國 立 交 通 大 學

統計學研究所

碩 士 論 文

有序事件之無母數存活分析

Bivarate Survival Estimation for Successive Events

研 究 生 ：吳侑峻

指導教授 ：王維菁 博士

中 華 民 國 九 十 六 年 六 月

Bivariate Survival Estimation for Successive events

研 究 生：吳侑峻　　Student：YuChun Wu
指導教授：王維菁　　Advisor：Weijing Wang

國 立 交 通 大 學
統計研究所
碩 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
Master
in
Statistics

June 2007

Hsinchu, Taiwan, Republic of China

中華民國九十六年六月

有序事件之無母數存活分析

學生：吳侑峻　　指導教授:王維菁博士

國立交通大學統計學研究所

摘　要

在論文中我們針對連續事件的間隔時探討聯合分配函數的推論方法。文獻中對這類的問題有兩個不同的處理方式。傳統的方法以隨機過程的角度切入，並對狀態的轉移率做模式的假設。近十年來有學者嘗試用多維存活分析的技巧處理此類的問題。

我們選出兩種無母數的估計方法並透過模擬比較其差異。第一個方法由 Frydman 所提出 (1992)，在馬可夫模型下建構無母數最大概似估計量。第二個方法是由 Wang and Wells (1998) 年提出，將感興趣的二維存活函數拆解成乘積極限 (product limit) 的形式，並透過加權的方法處理相關設限的問題。後者不需要任何模型假設。

透過模擬我們驗證了無母數最大概似估計量在資料符合假設時具有較好的效度，然而假設錯誤時則會出現偏誤。第二個方法因未用到任何模型的假設，所較為穩健。

關鍵詞:馬可夫模式，相關設限，無母數最大概似估計量

Bivariate Survival Estimation for Successive events

Student：YuChun Wu　　Advisors：Dr. Weijing Wang

Institute of Statistics
National Chiao Tung University

Abstract

Consider nonparametric analysis for successive events in which the joint survival function of the duration times is of major interest. Such a phenomenon is usually investigated under the framework of stochastic processes in which the transition rates are modeled by Markov-related properties. In the past decade, some authors have applied techniques of multivariate survival analysis to handle the problem.

In the thesis, we compare two different nonparametric estimators which are constructed based on different ideas. One estimator was proposed by Frydman (1992) who considered nonparametric MLE under a Markov assumption. The other estimator was proposed by Wang and Wells (1998) who suggested to decompose the target function in terms of estimable quantities. The latter does not make any model assumption. Via simulations, we want to verify our conjecture. Briefly speaking, we suspect that the NPMLE will be more efficient if the Markov property holds but will be biased if this assumption is violated. On the other hand, the estimator proposed by Wang and Wells (1998) should be more robust since it does not require any model assumption.

Key words: decomposition, dependent censoring, Markov model.

誌謝

感謝天父的恩典，能夠完成這份論文的過程中有太多值得感恩的人與事。

首先最想感謝的是在研究上指導我最多的王維菁教授，沒有她就絕對沒這份論文的誕生。在撰寫的這一年裡，模擬分析上的技巧和研究的態度上，教授不厭其煩地引導我很多。我很感謝教授的指導，雖然辛苦但我相信在未來的研究上這些經驗能使我受用無窮，也謝謝進見學長在論文的寫作過程中也為我解答了許多的問題。

在這二年的研究生活裡，很感謝牧師、師母及大同教會的每一位弟兄姐妹的代禱，生活中因著大家能使我有堅持到底的力量。也很幸運能認識統研９４的每一位同學，有文妤、建威和怡君在，讓我在過程中能夠彼分享寫論文的點滴，並且適時地給我一些建議和幫助。自己的父母親雖然不在身邊，但是總是時時地給我所需的一切，最感謝的是他們能讓自己能夠來到這裡學習。

領受得太多，唯有獻上滿心的感謝，祝福每一位曾幫助過我的朋友和家人們。

侑峻　謹誌于

台灣　風城

中華民國九十六年七月

Chapter 1 Introduction

Nonparametric analysis of the bivariate survival function has been a popular research area. Let (T_1, T_2) be a pair of failure time variables. Several estimators of $\Pr(T_1 > s, T_2 > t)$ have been proposed including the ones proposed by Campbell & Földer (1982), Dabrowska (1988), Prentice and Cai (1992), Lin & Ying (1993) and Wang and Wells (1997), just to name a few. In general situations, where no addition information about the path is given, the nonparametric MLE approach can not produce a reliable estimator in presence of censoring. Specifically this approach can estimate the mass of a given region that maximizes the likelihood function but the distribution within the region can not be determined without additional assumptions.

Figure 1-1 Evolution of AIDS

In the thesis, we consider a special situation such that the path information is known. Consider an example of AIDS, depicted in Figure 1-1, in which T_1 represents the time from HIV infection to AIDS and T_2 represents the time from AIDS to death. Traditionally this kind of phenomenon was analyzed under the framework of stochastic processes. Specific structures based on Markov or semi-Markov properties are often imposed on the process. For example, Frydman (1992) modeled the evolution of AIDS by a Markov process and then derived the nonparametric MLE when the data are interval censored. An alternative approach suggested to decompose $\Pr(T_1 > s, T_2 > t)$ into estimable quantities and then plug in

the estimator of each component. This approach has been taken by Visser (1996), Wang and Wells (1998) and Lin, Sun and Ying (1999) based on different decompositions. In the thesis we discuss the estimator proposed by Wang and Wells (1998) who suggested to decompose the joint survival function as follows:

$$\begin{aligned}&\Pr(T_1 > s, T_2 > t) = \Pr(T_1 > s)\Pr(T_2 > t \mid T_1 > s)\\&= \Pr(T_1 > s)\cdot\prod_{u\le t}\{1-\Pr(T_2 \in [u,u+du) \mid T_2 \ge u, T_1 > s)\}. \qquad (1.1)\end{aligned}$$

Note that each component in the right-hand side can be estimated separately. In presence of censoring, the challenge of statistical inference comes from the fact that T_2 is subject to dependent censoring by T_1. That is, the larger the value of T_1, the more likely that T_2 will be censored.

The major goal of the thesis is to compare the two approaches proposed by Frydman (1992) and Wang and Wells (1998). The former is constructed based on a Markov assumption which allows for the use of nonparametric MLE. The latter is derived without making any model assumption. When the Markov property holds, it is expected that the former estimator should yield better performance. However when this assumption fails, the latter approach should still perform reasonably since it is more robust. We aim to examine our conjecture via simulations.

Here is the outline of the thesis. In Chapter 2, we will introduce some common model structures under the framework of stochastic processes. More detailed discussions can be found in the book by Anderson et al. (1993). In Chapter 3, we review the two competing approaches, namely the methods proposed by Wang & Wells (1998) and Frydman (1992). Simulation results are presented in Chapter 4. We give some concluding remarks in Chapter 5.

Chapter 2 Review on Multi-state Models

There are two ways of describing a process with multiple states. One approach adopts the framework of stochastic processes under which the transitions between the states are modeled. Specifically let $\alpha_{ij}(t)$ be the transition rate from state i to state j at time t and define the transition probability

$$P_{ij}(s,t) = \Pr(X(t) = j \mid X(s) = i)$$

which measures the conditional probability that, given that a person is in state i at time s, he/she will be in state j at time t and $X(t)$ denotes the state number at time t. Specifically we can write $\alpha_{ij}(t)$ based on $P_{ij}(s,t)$ as follows

$$\alpha_{ij}(t) = \lim_{\Delta \to 0} \frac{\Pr(X(t+\Delta) = j \mid X(t) = i)}{\Delta}.$$

The other approach takes the viewpoint of survival analysis in which the failure times between different random events are of interest. Let T be the sojourn time of state i and $S(t) = \Pr(T > t)$ be the survival function of T. The corresponding hazard function is given by

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

$$= -\frac{\partial}{\partial t} \log(S(t)).$$

Note that under the framework of survival analysis, the variable of interest is often the length of survival time instead of the transition probability.

The following sections introduce several multistate models. First, the simplest model is a model with two states and one irreversible path. Then we extend the situation that allows for j different paths but only one path will happen. Then a more

complex model with three states and two paths. The thesis will focus on the model with three states and one path.

2.1 Two-state Model (The Classical Survival Framework)

Figure 2-1 depicts a simple model with two states, namely "alive" and "dead". There is only one path and the transition from state 0 to state 1 is not reversible. Here $\alpha_{01}(t)$ denotes the be the transition rate from alive to death at time t. Note that the phenomenon in Figure 2-1 actually falls in the classical framework of survival analysis.

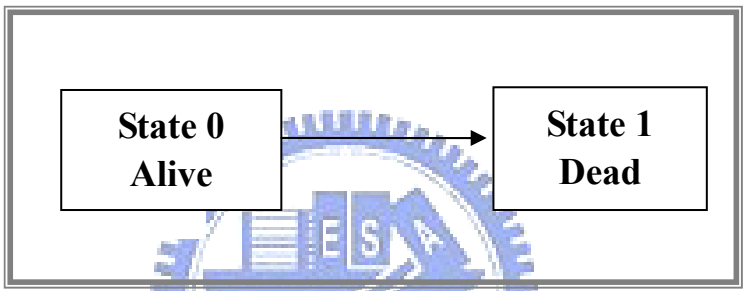


Figure 2-1 A simple two-state model

In this simplest model, it follows that $\lambda(t)=\alpha_{01}(t)$ which implies that the transition rate is the same as the hazard rate. The two approaches have no distinction.

2.2 Competing Risks Model

Figure 2-2 describes a competing risks model where a person can experience one of several different causes to the absorbing state. However a subject can only experience through one path. In this model, let the $\alpha_{0j}(t)$ be the transition rate from being alive to death due to cause j for $j=1,...,J$.

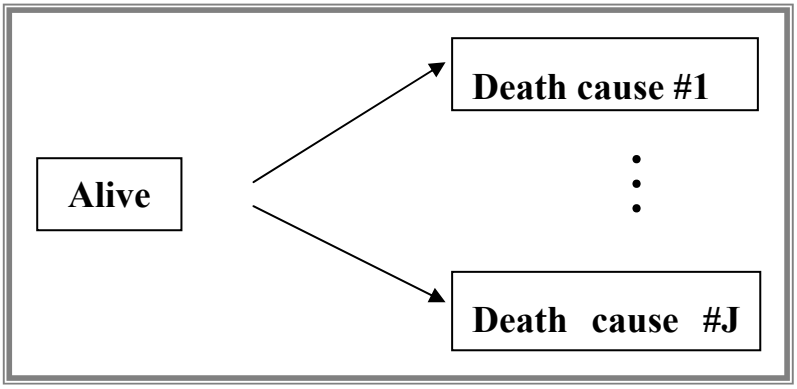


Figure 2-2 A competing risks model

In survival analysis, it follows that the hazard function from state 0 to state j can be defined as

$$\lambda_j(t) = \lim_{\Delta \to 0} \frac{\Pr(T \in [t, t+\Delta], B = j \mid T \geq t)}{\Delta}$$

where B is type of causes.

Let $\lambda(t)$ be the hazard rate escape from state 0. It follows that

$$\lambda(t) = \sum_{j=1}^{J} \lambda_j(t) = \sum_{j=1}^{J} \alpha_{0j}(t).$$

The survival function of T can be written as

$$S(t) = P(T \geq t) = \exp(-\int_0^t \sum_{j=1}^{J} \alpha_{0j}(u)du) = \exp(-\int_0^t \lambda(u)du).$$

Another useful measure in the context of competing risks is the cumulative incidence function. For cause j, the cumulative incidence function is defined as

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

$$= \int_{u=0}^{u=t} \Pr(T \geq u) \frac{\Pr(T = u, B = j)}{\Pr(T \geq u)}$$

$$= \int_{u=0}^{u=t} \Pr(T \geq u) \lambda_j(u)du$$

$$= P_{0j}(0,t).$$

Sometimes one may define the following random variable

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

Notice that T_j is an improper random variable since it carries positive mass $\Pr(B \neq j)$ at time ∞. It follows that $F_j(t) = \Pr(T_j \leq t)$ but

$$\Pr(T_j > t) \neq \exp(-\int_0^t \lambda_j(u)du).$$

The consequence of this property is that the relationship between a survival function and hazard rate no longer holds.

2.3 Three-state Model with Two Paths

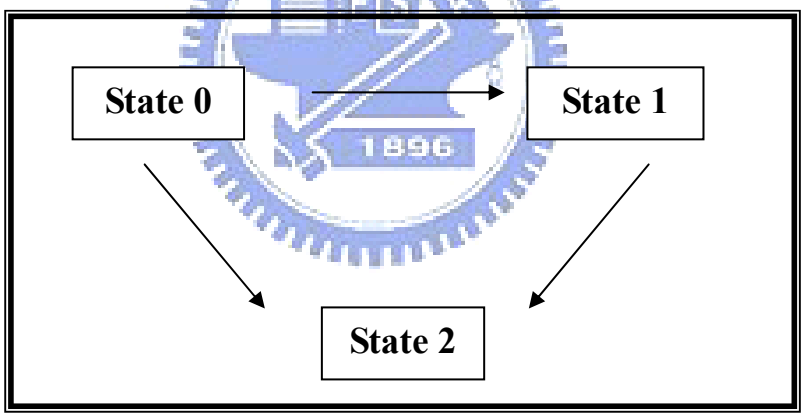


Figure 2-3 an illness-death model

The model depicted in Figure 2-3 is so called an illness-death model. An individual entering the model may take two different paths. State 1 can be viewed as an intermediate state and state 2 is an absorbing state. There are two possible paths. One path is state 0→state 1→state 2. The other is state 0→state2. Define $\lambda_{ij}(t)$ to be the hazard function from state i to state j for $(i, j) = (0,1),(0,2),(1,2)$. Define T to be the time to the first event. Using the terminology of survival analysis, the

survival function of T can be expressed as

$$\begin{aligned} \mathrm{P}(T > t) &= \exp(-\int_0^t \lambda(u)du) \\ &= \exp(-\int_0^t \lambda_{01}(u) + \lambda_{02}(u)du) \\ &= 1 - F_1(t) - F_2(t), \end{aligned}$$

where $F_j(t) = \Pr(T \le t, B = j)$. In the view of stochastic processes, recall that $P_{ij}(s,t) = \Pr(X(t) = j \mid X(s) = i)$, where $X(t)$ is the state number at time t. We have

$$P_{01}(0,t) = \Pr(X(t) = 1 \mid X(0) = 0) = \int_0^t P_{00}(0,u-)\alpha_{01}(u)P_{11}(u,t)du$$

and

$$P_{02}(0,t) = \int_0^t P_{00}(0,u-)\alpha_{02}(u)du + \int_{u \le t} [\int_{S \le u} P_{00}(0,s-)\alpha_{01}(s)ds] \cdot P_{11}(s,u-)\alpha_{12}(u)du .$$

Notice that $P_{0j}(0,t) = F_j(t)$ for $(j = 1,2)$.

Define the sojourn time T_{ij} which measures the length between state i and state j for $(i,j) = (0,1),(0,2),(1,2)$ Note that T_{ij} is undefined if a person never takes the path from state i to state j. Generally speaking, when there exists a proportion of subjects who can bypass the path, the definition of T_{ij} is not clear but $\alpha_{ij}(t)$ is well-defined.

2.4 Three-state Model with One Path

Figure 2-4 is simpler than the model discussed in Section 2.3 and is the focus of the thesis. Only the path, state 0→state 1→state 2, is possible. Let T_{01} be the

duration time between state 0 and state 1 and T_{12} be the duration from state 1 to state 2. Since only path is possible, T_{01} and T_{12} are well-defined.

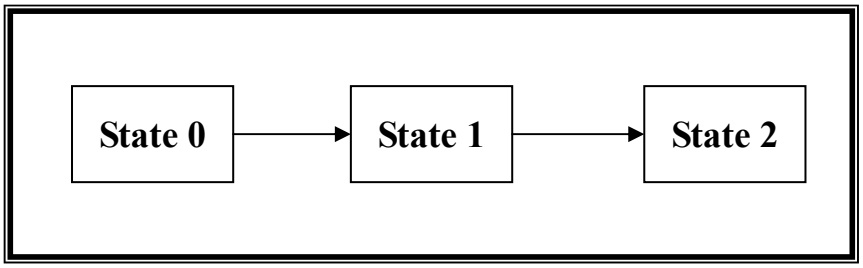


Figure 2-4 Three-state with one path

2.5 Model Properties

In the previous sections, we have seen that multi-state models can either be described under the framework of stochastic processes or in the context of survival analysis. Here we introduce some commonly seen model assumptions based on transition intensity functions.

2.5.1 Common Assumptions

One common assumption imposed on a stochastic process is the Markov assumption. Consider the continuous time Markov model (CTMC). A stochastic process $\{X(t): t \geq 0\}$ on the state space $\Phi = \{1, 2, 3, \ldots I\}$ is called a CTMC if, for all i and j in Φ, satisfy

$$\Pr(X(s+t) = j \mid X(s) = i, X(u), 0 \leq u \leq s)$$

$$= \Pr(X(s+t) = j \mid X(s) = i), \ \forall \ t, s \geq 0.$$

where s is the current time.

By the above definition, evolution in the future only depends on the present state. Recall that $P_{ij}(X(s+t) = j \mid X(s) = i)$ indicates the transition probability between state i and j. The transition matrix is

$$\mathrm{P} = \begin{bmatrix} P_{11}(s,t) & \cdots & P_{1I}(s,t) \\ \vdots & \ddots & \vdots \\ P_{I1}(s,t) & \cdots & P_{II}(s,t) \end{bmatrix}.$$

The transition rate from state i to state j at time s + t given the previous process is

$$\alpha_{ij}(s,t) = \lim_{\Delta \to 0} \frac{\Pr(X(s+t+\Delta) = j \mid X(s+t) = i, X(s) = i, X(u), 0 \le u < s)}{\Delta}$$

$$= \lim_{\Delta \to 0} \frac{\Pr(X(s+t+\Delta) = j \mid X(s+t) = i)}{\Delta}.$$

Since the future evolution only depends on present state at time s + t, then the intensity matrix becomes

$$\mathrm{R} = \begin{bmatrix} \alpha_{11}(s,t) & \cdots & \alpha_{1I}(s,t) \\ \vdots & \ddots & \vdots \\ \alpha_{I1}(s,t) & \cdots & \alpha_{II}(s,t) \end{bmatrix}.$$

In this case, longer past survival time may affect the hazard rate of the future state.

Now consider the simplest Markov process "time homogeneous Markov model". The CTMC is said to be time homogeneous if

$$P_{ij}(s,t) = P_{ij}(X(s+t) = j \mid X(s) = i) \ \forall \ t, s \ge 0$$

$$= \Pr(X(t) = j \mid X(0) = i) \ \forall \ t \ge 0.$$

Assume that $P_{ij}(s,t)$ is a function of t and $\alpha_{ij}(s,t)$ is a constant α_{ij} since

$$\alpha_{ij}(s,t) = \lim_{\Delta \to 0} \frac{\Pr(X(s+t+\Delta) = j \mid X(s+t) = i)}{\Delta}$$

$$= \lim_{\Delta \to 0} \frac{\Pr(X(\Delta) = j \mid X(0) = i)}{\Delta}$$

is a constant. Thus

$$\text{P}=\begin{bmatrix} P_{11}(t) & \cdots & P_{1I}(t) \\ \vdots & \ddots & \vdots \\ P_{I1}(t) & \cdots & P_{II}(t) \end{bmatrix} \text{ and R}= \begin{bmatrix} \alpha_{11} & \cdots & \alpha_{1I} \\ \vdots & \ddots & \vdots \\ \alpha_{I1} & \cdots & \alpha_{II} \end{bmatrix}.$$

We can ignore the information about s, the transition time of state 1, and only focus on the current time s + t in CTMC. The homogeneous Markov model means the transition rate is homogeneous for the current time. For example, if T_i is the sojourn time in state i follow $Exp(\alpha_i)$, the process $\{X(t): t \geq 0\}$ is a time homogeneous Markov process.

2.5.2 Markov Extension Model

The other well-known extension of Markov models is semi-Markov models. The transition rate of a Semi-Markov model is

$$\alpha_{ij}(s,t) = \lim_{\Delta \to 0} \frac{\Pr(X(s+t+\Delta) = j \mid X(s+t) = i, X(s) = i, X(u), 0 \leq u < s)}{\Delta}$$

$$= \lim_{\Delta \to 0} \frac{\Pr(X(t+\Delta) = j \mid X(t) = i)}{\Delta},$$

where $\alpha_{ij}(s,t)$ only depends on the duration time of state i and is independent of the previous state and the length of the past survival time. For more types of Markov Extension models, please refer to Hougaard (1999).

2.5.3 Example — Illness-death Model

Now we use the illness-death model depicted in Figure 2.3 to illustrate the Markov properties. Under the homogeneous Markov assumption, the transition rate matrix becomes

$$\text{R}= \begin{bmatrix} -\alpha_{01} - \alpha_{02} & \alpha_{01} & \alpha_{02} \\ 0 & -\alpha_{12} & \alpha_{12} \\ 0 & 0 & 0 \end{bmatrix}.$$

The lower triangular matrix is zero because it is an irreversible model. Extending to

the non-homogeneous Markov model, the transition intensity matrix becomes

$$\text{R}= \begin{bmatrix} \alpha_{00}(u) & \alpha_{01}(u) & \alpha_{02}(u) \\ 0 & \alpha_{11}(u) & \alpha_{12}(u) \\ 0 & 0 & 0 \end{bmatrix},$$

where u is the current time. Finally, the transition intensity matrix under the Semi-Markov model assumption is given by

$$\text{R}= \begin{bmatrix} \alpha_{00}(u) & \alpha_{01}(u) & \alpha_{02}(u) \\ 0 & \alpha_{11}(u-T_1) & \alpha_{12}(u-T_1) \\ 0 & 0 & 0 \end{bmatrix}$$

where T_1 is the transition time from state 0 to state 1. Bebchuk and Betensky (2001) proposed an estimator of the hazard function for such a multi-state model.

We can reduce illness-death model to a one path model with $\alpha_{02}(u)=0$ for all u. Note that this simplified case is the model that we will analyze. The transition matrix is given by:

$$\text{R}= \begin{bmatrix} \alpha_{00}(u) & \alpha_{01}(u) & 0 \\ 0 & \alpha_{11}(u-T_1) & \alpha_{12}(u-T_1) \\ 0 & 0 & 0 \end{bmatrix}.$$


E S A
1896

Chapter 3 Statistical Inference for Multi-state Model

As discussed in Chapter 2, we have seen that the process of multiple events data can either be described under the framework of stochastic processes or using the terminology of multivariate survival analysis. Now we discuss statistic inference. The difficulty of analysis usually comes from the complex structure of the events that may or may not have an orderable relationship. In the thesis, we consider a simpler case when the events are sequential. In this situation, applying Markov assumptions can simplify the analysis.

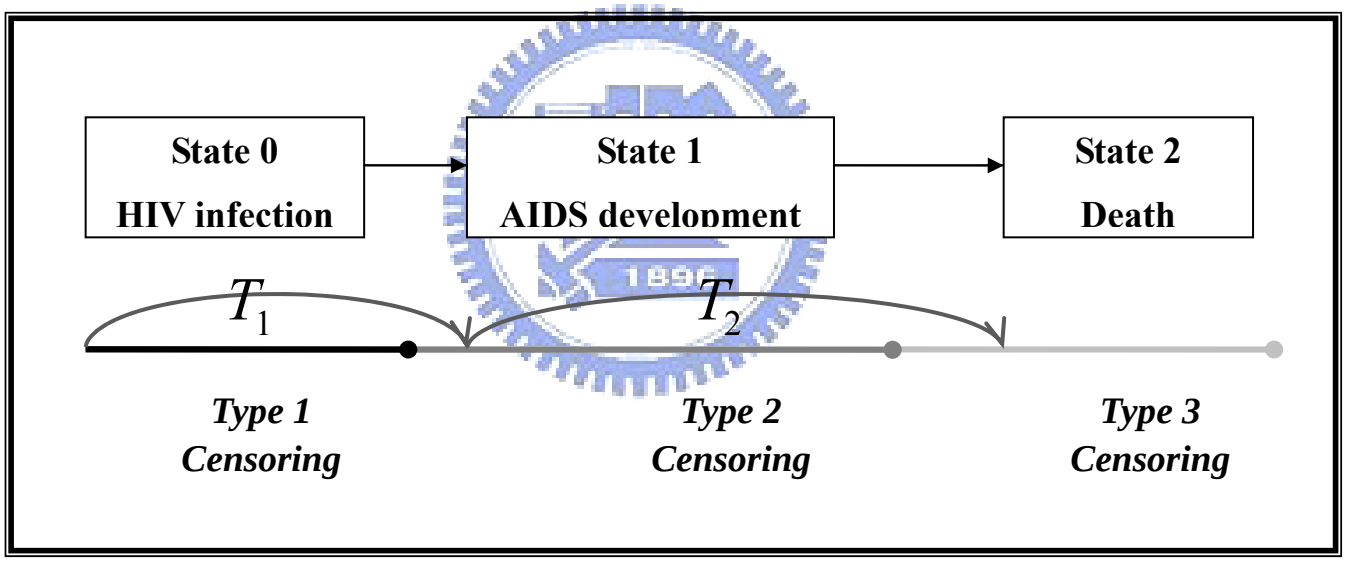


Figure 3-1 Observation of AIDS evolution subject to censoring

Using the evolution of AIDS as an example, state 0 refers to the event of HIV infection, state 1 refers to the development of AIDS and state 2 refers to death. Let T_1 be the time from HIV infection to AIDS development and T_2 be the time from AIDS development to death. Note that T_1 and T_2 may be correlated and their dependent relationship is of major interest. In the example, the intermediate event, AIDS, must proceed before death. If death occurs prior to AIDS, it is viewed as a

censoring event. In general external censoring may occur due to patients' withdrawal from the study or the end-of-effect. In Figure 3-1, we highlight three possible locations of the censoring event which would affect the resulting data structure. Let C be the censoring time measured from state 0 to the censoring event. Let $T = T_1 + T_2$ be the total survival time. If $C < T_1$, only partial information about T_1 is available and no information about T_2. In this case, we set

$$X_1 = T_1 \wedge C, \quad \delta_1 = I(T_1 \le C) = 0, \quad X_2 = 0 \text{ and } \quad \delta_2 = I(T_1 + T_2 \le C) = 0.$$

If $T_1 < C < T_1 + T_2$, we have complete information about T_1 and partial information about T_2. We set

$$X_1 = T_1, \quad \delta_1 = I(T_1 \le C) = 1, \quad X_2 = (T \wedge C) - X_1 \text{ and } \quad \delta_2 = I(T \le C) = 0.$$

If $T_1 + T_2 < C$, we have complete information about both T_1 and T_2. In this case, we set

$$X_1 = T_1, \quad \delta_1 = I(T_1 \le C) = 1, \quad X_2 = (T \wedge C) - X_1 \text{ and } \quad \delta_2 = I(T \le C) = 1.$$

We collect the data $(X_{1i}, \delta_{1i}, X_{2i}, \delta_{2i},)$ for i from 1 to N. Notice that the larger the value of T_1, the higher chance that T_2 will be censored. Since the two variables are correlated, we have dependent censoring.

3.1 Nonparametric MLE Based on Markov Assumption

Frydman (1992) used the evolution of AIDS as an example and analyzed the three-state with one-path model under the Markov assumption. Recall that T_1 represents the time from HIV infection to AIDS and T_2 represents the time from AIDS to death. Let $T = T_1 + T_2$. The Markov property is imposed on the transition probability such that

$$\Pr(T_2 = t \mid T_2 \ge t, T_1 = s) = \Pr(T = u \mid T \ge u).$$

were u = t + s.

The original paper considered the situation that T_1 is subject to interval censoring. Here we modify Frydman's discussion for right censored data. Following the notations of previous discussion, we have data ($X_{1i}, \delta_{1i}, X_{2i}, \delta_{2i}$,), (i=1,…,N). The likelihood function for right censored data can be written as:

$$\prod_{i=1}^{N} \Pr(T_1 \geq X_{1i})^{I_{1i}} \Pr(T_1 = X_{1i}, T_2 > X_{2i})^{I_{2i}} \Pr(T_1 = X_{1i}, T_2 = X_{2i})^{I_{3i}},$$

$$= \prod_{i=1}^{N} \Pr(T_1 \geq X_{1i})^{I_{1i}} \Pr(T_2 > X_{2i} \mid T_1 = X_{1i})^{I_{2i}} \Pr(T_1 = X_{1i})^{I_{2i}}$$

$$\times \Pr(T_2 = X_{2i} \mid T_1 = X_{1i})^{I_{3i}} \Pr(T_1 = X_{1i})^{I_{3i}},$$

where $I_{1i} = I(\delta_{1i} = 0, \delta_{1i} = 0)$, $I_{2i} = I(\delta_{1i} = 1, \delta_{1i} = 0)$ and $I_{3i} = I(\delta_{1i} = 1, \delta_{1i} = 1)$. Recall that $F_1(s) = \Pr(T_1 \leq s)$ and

$$\lambda_{2|1}(t \mid s) = \lim_{\Delta \to 0} \frac{\Pr(T_2 \in [t, t+\Delta) \mid T_1 = s, T_2 \geq t)}{\Delta}$$

is the transition rate which condition on $T_1 = s$ generally.

Under the Markov assumption,

$$\Pr(T_2 = u - s \mid T_2 \geq u - s, T_1 = s) = \Pr(T = u \mid T \geq u).$$

We can find the hazard rate can be simplified as a function of current time,

$$\lambda_T(u) = \lim_{\Delta \to 0} \frac{\Pr(T \in [u, u+\Delta) \mid T \geq u)}{\Delta}.$$

Hence the Markov property implies that $\lambda_{2|1}(u \mid t) = \lambda_T(u)$. Based on the product-limit decomposition and the Markov property we have

$$\Pr(T_2 > X_{2i} \mid T_1 = X_{1i}) = \prod_{(0, X_{2i}]} (1 - d\Lambda_{2|1}(u \mid X_{1i}))$$

$$= \prod_{(0, X_{2i}]} (1 - d\Lambda_T(u + X_{1i}))$$

$$= \prod_{(X_{1i}, X_{1i}+X_{2i}]} (1 - d\Lambda_T(v)),$$

where $\Lambda_{12}(t) = \int_0^t \lambda_{12}(x)dx$ and $\Lambda_T(t) = \int_0^t \lambda_T(x)dx$. Similarly, we have

$$\Pr(T_2 = X_{2i} \mid T_1 = X_{1i}) = \prod_{(0, X_{2i})} (1 - d\Lambda_{2|1}(u \mid X_{1i}))\Lambda_{2|1}(X_{2i} \mid X_{1i})$$

$$= \prod_{(0, X_{2i})} (1 - d\Lambda_{2|1}(u + X_{1i}))\Lambda_{2|1}(X_{2i} + X_{1i})$$

$$= \prod_{(X_{1i}, X_{1i}+X_{2i})} (1 - d\Lambda_T(v))\Lambda_T(X_{1i} + X_{2i}).$$

Finally, the likelihood function L can be rewritten as follows:

$$L(F_1, \Lambda_T) = \prod_{i=1}^{N} \Pr(T_1 \geq X_{1i})^{I_{1i}} \times \left\{ \prod_{(X_{1i}, X_{1i}+X_{2i}]} (1 - d\Lambda_T(u)) \right\}^{I_{2i}}$$

$$\times \Pr(T_1 = X_{1i})^{I_{2i}} \left\{ \prod_{(X_{1i}, X_{1i}+X_{2i})} (1 - d\Lambda_T(u))\Lambda_T(X_{2i} + X_{1i}) \right\}^{I_{3i}} \Pr(T_1 = X_{1i})^{I_{3i}}.$$

Now we discuss nonparametric likelihood estimation. The objective is to maximize

$$L(F_1, \Lambda_T) = \prod_{i=1}^{N} \Pr(T_1 \geq X_{1i})^{I_{1i}} \Pr(T_1 = X_{1i})^{I_{2i}} \Pr(T_1 = X_{1i})^{I_{3i}}$$

$$\times \left\{ \prod_{(X_{1i}, X_{1i}+X_{2i}]} (1 - d\Lambda_T(u)) \right\}^{I_{2i}} \left\{ \prod_{(X_{1i}, X_{1i}+X_{2i})} (1 - d\Lambda_T(u)) \right\}^{I_{3i}} \prod_{i=1}^{N} \{\Lambda_T(X_{2i} + X_{1i})\}^{I_{3i}} \qquad (3.1)$$

Let $\{X_{1(p)} \mid p = 1 \sim \tilde{M}\}$ be the collection of observed distinct values of T_1, where $\tilde{M}$ is the number of distinct failures of T_1. Let $\{X_{(n)} \mid n = 1 \sim \tilde{M}_D\}$ be the collection of observed distinct values of $T_1 + T_2$, where $\tilde{M}_D$ is the number of distinct failures of $T_1 + T_2$. Let $\{d_n\}$ be the number of ties for $X_{(n)}$. We define the $z_p = F_1(X_{1(p+1)}) - F_1(X_{1(p)}-)$ is the jump size of $F_1(t)$ at time $X_{1(p)}$ and

$h_n = \Lambda_T(X_{(n+1)}) - \Lambda_T(X_{(n)}-)$ is the jump size of $\Lambda_T(t)$ at time $X_{(n)}$. The first part $\prod_{i=1}^{N} \Pr(T_1 \geq X_{1i})^{I_{1i}} \Pr(T_1 = X_{1i})^{I_{2i}} \Pr(T_1 = X_{1i})^{I_{3i}}$ can be expressed as a function of z_p and the second part,

$$\prod_{i=1}^{N} \left\{ \prod_{(X_{1i}, X_{1i}+X_{2i}]} (1 - d\Lambda_T(u)) \right\}^{I_{2i}} \left\{ \prod_{(X_{1i}, X_{1i}+X_{2i})} (1 - d\Lambda_T(u)) \Lambda_T(X_{1i} + X_{2i}) \right\}^{I_{3i}}$$

can be expressed as a function of h_n.

Turnbull (1976) proved that the nonparametric MLE satisfies a property of self-consistency. Hence Frydman (1992) proposed a self-consistent algorithm to estimate z_p and h_n. Using the property of self-consistency, the estimation procedure can be stated as follows:

Step 0: Set initial values $Z^0 = \{z_p^0\}$ and $h^0 = \{h_n^0\}$ such that $\sum_{p=1}^{\tilde{M}} z_p^0 = 1$ and $0 < h_n^0 \leq 1$.

Step 1: Compute $\mu_{mi}(z,h)$, $\mu_{ij}^R(z,h)$ and $\gamma_{mn}(z,h)$.

$$\mu_{mi}(z,h) = I\{X_{1m} = X_{1(i)}, \delta_{1m} = 1\} \text{ for m=1~N, i=1~} \tilde{M}$$

$$\mu_{ij}^R(z,h) = \frac{I_{ij}}{\sum_{k=1}^{\tilde{M}} I_{kj}}, \text{ where } I_{ij} = \begin{cases} z_i & \text{if } X_{1(i)} \geq X_{1j} \text{ and } \delta_{1j} = 0 \\ 0 & \text{o.w.} \end{cases} \quad \text{for} \quad i=1\sim \tilde{M}$$

j=1~N

If $X_{(n)} \leq X_m$ and δ_{1m}=1, n=1~$\tilde{M}_D$, m=1~N, then

$$\gamma_{mn}(z,h) = \begin{cases} 0; \text{if } X_{(n)} \leq X_{1m} \\ \\ 1; \text{ if } X_{1m} < X_{(n)} \leq X_{1m} + X_{2m} \end{cases} \quad \text{for m=1~M, and n=1~} \tilde{M}_D$$

Step 2: Find new $\{\hat{z}_p\}$ and $\{\hat{h}_n\}$ based on $\mu_{mi}(z,h)$, $\mu_{ij}^R(z,h)$ and $\gamma_{mn}(z,h)$. Let

$$\hat{z}_p = \frac{\sum_{m=1}^{M}\mu_{mi}(z^0,h^0)+\sum_{j=1}^{J}\mu_{ij}^R(z^0,h^0)}{N} \quad \text{p= 1}\sim\tilde{\text{M}};$$

$$\hat{h}_n = \frac{d_n}{\sum_{m=1}^{M}\gamma_{mn}(z^0,h^0)} \quad \text{n=1}\sim\tilde{M}_D.$$

Step 3: Let Z^{p+1} and h^{p+1} be the new values of $\{\hat{z}_p\}$ and $\{\hat{h}_n\}$. To Repeat step1 ~ step 2 until $\{\hat{z}_p\}$ and $\{\hat{h}_n\}$ attain the required accuracy.

Note that a patient with survival time X_m with $X_{(n)} > X_m$ never enters the risk set at time $X_{(n)}$, so it is reasonable to set $\gamma_{mn}(z,h)$ equal to 0. If $\underset{m}{Max}\,\text{X}_{1m}$ is the censoring time, we should be careful that the denominator of $\mu_{ji}^R(z,h)$ becomes 0. We can directly get the $\Pr(T_1 \le s)$ and $\Pr(T_2 > t \mid T_1 = s)$ from $\{\hat{z}_p\}$ and $\{\hat{h}_n\}$ as follows:

$$\Pr(T_1 \le s) = \hat{F}(s) = \begin{cases} 0 & \\ \sum_{p=1}^{i}\hat{z}_p & \text{s}_{(i)} \le s < \text{s}_{(i+1)} \\ 1 & \end{cases}$$

$$\Pr(T_2 > t \mid T_1 = s) = \prod_{(s,s+t)}(1-d\Lambda_T) = \prod_{(s,s+t)}(1-\hat{h}_n) = \prod_{n=1}^{\tilde{M}_D}(1-\hat{h}_n I\{X_{(n)} \in (s,s+t]\}).$$

Since

$$\Pr(T_1 \le s, T_2 > t) = \int_0^s \Pr(T_2 > t \mid T_1 = u)\Pr(T_1 = u)du,$$

we can estimate $\Pr(T_1 \le s, T_2 > t)$ by

$$\hat{\Pr}(T_1 \le s, T_2 > t) = \sum_{0 \le u \le s} \Pr(T_2 > t \mid T_1 = u) \{ \sum_{i=1}^{\tilde{M}} \hat{z}_i I\{s_{(i)} \le u < s_{(i+1)}\} \}$$

$$= \sum_{0 \le s_{(p)} \le s} \prod_{n=1}^{\tilde{M}_D} (1 - \hat{h}_n I\{t_{(n)} \in (s_{(p)}, s_{(p)} + t]\}) \cdot \{ \sum_{i=1}^{\tilde{M}} \hat{z}_i I\{s_{(i)} \le s_{(p)} < s_{(i+1)}\} \} .$$

In the Appendix, we provide more detailed description about why Frydman's approach is self-consistent.

3.2 Estimators Based on Decomposition Approach

In the past decade, several nonparametric estimators of the bivariate survival function have been proposed. These estimators are constructed based on different ways of decomposing the survival function. For example Visser (1996) and Wang and Wells (1998) both decompose $\Pr(T_1 > t_1, T_2 > t_2)$ as the product of $\Pr(T_1 > t_1)$ and the other is $\Pr(T_2 > t_2 \mid T_1 > t_1)$. For example, Visser (1996) assumed that the variables are discrete and proposed to estimators for $\lambda_1(t_1)$ and $\lambda_{2|1}(t_2 \mid t_1)$. Finally, Visser maximize the likelihood function with respect to $\lambda_1(t_1)$ and $\lambda_{2|1}(t_2 \mid t_1)$.

Lin, Sun and Ying (1999) decompose the $\Pr(T_1 \le t_1, T_2 \le t_2)$ as $\Pr(T_1 \le t_1)$ - $\Pr(T_1 \le t_1, T_2 > t_2)$ and estimate the two parts separately. Another difficulty is how to estimate in presence of right censoring. Without right censoring data, $Pr(T_1 \le t_1, T_2 > t_2)$ can be estimated by empirical function, $\frac{\sum_{i=1}^{n} I(X_{i1} \le t_1, X_{i2} > t_2)}{n}$. If the presence of right censoring, they use the survival function of censoring time to adjust the empirical estimator as

$$\frac{1}{n}\sum_{i=1}^{n}\frac{I(X_{i1}\le t_1, X_{i2}> t_2)}{\hat{G}(X_{i1}+t_2)},$$

where $\hat{G}(t)$ is an estimator of $G(t)=\Pr(C>t)$.

In the thesis, we focus on Wang and Wells (1998). Wang and Wells (1998) proposed a method to estimate $\Pr(T_1>s, T_2>t)$. The idea can be described as follows. First of all, they consider the following decomposition:

$$\Pr(T_1>s,T_2>t)=\Pr(T_1>s,T_2>t)=\Pr(T_2>t\mid T_1>s)\Pr(T_1>s).$$

Then the two components involve only one-dimensional estimation, which can be handled based on the product-limit expression. Specifically, it follows that

$$\Pr(T_1>s) = \prod_{u\le s}\{1-\frac{\Pr(T_1\in(u,u+\Delta)}{\Pr(T_1\ge u)}\} = \prod_{u\le s}\{1-\lambda_{01}(u)du\},$$

where

$$\lambda_{01}(u) = \lim_{\Delta\to 0}\frac{\Pr(T_1\in[u,u+\Delta)\mid T_1\ge u)}{\Delta}.$$

Similarly

$$\Pr(T_2>t\mid T_1>s)= \prod_{u\le t}\{1-\frac{\Pr(T_2=u,T_1>s)}{\Pr(T_2\ge u,T_1>s)}\}.$$

In presence of censoring, we can get

$$\frac{\Pr(T_2=u,T_1>s)}{\Pr(T_2\ge u,T_1>s)} = \frac{\Pr(X_1>s,X_2=u,\delta_1=1,\delta_2=1)}{\Pr(X_1>s,X_2\ge u,\delta_1=1)}.$$

Estimation of the second component in presence of censoring is more difficult due to dependent censoring. They did the following derivations. For u , s > 0,

$$\Pr(X_1>s,X_2=u,\delta_1=1,\delta_2=1)$$

$$= \Pr(T_1>s,C\ge s,T_2=u,T_1\le C,C-T_1>u)$$

$$= \Pr(T_1 > s, T_2 = u, C - T_1 > u)$$

since $T_1 > s$ and $C - T_1 \geq u$ implies that $T_1 \leq C$, $C \geq s$. It follows that

$$\Pr(X_1 > s, X_2 = u, \delta_1 = 1, \delta_2 = 1) = \Pr(T_1 > s, T_2 = u)\Pr(C > u + T_1)$$

which is independence of C and T_1 、 T_2. Similarly,

$$\Pr(X_1 > s, X_2 \geq u, \delta_1 = 1)$$

$$= \Pr(T_1 > s, C > s, T_2 \geq u, C - T_1 \geq u, T_1 \leq C)$$

$$= \Pr(T_1 > s, T_2 \geq u, C \geq T_1 + u)$$

$$= \Pr(T_1 > s, T_2 \geq u)\Pr(C \geq T_1 + u)$$

Using the ratio of

$$\Pr(T_1 > s, T_2 = u) = \frac{\Pr(X_1 > s, X_2 = u, \delta_1 = 1, \delta_2 = 1)}{\Pr(C \geq T_1 + u)}$$

and

$$\Pr(T_1 > s, T_2 \geq u) = \frac{\Pr(X_1 > s, X_2 \geq u, \delta_1 = 1)}{\Pr(C \geq T_1 + u)}$$

as the basis to estimate $\lambda_{12}(u \mid T_1 > s)$, the effect of dependent censoring can be removed.

The survival function $\Pr(T_1 > s)$ is estimated easily by K-M estimator based on (X_{1i}, δ_{1i}) for i from 1 to N. Also we can use

$$\hat{\Pr}(T_1 > s, T_2 > t) = \prod_{u \leq t} \{1 - \hat{\lambda}_{12}(u \mid T_1 > s)du\}\hat{F}_1(s).$$

Now we discuss estimation of $\lambda_{12}(u \mid T_1 > s)du$ which need to account for the effect of dependent censoring. Let

$$(\mathrm{X}_\mathrm{i}^\mathrm{C}, \delta_i^C) = \begin{cases} if\ (\delta_{1i}, \delta_{2i}) = (0,0) : X_i^C = X_{1i} + 0,\ \delta_i^C = 1 - \delta_{1i}\delta_{2i} \\ if\ \ (\delta_{1i}, \delta_{2i}) = (1,0) : X_i^C = X_{1i},\quad \delta_i^C = 1 - \delta_{1i}\delta_{2i} \\ if\ \ (\delta_{1i}, \delta_{2i}) = (1,1) : X_i^C = X_{1i} + X_{2i},\ \ \delta_i^C = 1 - \delta_{1i}\delta_{2i}. \end{cases}$$

For u > 0, $\Pr(C \geq u + s)$ can be estimated by K-M method based on $(\mathrm{X}_\mathrm{i}^\mathrm{C}, \delta_i^C)$ for i from 1 to N. The corresponding estimator can be expressed as

$$\hat{G}_C(u+s) = \prod_{v<u+s} \{1 - \frac{\sum_{i=1}^{N_C} I(X_i^C = u+s, \delta_i^C = 1)}{\sum_{i=1}^{N_C} I(X_i^C > u+s)}\}.$$

Accordingly $\lambda_{12}(u \mid T_1 > s)du$ can be estimated by

$$\hat{\lambda}_{12}(u \mid T_1 > s)du = \frac{\sum_{i=1}^{N} I(X_{1i} > s, \delta_{1i} = 1, X_{2i} = u, \delta_{2i} = 1) / \hat{G}_C(X_{1i} + u)}{\sum_{i=1}^{N} I(X_{1i} > s, \delta_{1i} = 1, X_{2i} \geq u) / \hat{G}_C(X_{1i} + u)}.$$

Finally we can estimate the joint probability by

$$\hat{F}(s,t) = \prod_{u \leq t} \{1 - \hat{\lambda}_{12}(u \mid T_1 > s)du\} \hat{F}_1(s).$$

Now we compare the two estimators. They are both nonparametric. Frydman's approach uses the Markov property to simplify the likelihood expression. We expect that if the Markov property holds, the NPMLE would lead to a more efficient estimator. The approach proposed by Wang and Wells does not impose such a strong assumption and will be more robust. Our conjecture will be assessed via simulations.

Chapter 4 Simulation Study

In this chapter, we compare two different estimators via simulations. Specifically one is the path-dependent estimator proposed by Wang and Wells (1998) and the other is the non-parametric MLE proposed Frydman (1992) which assumes the Markov property. In our analysis, we generate two types of data, namely a non-homogeneous-Markov model and a Copula model. Note that under the first setting, the model assumption in the paper of Frydman (1992) is valid. Hence it is expected that this approach based on non-parametric MLE should work better than the method by Wang and Wells (1998) in which no model assumption is made. In the second setting, we generate the data from Clayton's model such that

$$\Pr(T_1 > t_1, T_2 > t_2) = \{\Pr(T_1 > t_1)^{1-\alpha} + \Pr(T_2 > t_2)^{1-\alpha} - 1\}^{\frac{-1}{\alpha-1}} \quad \alpha \in (1,\infty). \qquad (4.1)$$

4.1 Data Generation Algorithms

4.1.1 Data From a Non-homogeneous-Markov Process

Based on the non-homogeneous-Markov property, we have

$$\Pr(T \geq t \mid T_1 = s) = \exp\{-\int_0^t \Pr(T = u \mid T \geq u, T_1 = s)du\}$$

$$= \exp\{-\int_s^t \Pr(T = u \mid T \geq u)du\}$$

$$= \exp\{-\int_s^t \lambda_T(u)du\}.$$

We simplify the simulation algorithm in Judith and Betensky (2001), which is stated below.

1. Generate U follows a $uniform(0,1)$ distribution.

2. Let $U = F_1(T_1)$ and set $T_1 = F_1^{-1}(U)$, where the form of $F_1(s)$ can be derived if the form of $\lambda_1(s)$ is given based on $F_1(s) = \exp\{-\int_0^t \lambda_1(u)du\}$

3. Generate V which follows a $uniform(0,1)$ distribution, independent of U.

4. Let $V = F_T(T \mid s)$, then $T = F_T^{-1}(V \mid s)$ can be obtained by the probability integral transformation Theorem. Now we have T_1 = s and $T_2 = T - T_1 = T$-s.

5. Generate right censoring time variable C.

With (T_1, T_2, C), we can set the observed variables as $T_1 \le C$, $\delta_1 = I(T_1 \le C)$, $T_1 + T_2 \le C$ and $\delta_2 = I(T_1 + T_2 \le C)$.

λ \ α	1	3	6	10
1	0.331927	0.328921	0.329184	0.328956
3	-0.730228	-0.814889	-0.838167	-0.847920
6	-0.956323	-0.964759	-0.968716	-0.970787
10	-0.981231	-0.981207	-0.979310	-0.977902

Table 4-1 The number is each cell is Kendalls' τ

In our simulations, we generate T_1 from exponential (1) and set the hazard function $\lambda(t) = \alpha p(t)^{p-1}$ to follow the Weibull hazard function with parameter α and p. Table 4.1 lists the value of Kendalls' τ with different combination of (α, p). We observe that under the selected parameters, (T_1, T_2) can have large negative association but only permits low positive association.

4.1.2 Data from the Clayton Model

The Clayton model is given in (4.1) with $\alpha = \frac{1+\tau}{1-\tau}$. We adopt the algorithm in Prentice and Cai (1992) in which T_1 and T_2 are set to follow $Exp(\lambda_1)$ and $Exp(\lambda_2)$ respectively. The simulation procedure based on Clayton (1978) :

1. Generate U and V follow $uniform(0,1)$ distribution.

$$T_1 = -\frac{1}{\lambda_1}\log(1-U).$$

$$T_2 = \frac{1}{\alpha\lambda_2}\log(1-(1-U)^{-\alpha}+(1-U)^{-\alpha}(1-V)^{\frac{-\alpha}{\alpha+1}}).$$

2. Generate right censoring time variable C.

3. With (T_1, T_2, C), we can set the observed variables as $T_1 \le C$, $\delta_1 = I(T_1 \le C)$, $T_1 + T_2 \le C$ and $\delta_2 = I(T_1 + T_2 \le C)$.

4.2 Simulation Results

In our simulations, we report the bias and standard deviation (in the parenthesis) for the estimator $\Pr(T_1 \le t_1, T_2 \le t_2)$ based on 300 or 500 replications. Note that Wang and Wells (1998) provided their estimator based on $\Pr(T_1 > t_1, T_2 > t_2)$ and hence we need to calculate the joint probability by the following formula:

$$\Pr(T_1 \le t_1, T_2 \le t_2) = 1 - \Pr(T_1 > t_1) + \Pr(T_2 > t_2) - \Pr(T_1 > t_1, T_2 > t_2) \quad (4.2)$$

The numbers shown in the table are multiplied by 10^3. The results are given in Table 4-2 ~ Table 4-6.

4.2.1 Data from non-homogeneous Markov Process

First, we compare the performance of two estimators based on Markov data. Since the data satisfy the model assumption of Frydman's method, we expect that it will perform better than Wang and Wells' method. The bias and standard deviation of the two methods are reported corresponding to different parameters. In Table 4-2 with low correlation, we can see that both methods perform well and don't have obvious difference. In Table 4-3 and Table 4-4, the level of association becomes stronger. In the case, we can see that the bias of the two methods both are about the same, but the deviation of Wang and Wells' estimator can be larger in some case. Note that there are few observations of T_2 between 0.3 and 0.6 for generated Markov model and this affects the result of

Wang and Wells' estimation. Finally, the Table 4-4 with strong correlation structure and higher censoring rate shows the Frydman's estimator is better than Wang and Wells' estimator. The maximum bias is still about 4×10^{-3} but the bias of Wang and Wells' estimator is worse than in the low correlation structure. Even though Wang and Wells' estimator is worse, the maximum bias is about 9×10^{-3}. Note that the last grid points in Table 4-4 have unusual bias, it is caused by the bad selection of grid point with $\mathrm{P}(\mathrm{C}\geq t_1+t_2) = 0$. Since we only observe $C\wedge(T_1+T_2)$, choosing (t_1,t_2) in the tail area result in poor estimation.

4.2.2 Data from Clayton model.

Then we compare the two methods based on the Clayton model since this violates the assumption of Frydman's method, and we want to see how it behaves under model mis-specification. We report two different correlations, one is 0.005 and another is 0.818, in Table 4-5 and Table 4-6. We can find no large difference between two methods in low Kendall's τ. We increase the correlation to 0.818, and we find the bias of Frydman's method is larger at some grid points. So the wrong model assumption cause bad estimation of Frydman's approach as the correlation between the time variables is high.

T1	T2	True value	$P(C \geq t_1 + t_2)$	Bias (F)	Standard deviation (F)	Bias (W)	Standard deviation (W)
0.050000	0.130000	0.004343	0.964660	-0.284026	8.755885	-0.434063	13.549313
0.050000	0.150000	0.005498	0.961727	-0.407624	10.011397	-0.611689	14.821236
0.050000	0.250000	0.012610	0.941480	-1.262630	14.660030	-0.785667	20.901781
0.050000	0.450000	0.028900	0.881880	-2.001756	15.231161	-0.345468	24.267647
0.150000	0.150000	0.022526	0.941480	-0.661658	32.106002	-1.854896	46.602093
0.150000	0.200000	0.033448	0.934807	-1.879834	38.637600	-1.876408	54.963911
0.150000	0.250000	0.045220	0.914733	-3.144643	42.919818	-2.285402	60.973611
0.150000	0.450000	0.091090	0.861807	-4.799818	37.289896	-1.480140	64.176412
0.300000	0.650000	0.234054	0.785373	-2.211174	20.381585	-0.039404	49.780053
0.300000	0.700000	0.240786	0.772927	-1.355026	15.493690	0.342529	37.723636
0.800000	0.700000	0.529438	0.667293	-1.199453	16.874751	0.773245	38.235751
0.800000	0.750000	0.535670	0.655893	-0.483195	13.559709	0.890078	29.091310
1.200000	0.750000	0.683752	0.567680	-1.188078	13.272199	0.098142	28.755435
1.500000	0.800000	0.766423	0.490700	-0.189812	11.775102	0.421018	22.058600
2.000000	0.850000	0.857540	0.358560	0.171229	10.550403	-0.460072	16.634423
2.500000	1.000000	0.915880	0.222387	3.445247	10.441048	0.759025	11.565810

Table 4-2 Markov model α=2 λ=4 repeat 500 sample size 300 Kendall's τ=0.029227

Censoring rate of T1 is 21.87%, Censoring rate of T2 is 25.95%.

T1	T2	True value	$P(C \geq t_1 + t_2)$	Bias (F)	Standard deviation (F)	Bias (W)	Standard deviation (W)
0.050000	0.130000	0.002223	0.957753	-0.056518	5.387089	-0.301042	9.1041320
0.100000	0.500000	0.042762	0.865847	-2.771526	28.691298	-0.787734	46.704400
0.200000	0.750000	0.136575	0.788027	-3.050022	27.169118	-0.477552	73.369998
0.350000	1.000000	0.272948	0.699880	-0.281012	12.405852	0.0925330	38.961856
0.150000	0.150000	0.011802	0.928633	-0.263205	20.688308	-1.456980	31.540952
0.400000	0.350000	0.128425	0.829847	-2.226206	81.576866	-6.747808	141.35072
0.150000	0.450000	0.057598	0.865847	-3.873429	41.251242	-2.501958	65.275824
0.500000	0.750000	0.323290	0.719333	-2.965783	35.222427	-2.454568	99.687207
0.300000	0.350000	0.091797	0.854240	-3.082721	69.396949	-5.182065	112.14946
0.400000	0.450000	0.169196	0.808840	-4.147652	78.214676	-6.840380	146.43111
0.500000	0.750000	0.323290	0.719333	-2.965783	35.222427	-2.454568	99.687207
0.600000	0.800000	0.390315	0.689687	-2.645174	28.867423	-3.303129	84.626810
0.900000	0.750000	0.514485	0.628160	-2.695100	34.695753	-2.632240	97.565389
1.500000	0.800000	0.710792	0.489780	-2.755712	28.201743	-2.789572	84.507931
2.000000	0.850000	0.811094	0.362633	-2.279274	22.735904	-3.116583	70.819558
2.500000	1.000000	0.890848	0.215300	1.564353	14.766494	-0.336915	39.136745

Table 4-3 Markov model α=2 λ=2 repeat 500 sample size 300 Kendall's τ=0.146178
Censoring rate of T1 is 22.15%, Censoring rate of T2 is 28.75%.

T1	T2	True value	$P(C \geq t_1 + t_2)$	Bias (F)	Standard deviation (F)	Bias (W)	Standard deviation (W)
0.050000	0.130000	0.000730	0.939160	0.055047	2.058647	-1.797911	4.2096050
0.050000	0.150000	0.001045	0.935733	0.068707	2.701295	-1.890555	5.6914190
0.050000	0.250000	0.003905	0.903167	0.008385	6.816600	-2.331336	13.358093
0.050000	0.450000	0.016995	0.826640	-0.395545	14.823341	-1.935060	24.859973
0.150000	0.150000	0.006257	0.903167	0.120047	12.023384	-6.588907	17.569704
0.150000	0.200000	0.011192	0.890580	0.020023	18.410499	-6.719424	31.943418
0.150000	0.250000	0.017913	0.862573	-0.249468	25.211740	-6.965533	43.752489
0.150000	0.450000	0.060635	0.792413	-2.029990	41.228056	-5.308293	67.590241
0.300000	0.450000	0.143088	0.732047	-3.739897	64.850595	-9.455314	113.99729
0.300000	0.550000	0.185506	0.696940	-3.463777	49.132670	-5.827240	102.46529
0.800000	0.550000	0.469792	0.517273	-1.981527	54.713053	-6.178835	116.23843
0.800000	0.600000	0.490822	0.509773	-1.541588	43.888784	-4.712470	98.946725
1.200000	0.600000	0.638953	0.370707	0.235860	44.276640	-4.677959	97.504928
1.500000	0.650000	0.734026	0.243287	3.848780	33.938284	-2.262546	78.687287
2.000000	0.700000	0.835086	0.036567	11.287361	29.479361	-1.595592	64.155822
3.000000	0.850000	0.942616	0.000000	34.420339	17.921452	49.032459	19.881606

Table 4-4 Markov model α =3 λ =4 repeat 500 sample size 300 Kendall's τ=-0.827893
Censoring rate of T1 is 33.18%, Censoring rate of T2 is 40.11%.

T1	T2	True value	$P(C \geq t_1 + t_2)$	Bias (F)	Standard deviation (F)	Bias (W)	Standard deviation (W)
0.050000	0.020000	0.004684	0.984671	-0.076867	3.205136	-0.127983	3.990142
0.050000	0.100000	0.019333	0.966761	-0.179919	7.327877	-0.361669	8.077350
0.050000	0.150000	0.025900	0.955086	-0.131419	8.488430	-0.124885	9.327232
0.050000	0.200000	0.031003	0.944022	-0.087343	9.430975	-0.321557	10.467596
0.200000	0.020000	0.017398	0.950790	-0.140047	4.873253	-0.269865	7.811502
0.200000	0.100000	0.071819	0.932254	-0.339722	14.159066	-0.260133	16.361615
0.200000	0.150000	0.096221	0.921393	-0.416783	16.707775	-0.166718	17.727617
0.200000	0.500000	0.166722	0.843404	-0.436926	21.640716	-0.418824	22.577976
1.000000	0.150000	0.334825	0.743318	-0.862412	24.945594	-0.055581	29.639565
1.000000	0.250000	0.452326	0.720649	-1.167206	29.045440	-0.531579	31.807811
1.000000	0.350000	0.523388	0.698333	-1.192295	30.331127	-1.668054	32.919563
1.000000	0.450000	0.566364	0.675994	-1.245497	30.513365	-1.816170	32.345590
2.000000	0.150000	0.457179	0.519444	0.002288	25.398337	-0.732071	32.212080
2.000000	0.300000	0.672636	0.486174	0.232664	28.076394	-0.971929	32.102910
2.000000	0.800000	0.849028	0.377316	0.185231	24.794812	-0.079348	27.359247

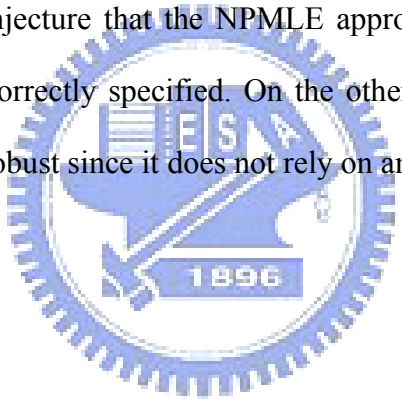
Table 4-5 Clayton model exp(1) exp(5) τ=0.005 repeat 500 sample size 300 Censoring rate of T1 is 22.29%, Censoring rate of T2 is 26.370%.

T1	T2	True value	$P(C \geq t_1 + t_2)$	Bias (F)	Standard deviation (F)	Bias (W)	Standard deviation (W)
0.050000	0.020000	0.028111	0.984663	0.033766	9.269063	-0.055722	9.907238
0.050000	0.040000	0.040643	0.980024	-0.949622	11.586493	-0.223587	12.001098
0.050000	0.080000	0.047624	0.971106	-0.948741	12.835534	0.140512	13.173680
0.050000	0.100000	0.048347	0.966859	-0.704267	13.017319	0.054889	13.246833
0.200000	0.020000	0.075735	0.950865	10.736515	14.481595	0.109143	16.573450
0.200000	0.040000	0.127881	0.946090	1.647625	18.899140	0.275777	20.735686
0.200000	0.080000	0.171711	0.937188	-8.585814	21.836477	0.104741	23.370917
0.200000	0.100000	0.177602	0.932508	-7.906743	22.325337	-0.288541	23.340817
0.500000	0.020000	0.094080	0.883341	44.986948	15.359193	0.080468	18.248861
0.500000	0.040000	0.177602	0.878678	43.228557	20.483141	-0.333964	24.063984
0.500000	0.080000	0.307634	0.870401	-3.254256	24.717086	-0.932267	29.544991
0.500000	0.100000	0.348858	0.866106	-22.264862	25.946045	-0.952399	30.634325
1.000000	0.100000	0.393023	0.753870	52.260769	24.965144	-0.621747	33.646507
1.000000	0.500000	0.632121	0.665104	-15.822411	30.558998	-1.672707	30.612189

Table 4-6 Clayton model exp(1) exp(5) τ=0.818 repeat500 sample size 300 Censoring rate of T1 is 22.29%, Censoring rate of T2 is 26.75%.

Chapter 5 Conclusion

In the thesis, we review several multi-state models and some properties that are often imposed to describe the underlying process. In particular we focus on a simple model with three states and one path that can be used to describe the evolution of AIDS. Two nonparametric estimators for estimating the joint distribution of the two duration times are discussed. Specifically Frydman (1992) proposed a nonparametric maximum likelihood estimator under the assumption that the underlying stochastic processes the Markov property. The other estimator proposed by Wang and Wells (1998) is constructed by expressing the joint survival function in terms of estimable quantities. By using the idea of weighting, the bias due to dependent censoring can be removed. Our simulations confirm our conjecture that the NPMLE approach yields a more efficient result when the model assumption is correctly specified. On the other hand, the estimator proposed by Wang and Wells (1998) is more robust since it does not rely on any model assumption.



Reference

Andersen P. K., Borgan, O., Gill, R. D., and Keding, N. (1992) *Statistical models based on counting processes*. New York: Springer-Verlag.

Andersen P. K. (1988) Multistate models in survival analysis: a study of nephropathy and mortality in diabetes. *Stat Med* **7**, 661-670.

Andersen P. K., and Keiding, N. (2002) Multi-state models for event history analysis. *Statistical Methods in Medical Research* **11**, 91-115.

Commenges, D. (1999). Multi-state Models in Epidemiology. *Lifetime Data Analysis,* **5**, 315-327.

Clayton, G. (1978) A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika* **65**, 141-51.

Campbell, G. and Földers, A. (1982) Large sample properties of nonparametric statistical inference. In *Colloquia Methemetica-Societatis, János Bolyai*, Ed. B. V. Gnedenko, M. L. Puri and I. Vincze, 103-122. Amsterdam: North Holland.

Dabrowska, D. M. (1988) Kaplan-Meier estimation on the plane. *Ann. Statist* **16**, No.4, 1475-1489.

Frydman, H. (1992). A nonparametric estimation procedure for a periodically observed three-state Markov process, with application to AIDS. *JRSSB* **54**, No.3, 853-866.

Hougaard, P. (1999) Multi-state models: a review. *Lifetime Data Analysis*, **5**, 239-264.

Hougaard, P. (2000) *Statistics for biology and health*. New York :Springer.

Bebchuk, J. D. and Betensky, R. A. (2001) Local likelihood analysis of survival data with censored intermediate events. *JASA* **96**, No.454, 449-457.

Kalbfleisch, John D. and Prentice, Ross L. (1980) *The statistical analysis of failure time data.* New York: Wiley.

Lin, D. Y., and Ying, Z. (1993) A simple nonparametric estimator of the bivariate survival function under univariate censoring. *Biometrika* **80**, No.3, 573-581.

Lin, D. Y., Sun, W., and Ying, Z. (1999) Nonparametric estimation of the gap time distribution for serial events with censored data. *Biometrika* **86**, No.1, 59-70.

Prentice, R. L. and Cai, J. (1992) Covariance and survivor function estimation using censored multivariate failure time data. *Biometrika* **79**, No.3, 495-512.

Regina, C., Elandt, Johnson., Norman, L., Johnson. (1980) *Survival models and data analysis.* New York : Wiley.

Scholz, F. W. (1980) Towards a unified definition of maximum likelihood. *The Canadian Journal of Statistics* **8**, No.2, 193-203.

Sun, J. (2006) *The statistical analysis of interval-censored failure time data.* New York : Springer.

Turnbull, B. W. (1976) The empirical distribution function with arbitrarily grouped, censored and truncated data. *JRSSB* **38**, No.3, 290-295.

Visser, M. (1996) Nonparametric estimation of the bivariate survival function with an application to vertically transmitted AIDS. *Biometrika* **83**, No.3, 507-518.

Wang, W. and Wells, M. T. (1997) Nonparametric estimators of the bivariate survival function under simplified censoring conditions. *Biometrika* **84**, No.4, 863-880.

Wang, W. and Wells, M. T. (1998) Nonparametric estimation of successive duration times under dependent censoring. *Biometrika* **85**, No.3, 561-572.

Appendix: Self-consistency of Frydman's estimator

Definition: self-consistency estimator (Turnbull 1976)

$Z = \{z_i,\ \mathrm{j}=1\sim\widetilde{\mathrm{M}}\}$ is a set of probability mass. Z is a self-consistency if Z is defined to be any solution of the equation,

$$z_j = f(Z) \quad \forall\ \mathrm{j}=1\sim\widetilde{\mathrm{M}}.$$

Under Z and $H=\{h_n,\ \mathrm{n}=1\sim\mathrm{Q}\}$ are known and

$$z_j = \Pr(X_{1(j)} \le T_1 < X_{1(j+1)}) \quad \forall\ \mathrm{j}=1\sim\widetilde{\mathrm{M}},$$

$$h_n = \Pr(T = X_{(n)} \mid T \ge X_{(n)}) \quad \forall\ \mathrm{n}=1\sim\widetilde{\mathrm{M}}_{\mathrm{D}}.$$

Let

$$\mu_{jm}(z,h) = I\{X_{1m} = X_{1(j)}, \delta_{1m} = 1\} \quad \text{for j=1}\sim\widetilde{\mathrm{M}}\text{, m=1}\sim\mathrm{N},$$

which is the probability that m-th observation fail at time $X_{1(i)}$ exactly. Let

$$\mu_{jm}^{R}(z,h) = \frac{\mathrm{I}_{jm}}{\sum_{k=1}^{N}\mathrm{I}_{km}},\ \text{where}\ \mathrm{I}_{\mathrm{jm}} = \begin{cases} z_j & \text{if } \mathrm{X}_{1(\mathrm{j})} \in (X_{1m},\infty) \text{ and } \delta_{1m}=0 \\ 0 & \text{o.w.} \end{cases} \quad \text{for j=1}\sim\widetilde{\mathrm{M}},$$

m=1~N,

which is the probability that m-th observation fail at time $X_{1(j)}$ if $T_1 > X_{1m}$. If $X_{(n)} \le X_m = X_{1m} + X_{2m}$ and δ_{1m}=1, n=1~ $\widetilde{\mathrm{M}}_{\mathrm{D}}$, m=1~N, then

$$\gamma_{mn}(z,h) = \begin{cases} 0; \text{if } \mathrm{X}_{(\mathrm{n})} \le X_{1m} \\ 1; \text{ if } \mathrm{X}_{(\mathrm{n})} \in (X_{1m}, X_m] \end{cases} \quad \text{for m=1}\sim\mathrm{N}\text{, and n=1}\sim\widetilde{\mathrm{M}}_{\mathrm{D}},$$

which is the number of risk set at time $X_{(n)}$. Let

$$d_{mn} = I\{\delta_{1m}=1,\ \delta_{2m}=1, X_m = X_{(n)}\} \quad \text{for m=1}\sim\mathrm{N}\text{, n=1}\sim\widetilde{\mathrm{M}}_{\mathrm{D}},$$

And $d_n = \sum_{n=1}^{\tilde{M}_D} d_{in}$ is number of ties at $X_{(n)}$.

Based on the above definition, we can calculate following expected value,

$$E(N_{01}(X_{1(i)}) \mid Data) = \sum_{m=1}^{N} \mu_{jm}(z,h) + \sum_{m=1}^{N} \mu_{jm}^{R}(z,h)$$

$N_{01}(X_{1(i)}) =$ # of AIDS development at time t.

$$E(N_{12}(X_{(n)}) \mid Data) = \sum_{m=1}^{N} d_{mn}$$

$N_{12}(X_{(n)})$ = # of death at time t.

$$E(Y_2(X_{(n)}^{-}) \mid Data) = \sum_{m=1}^{N} \gamma_{mn}(z,h)$$

$Y_2(X_{(n)}^{-})$ = # of risk set at time t.

According to the equations in Frydman (1992),

$$E(N_{01}(X_{1(i)}) \mid Data) = N \times z_i$$

$$E(N_{12}(X_{(n)}) \mid Data) = h_n \times E(Y_2(X_{(n)}^{-}) \mid Data).$$

This implies that

$$\hat{z}_j = \frac{\sum_{m=1}^{N} \mu_{jm}(z,h) + \sum_{m=1}^{N} \mu_{jm}^{R}(z^0,h^0)}{N} = f(Z,H)$$ satisfy the definition of self-consistency.

$$\hat{h}_n = \frac{\sum_{m=1}^{N} d_{mn}}{\sum_{m=1}^{N} \gamma_{mn}(z^0,h^0)} = f(Z,H).$$

We will prove the NPMLE of equation (3.1) can be maximized by self-consistent estimator.

First, define $A_m = \begin{cases} (X_{1m},\infty] & \text{if } \delta_{1m} = 0 \\ (X_{1m},X_{1m}] & \text{if } \delta_{1m} = 1 \end{cases} = \begin{cases} (X_{1m},\infty) & \text{otherwise} \\ (X_{1m},X_{1m}] & \text{if } \delta_{1m} = 1 \end{cases}$ $\forall$ m = 1 ~ N

$$\alpha_{mj} = \begin{cases} 1 & \text{if } X_{1(j)} \in A_m \\ 0 & \text{otherwise} \end{cases}.$$

$B_m = (X_{1m}, X_m)$ $\forall$ m = 1 ~ N

$$\beta_{in} = \begin{cases} 1 & \text{if } \delta_{1i} = 1,\ X_{(n)} \in B_i \\ 0 & \text{otherwise} \end{cases}$$

Recall that $z_j = F_1(X_{1(j+1)}) - F_1(X_{1(j)}-)$ is the jump size of $F_1(t)$ at time $X_{1(j)}$ and $h_n = \Lambda_T(X_{(n+1)}) - \Lambda_T(X_{(n)}-)$ is the jump size of $\Lambda_T(t)$ at time $X_{(n)}$.

We write down the log likelihood as following:

$$\begin{aligned}\log L &= \sum_{i=1}^{N}\{I_{1i}\log[\sum_{j=1}^{P}\alpha_{ij}z_j] + (I_{2i}+I_{3i})\log[\sum_{j=1}^{P}\alpha_{ij}z_j]\} \\ &+ \sum_{i=1}^{N} I_{2i}\{\sum_{n=1}^{\tilde{M}_D}\beta_{in}\log(1-h_n)\} + \sum_{i=1}^{N} I_{3i}\{\sum_{n=1}^{\tilde{M}_D}\beta_{in}\log(1-h_n) + \sum_{n=1}^{\tilde{M}_D} d_{in}\log h_n\} \\ &= \sum_{i=1}^{N}\{\log[\sum_{j=1}^{P}\alpha_{ij}z_j]\} + \sum_{i=1}^{N}\{\sum_{n=1}^{\tilde{M}_D}\beta_{in}\log(1-h_n)\} + \sum_{n=1}^{\tilde{M}_D} d_n \log h_n\end{aligned}$$

Let $d_j(Z) = \frac{\partial}{\partial\varepsilon}\sum_{i=1}^{N}\{\log[\sum_{k\neq j}^{\tilde{M}}\alpha_{ik}\frac{z_k}{1+\varepsilon} + \alpha_{ij}\frac{z_j+\varepsilon}{1+\varepsilon}]\} + \frac{\partial}{\partial\varepsilon}\sum_{i=1}^{N}\{\sum_{n=1}^{\tilde{M}_D}\beta_{in}\log(1-h_n)\} + \frac{\partial}{\partial\varepsilon}\sum_{n=1}^{\tilde{M}_D} d_n\log h_n$

$$= \frac{\partial}{\partial\varepsilon}\sum_{i=1}^{N}\{\log[\sum_{k\neq j}^{\tilde{M}}\alpha_{ik}\frac{z_k}{1+\varepsilon} + \alpha_{ij}\frac{z_j+\varepsilon}{1+\varepsilon}]\} + 0 = \sum_{i=1}^{N}\{\frac{\sum_{k\neq j}^{\tilde{M}}\alpha_{ik}\frac{-z_k}{(1+\varepsilon)^2} + \alpha_{ij}\frac{(1-z_j)}{(1+\varepsilon)^2}}{\sum_{k\neq j}^{\tilde{M}}\alpha_{ik}\frac{z_k}{1+\varepsilon} + \alpha_{ij}\frac{z_j+\varepsilon}{1+\varepsilon}}\}$$

Set $\varepsilon = 0$. $d_j(Z) = \sum_{i=1}^{N}\{\frac{-\sum_{k=1}^{\tilde{M}}\alpha_{ik}z_k + \alpha_{ij}}{\sum_{k=1}^{\tilde{M}}\alpha_{ik}z_k}\}$ and find the relationship between $d_j(Z)$ and $\hat{z}_j$ is

$$\hat{z}_j = \frac{1}{N}\sum_{i=1}^{N}\frac{\alpha_{ij}z_j}{\sum_{k=1}^{\tilde{M}}\alpha_{ik}z_k} = \frac{z_j}{N}\sum_{i=1}^{N}\frac{\alpha_{ij}}{\sum_{k=1}^{\tilde{M}}\alpha_{ik}z_k} = \frac{z_j}{N}\sum_{i=1}^{N}\{\frac{\alpha_{ij}}{\sum_{k=1}^{\tilde{M}}\alpha_{ik}z_k} + 1 - 1\} = \frac{z_j}{N}\{\frac{d_j(Z)}{N} + 1\}$$

So $d_j(Z) = 0$, the self-consistent estimator is equivalent to NPMLE.

We consider the NPMLE of h_n.

$$\frac{\partial}{\partial h_g}\sum_{i=1}^{N}\{\log[\sum_{k\neq j}^{\tilde{M}}\alpha_{ik}\frac{z_k}{1+\varepsilon} + \alpha_{ij}\frac{z_j+\varepsilon}{1+\varepsilon}]\}$$

$$+\frac{\partial}{\partial h_g}\sum_{i=1}^{N}\{\sum_{n=1}^{\tilde{M}_D}\beta_{in}\log(1-h_n)\}+\frac{\partial}{\partial h_g}\sum_{n=1}^{\tilde{M}_D}d_n\log h_n$$

$$=0+\frac{d_g}{h_g}+\sum_{i=1}^{N}\{\frac{-\beta_{ig}}{1-h_g}\}=0.$$

Finally, We solve above equation and find $\hat{h}_g^{MLE}=\frac{d_g}{d_g+\sum_{i=1}^{N}\beta_{ig}}$ is equivalent to

$$\hat{h}_g=\frac{\sum_{m=1}^{N}d_{mg}}{\sum_{m=1}^{N}\gamma_{mg}(z^0,h^0)}.$$