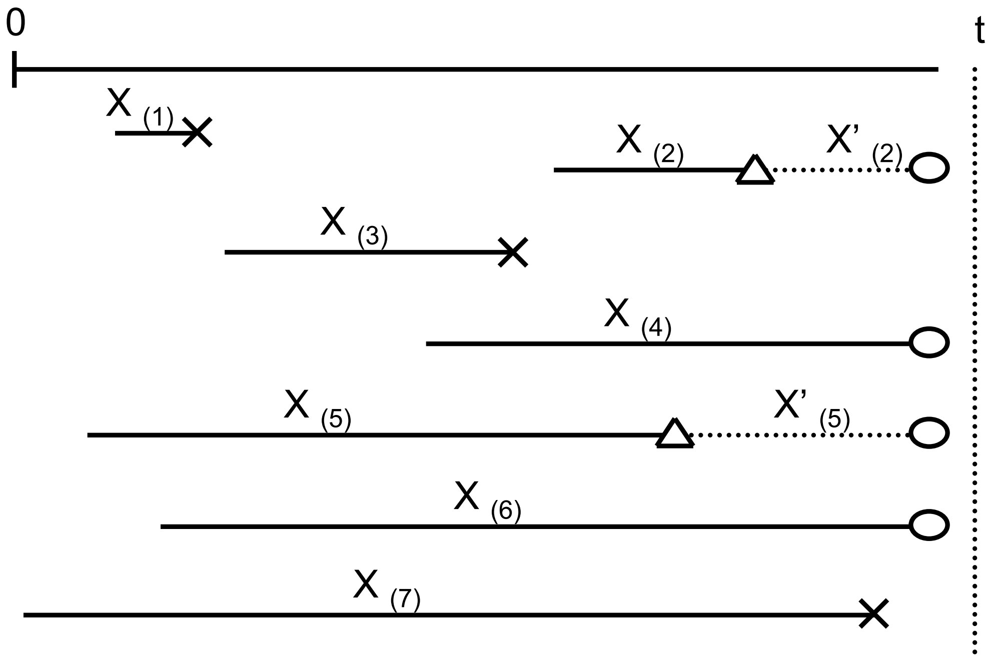

Figure 4: Relationship between the two data structures. The difference is that the second type of data extends the endpoint of the competing risk event marked by △ to the endpoint of external censoring marked by ○.

By independence between T_j and C and equation (16), $S_j(t)$ can be consistently estimated by the Kaplan-Meier estimator,

$$\check{S}_j(t) = \prod_{u \le t} \{1 - \frac{\sum_{i=1}^{n} I(\check{X}_i^j = u, \check{\delta}_i^j = 1)}{\sum_{i=1}^{n} I(\check{X}_i^j \ge u)}\}$$

and hence $F_j(t)$ can be estimated by $\check{F}_j(t) = 1 - \check{S}_j(t)$. As $\hat{F}_j(t)$, $\check{F}_j(t)$ also has different but equivalent representations. By similar arguments, one can show that

$$\check{F}_j(t) \quad = \quad \sum_{i=1}^{n} \frac{I(\check{X}_i^j \le t, \check{\delta}_i^j = 1)}{n\check{G}(\check{X}_i^j -)} \qquad (17)$$

$$= \int_{u \le t} \frac{\sum_{i=1}^{n} dI(\check{X}_i^j \le u, \check{\delta}_i^j = 1)}{n - \sum_{i=1}^{n} I(\check{X}_i^j < u, \check{\delta}_i^j = 0)/\check{S}_j(\check{X}_i^j)}, \tag{18}$$

which is also the nonparametric MLE, where

$$\check{G}(t) = \prod_{u \le t} \{1 - \frac{\sum_{i=1}^{n} I(\check{X}_i^j = u, \check{\delta}_i^j = 0)}{\sum_{i=1}^{n} I(\check{X}_i^j \ge u)}\}.$$

It should be mentioned that sometimes practitioners may estimate $F_j(t)$ by the complement of the wrong Kaplan-Meier estimator

$$\check{S}_j^0(t) = \prod_{u \le t} \{1 - \frac{\sum_{i=1}^{n} I(X_i = u, \tilde{B}_i = j)}{\sum_{i=1}^{n} I(X_i \ge u)}\},$$

which actually estimates $\exp\{-\Lambda_j(t)\}$ but $S_j(t) \ge \exp\{-\Lambda_j(t)\}$ as claimed in (16). The misuse of $1 - \check{S}_j^0(t)$ to estimate $F_j(t)$ has been noticed by many authors (Farley et al., 2001; Gooley et al.,1999; Lin, 1997; Wang, 2003), just to name a few.

4.2. Comparison of the Two Estimators

Now we compare the two estimators derived under different data structures based on their weighted expressions given in (6) and (17) respectively. We find that

$$I(\check{X}_i^j = u, \check{\delta}_i^j = 1) = I(X_i = u, \tilde{B}_i = j).$$

Hence these estimators only differ in the denominator that involves the estimator of $G(t)$. If the main purpose of inference was to estimate $G(t)$, $\check{G}(t)$ is clear a better choice since $\Pr(\check{\delta} = 0) \ge \Pr(\tilde{B} = 0)$. In fact, $\check{G}(t)$ is less variable than $\hat{G}(t)$. However the estimator of $G(t)$ is now used as a reciprocal weight to obtain an estimator of $F_j(t)$. It turns out that a better estimator of $G(t)$ does not necessarily yield a better estimator of $F_j(t)$. In the following, we show that $\hat{F}_j(t)$ has smaller asymptotic variance than $\check{F}_j(t)$. This phenomenon was first noticed by Tsai and Crowley (1998) in a different estimation problem that also involves inverse-probability-weighting. To better understand the effect of the estimator of $G(t)$, we also evaluate two hypothetical estimators,

$$\bar{F}_j(t) = \sum_{i=1}^{n} \frac{I(X_i \leq t, \tilde{B}_i = j)}{n\bar{G}(X_i-)} \tag{19}$$

where $\bar{G}(t) = \sum_{i=1}^{n} I(C_i > t)/n$, and

$$\bar{F}_j^*(t) = \sum_{i=1}^{n} \frac{I(X_i \leq t, \tilde{B}_i = j)}{nG(X_i-)}. \tag{20}$$

Now, we show that although

$$0 = \text{AVar}(G(t)) \leq \text{AVar}(\bar{G}(t)) \leq \text{AVar}(\check{G}(t)) \leq \text{AVar}(\hat{G}(t)),$$

the relationship for the estimators of $F_j(t)$ is reverse such that

$$\text{AVar}(\bar{F}_j^*(t)) \geq \text{AVar}(\bar{F}_j(t)) \geq \text{AVar}(\check{F}_j(t)) \geq \text{AVar}(\hat{F}_j(t)).$$

Theorem 4: AVar$(\bar{F}_j^*(t)) \geq$ AVar$(\bar{F}_j(t)) \geq$ AVar$(\check{F}_j(t)) \geq$ AVar$(\hat{F}_j(t))$.

Proof:

Based on the martingale representation of $\hat{G}(t)$, one can write

$$\sqrt{n}\{G(t) - \hat{G}(t)\} = n^{-1/2} G(t) \sum_{i=1}^{n} \int_0^t \frac{dM_{ci}(u)}{\Pr(X \geq u)} + o_p(1),$$

where

$$M_{ci}(u) = I(X_i \le u, \tilde{B}_i = 0) - \int_0^u I(X_i \ge s)\Lambda_c(ds)$$

and $\Lambda_c(s)$ is the cumulative hazard function of C. It follows that

$$\sqrt{n}\{\hat{F}_j(t) - F_j(t)\} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^t \frac{dI(X_i \le u, \widetilde{B}_i = j) - dH_j(u)}{G(u-)}$$

$$+ \quad \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^t [\int_0^u \frac{dI(X_i \le s, \widetilde{B}_i = 0) - I(X_i \ge s)d\Lambda_c(s)}{\Pr(X_i \ge s)}]dF_j(u) + o_p(1),$$

where $H_j(t) = \Pr(X \le t, \widetilde{B} = j)$. The asymptotic variance of $\hat{F}_j(t)$ equals $\mathrm{Var}(A(t)) + 2\mathrm{Cov}(A(t), B(t)) + \mathrm{Var}(B(t))$, where

$$\begin{aligned} A(t) &= \int_0^t \frac{dI(X \le u, \widetilde{B} = j) - dH_j(u)}{G(u-)}, \\ B(t) &= \int_0^t \int_0^u \frac{dM_c(s)}{\Pr(X \ge s)} dF_j(u) \\ M_c(u) &= I(X \le u, \tilde{B} = 0) - \int_0^u I(X \ge s)\Lambda_c(ds). \end{aligned}$$

It follows that

$$Var(A(t)) = E[A^2(t)] = \int_0^t \frac{dF_j(u)}{G(u-)} - F_j^2(t),$$

$$\begin{aligned} Var(B(t)) &= \int_0^t \int_0^t E\left[\int_0^u \frac{dM_c(s)}{\Pr(X \ge s)} \cdot \int_0^v \frac{dM_c(w)}{\Pr(X \ge w)}\right] dF_j(u)dF_j(v) \\ &= \int_0^t \int_0^t \int_0^{u \wedge v} \frac{d\langle M_c, M_c\rangle(s)}{\Pr^2(X \ge s)} dF_j(u)dF_j(v), \end{aligned}$$

$$= \int_0^t \int_0^t \int_0^{u \wedge v} \frac{d\Lambda_c(s)}{\Pr(X \geq s)} dF_j(u) dF_j(v).$$

The covariance term equals

$$
\begin{aligned}
& E[A(t)B(t)] \\
= \; & E\left\{ \int_0^t \frac{dI(X \leq v, \widetilde{B} = j) - dH_j(v)}{G(v-)} \cdot \int_0^t \int_0^u \frac{dM_c(s)}{\Pr(X \geq s)} dF_j(u) \right\} \\
= \; & \int_0^t \int_0^t \frac{1}{G(v-)} \\
& \cdot E\left\{ \left[dI(X \leq v, \widetilde{B} = j) - dH_j(v) \right] \cdot \int_0^u \frac{dI(X \leq s, \widetilde{B} = 0) - I(X \geq s) d\Lambda_c(s)}{\Pr(X \geq s)} \right\} dF_j(u) \\
= \; & \int_0^t \int_0^t \frac{1}{G(v-)} \\
& \cdot E\left\{ dI(X \leq v, \widetilde{B} = j) \cdot \int_0^u \frac{dI(X \leq s, \widetilde{B} = 0) - I(X \geq s) d\Lambda_c(s)}{\Pr(X \geq s)} \right\} dF_j(u) \\
= \; & -\int_0^t \int_0^t \frac{1}{G(v-)} E\left\{ dI(X \leq v, \widetilde{B} = j) \cdot \int_0^u \frac{I(X \geq s) d\Lambda_c(s)}{\Pr(X \geq s)} \right\} dF_j(u) \\
= \; & -\int_0^t \int_0^t \frac{1}{G(v-)} E\left\{ dI(X \leq v, \widetilde{B} = j) \cdot \int_0^u \frac{d\Lambda_c(s)}{\Pr(X \geq s)} \right\} dF_j(u) \\
= \; & -\int_0^t \int_0^t \int_0^{u \wedge v} \frac{d\Lambda_c(s)}{\Pr(X \geq s)} dF_j(u) dF_j(v).
\end{aligned}
$$

Asymptotic variance of $\sqrt{n}\{\hat{F}_j(t) - F_j(t)\}$ equals

$$\left[\int_0^t \frac{dF_j(u)}{G(u-)} - F_j^2(t)\right] - \int_0^t \int_0^t \int_0^{u \wedge v} \frac{d\Lambda_c(s)}{\Pr(X \geq s)} dF_j(u) dF_j(v). \tag{21}$$

Hence asymptotic variance of $\sqrt{n}\{\check{F}_j(t) - F_j(t)\}$ can be derived in the same way, which is given by

$$\left[\int_0^t \frac{dF_j(u)}{G(u-)} - F_j^2(t)\right] - \int_0^t \int_0^t \int_0^{u \wedge v} \frac{d\Lambda_c(s)}{\Pr(\check{X} \geq s)} dF_j(u) dF_j(v). \tag{22}$$

Notice that (21) and (22) only differ by one term in the the second component. Because $\Pr(\check{X} \geq s) \geq \Pr(X \geq s)$, it is easy to see that

$$\mathrm{AVar}(\sqrt{n}\{\check{F}_j(t) - F_j(t)\}) \geq \mathrm{AVar}(\sqrt{n}\{\hat{F}_j(t) - F_j(t)\}).$$

The asymptotic variance of $\sqrt{n}\{\bar{F}_j(t) - F_j(t)\}$ is

$$\left[\int_0^t \frac{dF_j(u)}{G(u-)} - F_j^2(t)\right] - \int_0^t \int_0^t \int_0^{u \wedge v} \frac{d\Lambda_c(s)}{\Pr(C \geq s)} dF_j(u) dF_j(v), \tag{23}$$

which can be further simplified as

$$\left[\int_0^t \frac{dF_j(u)}{G(u-)} - F_j^2(t)\right] - \int_0^t \int_0^t (\frac{1}{G(u \wedge v)} - 1) dF_j(u) dF_j(v),$$

and the asymptotic variance of $\sqrt{n}\{\bar{F}_j^*(t) - F_j(t)\}$ is

$$\int_0^t \frac{dF_j(u)}{G(u-)} - F_j^2(t). \tag{24}$$

It is clear that

$$\mathrm{AVar}(\bar{F}_j^*(t)) \geq \mathrm{AVar}(\bar{F}_j(t)) \geq \mathrm{AVar}(\check{F}_j(t)) \geq \mathrm{AVar}(\hat{F}_j(t)). \tag{25}$$

4.3. Estimation of long-term incidence rate

In presence of external censoring, the non-identifiable region for competing risks data refers to the set, $\{t : t > \tau_T \wedge \tau_C\}$ and the non-identifiable region for data of the cure model is $\{t : t > \tau_T^j \wedge \tau_C\}$. It is important to note that $\tau_T^j < \tau_T$. Clearly the second set is larger than the first one. When follow-up is sufficient, we have $\tau_T^j < \tau_T < \tau_C$ and both data structures provide enough information for estimating π. However when follow-up is not sufficient, the two data structures contain different level of information about the tail distribution.

Based on the second type of data $\{(\breve{X}_i^j, \breve{\delta}_i^j)\ (i = 1, \ldots, n)\}$, sufficient follow-up requires $\tau_T^j \leq \tau_C$, which implies that the length of follow-up is long enough to observe all the susceptible with $B = j$. If $\tau_T^j \leq \tau_C$, π_j can be consistently estimated by the following three expressions

$$\breve{\pi}_j = 1 - \breve{S}_j(x^j_{\max}) = \sum_{i=1}^{n} \frac{I(\breve{\delta}_i^j = 1)}{n\breve{G}(\breve{X}_i^j -)} = \frac{1}{n}\sum_{i=1}^{n} \left\{ I(\breve{\delta}_i^j = 1) + I(\breve{\delta}_i^j = 0)\breve{w}(\breve{X}_i^j) \right\},$$

where

$$\breve{w}(t) = \frac{\sum_{k=1}^{n} I(\breve{X}_k^j > t, \breve{\delta}_k^j = 1)\breve{F}_j(\Delta \breve{X}_k^j)}{\breve{S}_j(t)}.$$

For the cure model, insufficient follow-up means that $\tau_T^j > \tau_C$. Consequently $\breve{\pi}_j$ also underestimates π_j. For competing risks data, the probability of the non-identifiable region $\{t : t > \tau_T \wedge \tau_C\}$ is identifiable and is estimated by $\hat{S}(x_{(n)})$ which provides the foundation for us to re-allocate the mass under the artificial assumption $\Pr(B = j|T >$

$\tau_X) = \pi_j$. However for the second data structure, the mass of the non-identifiable region $\{t : t > \tau_T^j \wedge \tau_C\}$ is still not identifiable without imposing distributional assumptions. It seems that it is impossible to construct a valid estimator of π_j nonparametrically. The book by Maller and Zhou (1996) has thorough discussions on the problem of non-identifiability in the context of cure models.

Chapter 5. Extension to Truncation Data

We discuss whether the two suggested estimation principles can be extended to incorporate left truncation. Suppose that T is the age when the first failure occurs and B is the corresponding failure type. Let L be the age that a subject enters the study. If the sample only includes those who have not developed any type of failure, we can imagine that the sample is subject to left truncation such that only those with $T > L$ can be included in the sample. Let (T_i, B_i, L_i) $(i = 1, \ldots, m)$ be a random sample of (T, B, L). Under left truncation, we only observe the sample (T_i, B_i, L_i) $(i = 1, \ldots, n)$ with $T_i > L_i$ and m is unknown. Let $F_L(t) = \Pr(L \leq t)$ be the distribution function of L. For truncated data, the marginal functions $S(t)$ and $F_L(t)$ can be estimated by the Lynden-Bell estimators

$$\tilde{S}(t) = \prod_{u \leq t} \{1 - \frac{\sum_{i=1}^{n} I(T_i = u, L_i \leq u)}{\sum_{i=1}^{n} I(L_i \leq u \leq T_i)}\},$$

$$\tilde{F}_L(t) = \prod_{u > t} \{1 - \frac{\sum_{i=1}^{n} I(L_i = u, T_i \geq u)}{\sum_{i=1}^{n} I(L_i \leq u \leq T_i)}\}.$$

We first study estimation of m and $\alpha = \Pr(T \geq L)$. Let $F(t) = 1 - S(t)$. Based on the decomposition that $\alpha = \int F_L(u) dF(u)$, we have $\tilde{\alpha}^D = \int \tilde{F}_L(u) d\tilde{F}(u)$, where $\tilde{F}(t) = 1 - \tilde{S}(t)$. Based on another less straightforward decomposition,

$$F_L(t)S(t-) = \Pr(L \leq t)\Pr(T \geq t) = \alpha \Pr(L \leq t \leq T | T \geq L),$$

we may estimate α by

$$\tilde{\alpha}(t) = \frac{\tilde{F}_L(t)\tilde{S}(t-)}{\sum_{i=1}^{n} I(L_i \le t \le T_i)/n}.$$

The weighting approach can be applied based on the relationship

$$E[\frac{1}{F_L(T)}|T \ge L] = \frac{1}{\Pr(T \ge L)}E[\frac{I(T \ge L)}{F_L(T)}] = \frac{1}{\alpha},$$

which implies that α can be estimated by

$$\tilde{\alpha}^w = \{\frac{1}{n}\sum_{i=1}^{n}\frac{1}{\tilde{F}_L(T_i)}\}^{-1}.$$

It is important to mention that, in presence of truncation, the method of imputation seems not applicable since an truncated observation is completely missing and even its existence is unknown. He and Yang (1998) considered estimation of α and, in their Theorem 2.2, it is shown that $\tilde{\alpha}(t) = \tilde{\alpha}^D$ for any t such that $\sum_{i=1}^{n} I(L_i \le t \le T_i) > 0$. Now, we show that $\tilde{\alpha}(t) = \tilde{\alpha}^D = \tilde{\alpha}^w$.

Theorem 5: $\tilde{\alpha}(t) = \tilde{\alpha}^D = \tilde{\alpha}^w$

Proof:

From (2.8)of He and Yang (1998), we have $\sum_{i=1}^{n} I(T_i = t)/n = \tilde{F}_L(t)\Delta\tilde{F}(t)/\tilde{\alpha}^D$ and hence $\Delta\tilde{F}(t) = \tilde{\alpha}^D \sum_{i=1}^{n} I(T_i = t)/n\tilde{F}_L(t)$. By taking summation over all t on both sides of the previous identity, we have

$$1 = \sum_{i=1}^{n}\Delta\tilde{F}(T_i) = \tilde{\alpha}^D\sum_{i=1}^{n}\frac{1}{n\tilde{F}_L(T_i)},$$

which means that $\tilde{\alpha}^D = \{\frac{1}{n}\sum_{i=1}^{n}\frac{1}{\tilde{F}_L(T_i)}\}^{-1} = \tilde{\alpha}^w$.

From the above two estimators, we see the jump sizes at time u are given by

$$\tilde{F}_j^w(\Delta u) = \{\frac{1}{n}\sum_{i=1}^{n}\frac{1}{\tilde{F}_L(T_i)}\}^{-1}\cdot\{\frac{\sum_{i=1}^{n} I(T_i = u, B_i = j, T_i \geq L_i)}{n\tilde{F}_L(u)}\},$$

and

$$\tilde{F}_j^D(\Delta u) = \{\frac{\sum_{i=1}^{n} I(T_i = u, B_i = j, T_i \geq L_i)}{\sum_{i=1}^{n} I(L_i \leq u \leq T_i)/\tilde{S}(u-)}\}.$$

Proving $\tilde{F}_j^w(\Delta u) = \tilde{F}_j^D(\Delta u)$ is the same as showing $\tilde{\alpha}^D = \tilde{\alpha}^w$.

It follows that the total sample size m can be estimated by

$$\tilde{m} = \frac{n}{\int \tilde{F}_L(u)d\tilde{F}(u)} = \sum_{i=1}^{n}\frac{1}{\tilde{F}_L(T_i)} = \frac{\sum_{i=1}^{n} I(L_i \leq t \leq T_i)}{\tilde{F}_L(t)\tilde{S}(t-)}. \qquad (26)$$

Since $\sum_{i=1}^{n} I(T_i = u, B_i = j)/n$ is an estimator of $F_j(\Delta u)\Pr(L < u)/\alpha$, it follows that $F_j(\Delta u)$ can be estimated by $\sum_{i=1}^{n} I(T_i = u, B_i = j)/\tilde{m}\hat{F}_L(u)$, where $\tilde{m}$ can be any of the above estimators in (26). The method of decomposition yields the estimator

$$\tilde{F}_j(t) = \int_{u\leq t}\tilde{S}(u-)d\tilde{\Lambda}_j(u),$$

where

$$\tilde{\Lambda}_j(u) = \int\frac{\sum_{i=1}^{n} I(T_i = u, B_i = j, L_i \leq u)}{\sum_{i=1}^{n} I(L_i \leq u \leq T_i)}.$$

Chapter 6. Numerical Analysis

6.1. Simulation Results

The purpose of the simulation studies was to examine the performance of $\hat{F}_1(t)$ and $\check{F}_1(t)$ in finite samples (i.e. $n = 100$ and $n = 200$). Suppose that there are two types of failure. The indicator $I(B = 1)$ was generated from a Bernoulli random variable with probability π_1 and the latency variable for each type $T_j|B = j$ was generated from an exponential distribution such that $\Pr(T_j > t|B = j) = \exp(-\lambda_j)$ $(j = 1, 2)$. The failure time of the first event was set to be $T = I(B = 1)T_1 + I(B = 2)T_2$. The independent censoring variable C was generated from $U(0, m)$. Hence

$$\Pr(\tilde{B} = 1) = \Pr(\check{\delta} = 1) = \pi_1 \cdot \Pr(T \leq C|B = 1),$$

which measures the probability of observing the first type of failure by time t in presence of external censoring. In addition to the two proposed estimators $\hat{F}_1(t)$ and $\check{F}_1(t)$, the hypothetical estimators $\bar{F}_1(t)$ and $\bar{F}_1^*(t)$ in (19) and (20) were also evaluated.

Table 1 displays the average bias and standard deviation of each estimator, based on 500 replications, when $\pi_1 = 0.9$, $\lambda_1 = 0.8$, $\lambda_2 = 1$ and $m = 5$ which gives $\Pr(\tilde{B} = 1) \approx 0.67$. We see that the correct order of the asymptotic variances for the four estimators shown in (25) has already appeared when $n = 100$. The discrepancy gets larger when t approaches to the tail. Table 2 shows the results when $\pi_1 = 0.9$, $\lambda_1 = 0.8$, $\lambda_2 = 1$ and $m = 10$ which gives $\Pr(\tilde{B} = 1) \approx 0.79$. We see that the performance of all the estimators improves as n or $\Pr(\tilde{B} = 1)$ increases. Table 3 displays the results when $\pi_1 = 0.5$, $\lambda_1 = 0.8$, $\lambda_2 = 1$ and $m = 10$ which gives $\Pr(\tilde{B} = 1) \approx 0.44$. Although the relative performances of the four estimators are consistent with the previous two settings, the differences become less obvious when π_1 and $\Pr(\tilde{B} = 1)$ decrease.

6.2. Analysis of Stanford Heart Transplant Data

For illustration, the methods discussed in the article were applied to analyze the Stanford transplant data (Crowly and Hu, 1977). The event of interest $(B = 1)$ is rejection and the competing risk $(B = 2)$ is death without rejection. Among 65 patients, 29 with rejection $(\tilde{B} = 1)$, 12 died without rejection $(\tilde{B} = 2)$ and 24 were censored $(\tilde{B} = 0)$. For the 12 patients without rejection $(\tilde{B} = 2)$, we set $\check{X} = C$, which is the time from the date of acceptance to the end of the study in April 1974. The quantity of interest is $F_1(t)$ which represents the cumulative incidence probability of experience rejection by time t. The two estimated cumulative incidence functions, $\hat{F}_1(t)$, and $\check{F}_1(t)$, were shown in Table 4 and plotted in Figure 5. When $t > 500$, we see that the discrepancy between $\hat{F}_1(t)$ and $\check{F}_1(t)$ increases as t gets larger. It seems that $\check{F}_1(t)$ is less reliable because of heavy censoring in T_1, the time to death without rejection.

To estimate π_1, the probability of rejection, a naive estimate is $29/65 = 0.446$. If $\Pr(\tilde{B} = 0)$ is very small, this estimate may be reliable. However in the present example, $\Pr(\tilde{B} = 0) \approx 24/65 \approx 0.37$, which implies that this naive estimator is not reasonable. Applying our methods, the evidence of insufficient follow-up is clear since $\hat{S}(x_{(n)}) = \hat{S}(1350) \approx 0.209$. We found that $x^1_{\text{max}} = 1350$ and $\hat{\pi}_1 = \hat{F}_1(x^1_{\text{max}}) = 0.548$. The modified estimator in (12), which accounts for the unassigned mass in the tail, is $\breve{\pi}_1 = 0.693$. Which is a more reasonable estimator of π_1 than the naive estimator.

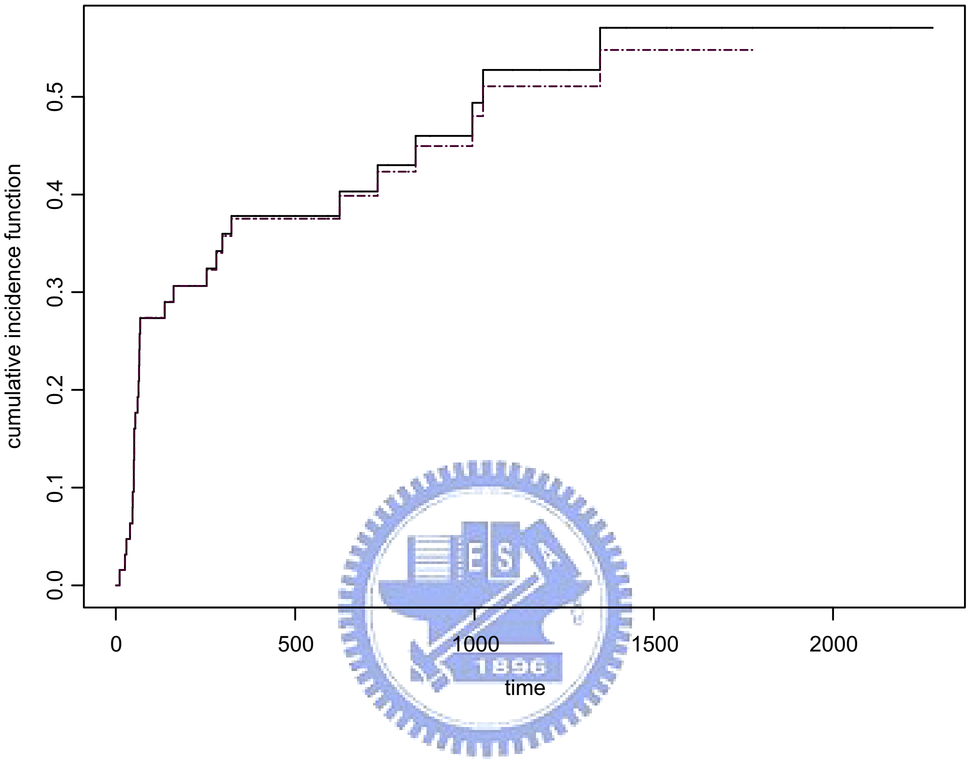

Figure 5: Two estimators of the cumulative incidence probability for the time to rejection by time t based on the Stanford Heart Transplant data. $\hat{F}_1(t)$: the dashed line; $\check{F}_1(t)$: the solid line.

Chapter 7. Discussion

We have demonstrated that, as the Kaplan-Meier estimator, the cumulative incidence function can be estimated via different approaches. We prove that these representations are equivalent under the settings discussed in the thesis. For future applications, these techniques may offer different alternatives in other contexts such as interval censoring or regression problems.

The inverse-probability-weighting representation is particularly useful in simplifying analytic work which has led to closed-form expressions of the asymptotic variances as shown in (21) and (22). It also allows us to compare the two estimators $\hat{F}_j(t)$ and $\check{F}_j(t)$ analytically. The second structure allows one to use existing statistical packages to estimate $1 - F_j(t)$ by the Kaplan-Meier estimator. However since the data ignores the information provided by other competing risks, the resulting estimator has larger variance and also it seems unlikely to estimate the long-term incidence rate if the follow-up period is not long enough. In fact, the simulations indicate that the two estimators have larger difference in the tail region. In the extension to truncated data we found that the imputation fails but the weighting method still works.

Shen (2003) showed that the Lynden-Bell estimator for the left-truncation data can be expressed as inverse-probability-weighted average. To estimate the truncation probability with left-truncated and right-censored data, Shen (2005) proposed two different estimators by using the inverse-probability-weighting method.

References

Andersen, P. K., Borgan, ∅., Gill, R. D., and Keiding, N. (1993). *Statistical Models Based on Counting Processes*, New York: Springer-Verlag.

Betensky, R. A. and Schoenfeld, D. A. (2001). Nonparametric estimation in a cure model with random cure times. *Biometrics*, **57**, 282-286.

Crowley, J. and Hu, M. (1977). Covariance analysis of heart transplant survival data, *Journal of the American Statistical Association*, **72**, 27-36.

Farley, T. M. M., Ali, M. M. and Slaymaker, E. (2001). Competing approaches to analysis of failure times with competing risks. *Statistics in Medicine*, **20**, 3601-3610.

Fine, J. P. and Gray, R. J. (1999). A proportional hazards model for the subdistribution of a competing risk, *Journal of the American Statistical Association*, **94**, 496-509.

Fine, J. P. (1999). Analyzing competing risks data with transformation models. *J. R. Statist. Soc. B*, **61**, 817-830.

Gooley, Leisenring, W., Crowley, J. and Storer, B. E. (1999). Estimation of failure probabilities in the presence of competing risks: mew representation of old estimators. *Statistics in Medicine*, **18**, 695-706.

Kalbfleisch, J. D. and Prentice, R. L. (2002). *The Statistical Analysis of Failure Time Data*, Wiley-Interscience.

Lin, D. Y. (1997). Non-parametric inference for cumulative incidence functions in competing risks studies. *Statistics in Medicine*, **16**, 901-910.

Maller, R. A. and Zhou, S. (1996). *Survival Analysis with Long-Term Survivors,* Wiley: New York.

Satten, G. A. and Datta, S. (1999). Kaplan-Meier representation of competing risk estimates. *Statistics and Probability Letters*, **42**, 299-304.

Satten, G. A. and Datta, S. (2001). The Kaplan-Meier Estimator as an Inverse-Probability-of-Censoring Weighted Average. *The American Statistician*, **55**, 207-210.

Shen, P.-S. (2003). The product-limit estimates as an Inverse-Probability-Weighted Average. *Communications in statistics, Part A-Theory and Methods*, **32**, 1119-1133.

Shen, P.-S. (2005). Estimation of the truncation probability with left-truncated and right-censored data. *Nonparametric Statistics*, **17**, 957-969.

Tsai, W.-Y. and Crowley, J. (1998). A note on nonparametric estimators of the bivariate survival function under univariate censoring. *Biometrika*, **85**, 573-580.

Wang, W. (2003). Nonparametric estimation of the sojourn time distributions for a multi-path Model. *J. R. Statist. Soc. B*, **65**, 921-936.

Table 1: Performance of $\hat{F}_1(t)$, $\check{F}_1(t)$, $\bar{F}_1(t)$, and $\bar{F}_1^*(t)$ when $\pi_1 = 0.9$ and $\Pr(\tilde{B} = 1) \approx 0.67$.

	n=100				n=200			
$F_n(t)$	$\hat{F}_1(t)$	$\check{F}_1(t)$	$\bar{F}_1(t)$	$\bar{F}_1^*(t)$	$\hat{F}_1(t)$	$\check{F}_1(t)$	$\bar{F}_1(t)$	$\bar{F}_1^*(t)$
0.1	0.339	0.340	0.340	0.342	0.272	0.272	0.273	0.275
	(-0.982)	(-0.982)	(-0.983)	(-1.051)	(-0.522)	(-0.522)	(-0.522)	(-0.587)
0.3	1.05	1.05	1.05	1.13	0.798	0.798	0.812	0.86
	(-0.984)	(-0.987)	(-0.978)	(-1.66)	(-0.521)	(-0.524)	(-0.523)	(-1.21)
0.5	1.73	1.74	1.82	1.98	1.28	1.29	1.35	1.57
	(-1.06)	(-1.07)	(-1.06)	(-3.27)	(-0.530)	(-0.538)	(-0.535)	(-2.8)
0.7	2.37	2.46	2.78	3.05	1.70	1.78	1.99	2.35
	(-1.02)	(-1.06)	(-1.05)	(-6.48)	(-0.46)	(-0.47)	(-0.49)	(-5.9)

The first number in each cell is the standard deviation ($\times 10^{-2}$), and the second number in parentheses is the average bias ($\times 10^{-3}$).

Table 2: Performance of $\hat{F}_1(t)$, $\check{F}_1(t)$, $\bar{F}_1(t)$, and $\bar{F}_1^*(t)$ when $\pi_1 = 0.9$ and $\Pr(\tilde{B} = 1) \approx 0.79$.

	n=100				n=200			
$F_n(t)$	$\hat{F}_1(t)$	$\check{F}_1(t)$	$\bar{F}_1(t)$	$\bar{F}_1^*(t)$	$\hat{F}_1(t)$	$\check{F}_1(t)$	$\bar{F}_1(t)$	$\bar{F}_1^*(t)$
0.1	0.252	0.252	0.253	0.260	0.173	0.173	0.174	0.176
	(-0.99)	(-0.99)	(-0.99)	(-0.99)	(-0.493)	(-0.493)	(-0.492)	(-0.496)
0.3	0.758	0.759	0.768	0.831	0.529	0.529	0.530	0.570
	(-0.992)	(-0.99)	(-0.99)	(-0.97)	(-0.50)	(-0.501)	(-0.496)	(-0.517)
0.5	1.22	1.23	1.28	1.47	0.82	0.824	0.868	1.05
	(-1.0)	(-0.994)	(-0.98)	(-0.94)	(-0.496)	(-0.493)	(-0.481)	(-0.528)
0.7	1.62	1.67	1.86	2.30	1.14	1.17	1.33	1.66
	(-0.994)	(-0.975)	(-0.931)	(-0.86)	(-0.492)	(-0.475)	(-0.431)	(-0.513)
0.9	2.19	2.65	3.19	3.92	1.67	2.25	2.58	3.05
	(-1.98)	(-1.81)	(-1.79)	(-1.67)	(-1.19)	(-0.98)	(-0.92)	(-1.02)

The first number in each cell is the standard deviation ($\times 10^{-2}$), and the second number in parentheses is the average bias ($\times 10^{-3}$).

Table 3: Performance of $\hat{F}_1(t)$, $\check{F}_1(t)$, $\bar{F}_1(t)$, and $\bar{F}_1^*(t)$ when $\pi_1 = 0.5$ and $\Pr(\tilde{B} = 1) \approx$ 0.44.

	n=100				n=200			
$F_n(t)$	$\hat{F}_1(t)$	$\check{F}_1(t)$	$\bar{F}_1(t)$	$\bar{F}_1^*(t)$	$\hat{F}_1(t)$	$\check{F}_1(t)$	$\bar{F}_1(t)$	$\bar{F}_1^*(t)$
0.1	0.355	0.357	0.357	0.360	0.267	0.268	0.269	0.281
	(-0.99)	(-0.99)	(-0.99)	(-1.0)	(-0.530)	(-0.529)	(-0.529)	(-0.527)
0.2	0.680	0.683	0.684	0.727	0.497	0.497	0.5	0.533
	(-0.96)	(-0.96)	(-0.96)	(-0.99)	(-0.497)	(-0.494)	(-0.495)	(-0.485)
0.3	1.122	1.130	1.137	1.220	0.771	0.791	0.798	0.865
	(-0.93)	(-0.93)	(-0.94)	(-0.98)	(-0.48)	(-0.472)	(-0.481)	(-0.454)
0.4	1.54	1.65	1.69	1.90	1.09	1.13	1.15	1.30
	(-1.01)	(-1.01)	(-1.03)	(-1.08)	(-0.416)	(-0.41)	(-0.434)	(-0.385)
0.5	3.36	3.45	3.51	3.78	2.2	2.37	2.39	2.49
	(-2.85)	(-2.79)	(-2.82)	(-2.86)	(-1.6)	(-1.54)	(-1.59)	(-1.52)

The first number in each cell is the standard deviation ($\times 10^{-2}$), and the second number in parentheses is the average bias ($\times 10^{-3}$).

Table 4: Comparison of $\hat{F}_1(t)$, $\check{F}_1(t)$, based on Stanford heart transplant data.

t	$\hat{F}_1(t)$	$\check{F}_1(t)$	t	$\hat{F}_1(t)$	$\check{F}_1(t)$
13	0.016	0.016	161	0.306	0.306
29	0.047	0.047	280	0.340	0.342
39	0.063	0.063	589	0.375	0.378
50	0.128	0.128	624	0.399	0.403
54	0.177	0.177	994	0.480	0.494
63	0.209	0.209	1106	0.511	0.528
66	0.258	0.257	1350	0.548	0.571
68	0.274	0.273			
136	0.29	0.29			