

國立交通大學

統計學研究所

碩士論文

針對半競爭風險與截切資料
估計累積發生函數



Nonparametric Estimation of Cumulative Incidence Function
Based on Truncated Semi-Competing Risk Data

研究生：楊舒雯

指導教授：王維菁 教授

中華民國九十五年六月

針對半競爭風險與截切資料

估計累積發生函數

Nonparametric Estimation of Cumulative Incidence Function
Based on Truncated Semi-Competing Risk Data

研究生：楊舒雯

Student : Shu-Wen Yang

指導教授：王維菁 教授

Advisor : Dr. Weijing Wang

國立交通大學

統計學研究所

碩士論文

A Thesis

1896

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 2006

Hsinchu, Taiwan, Republic of China

中華民國九十五年六月

針對半競爭風險與截切資料 估計累積發生函數

研究生：楊舒雯

指導教授：王維菁 教授

國立交通大學統計研究所

摘 要

本篇論文針對半競爭風險與截切資料探討累積發生函數的無母數估計方法。發現文獻中已有 Peng & Fine (2006) 討論過這個問題，並以“分解代入法”提出估計量。我們先分析競爭風險與半競爭風險資料所提供訊息之差異，並探討截切限制對估計造成的影響，最後以“加權修正偏誤”的方法提出新的無母數估計量。藉著電腦模擬實驗我們比較兩種估計量的表現，並探討造成優劣之可能原因。

關鍵字：累積發生函數、無母數分析、發病時間、半競爭風險、截切、

加權法

Nonparametric Estimation of Cumulative Incidence Function Based on Truncated Semi-Competing Risk Data

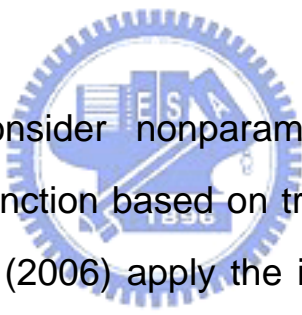
Student : Shu-Wen Yang

Advisor : Dr. Weijing Wang

Institute of Statistics

National Chiao Tung University

Abstract



In this thesis, we consider nonparametric estimation of the cumulative incidence function based on truncated semi-competing risk data. Peng & Fine (2006) apply the idea of decomposition to construct an estimator. Under a similar framework, we propose an alternative weighting approach. The two estimators are compared via simulations. The results show that our proposed estimator performs better which is possibly due to the fact that we make slightly stronger assumptions. We also investigate the difference between the information given by competing risks data and that given by semi-competing risks data.

Key words : cumulative incidence function 、 nonparametric analysis 、 onset age 、 semi-competing risks 、 truncation 、 weighting

誌謝

時光匆匆過去，經歷許多成功與挫折，歡笑與悲傷，兩年的時間讓自己茁壯許多，對未來也更能勇敢面對，而能擁有這些勇氣與能力，首先必須感謝指導教授王維菁老師，因為她的耐心指導，她的真誠對待，讓我不只在課業研究上學習許多專業知識，面對人生與生活也有更多的體悟，真的非常感謝老師不辭辛苦地指導。同時也要感謝口試委員洪慧念教授、陳珍信教授和張淑惠教授，於口試時不吝指教並提供誠懇建議。更要感謝許秋婷學姊像姊姊一樣的照顧，在生活上與課業上給予許多的教導與幫助。

在交大統研所兩年，學生老師、學長姐與同學間就像一個大家庭，大家共同扶持協助，很高興能遇見妳們，謝謝大家。

家，永遠都是家，能擁有現在的一切都要感謝我的父母家人，因為他們的辛苦工作，讓我可以開心地學習，因為他們的關心支持，讓我勇敢面對困境，因為他們無私地付出，所以我幸福地成長。我愛你們，親愛的家人。

舒雯

交通大學統計學研究所

民國九十五年六月

論文目錄

第一章	緒論	01
1.1	背景	01
1.2	Cache 資料	03
1.3	論文方向	05
第二章	文獻研讀 - 不完整資料型態之簡介	06
2.1	截切資料	06
2.2	競爭風險資料	10
2.3	半競爭風險資料	10
第三章	文獻研讀 - 累積發生函數之估計	12
3.1	競爭風險架構	12
3.2	競爭風險資料下的估計方法	12
3.3	截切後的競爭風險架構	15
3.4	截切後的半競爭風險架構	17
第四章	我們提出的估計量	21
4.1	未截切資料	21
4.2	競爭風險下的截切資料	23
4.3	半競爭風險下的截切資料	24

第五章	模擬實驗	28
5.1	資料生成方法	28
5.2	競爭風險資料之估計量比較	29
5.3	半競爭風險資料之估計量比較	29
第六章	結論	53
參考文獻與附錄		

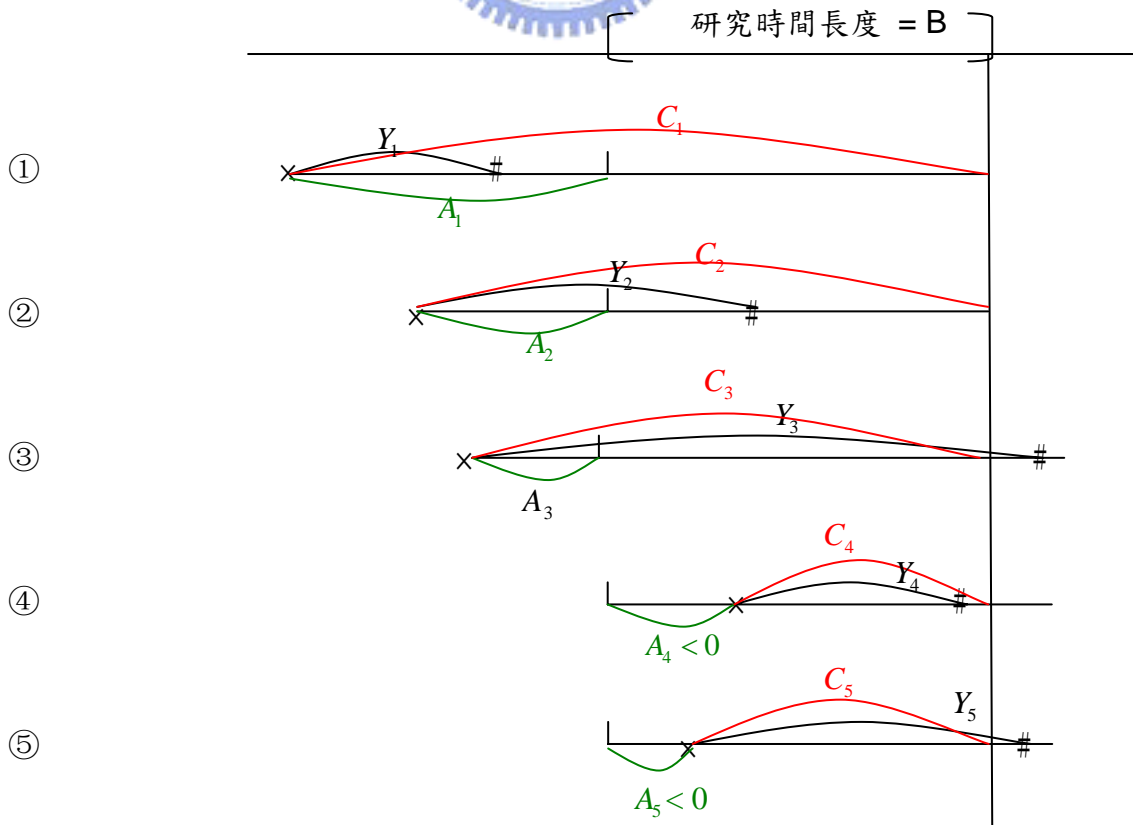


第一章 緒論

1.1 背景

我們由約翰霍普金斯大學 Dr. Peter P. Zandi 處獲得一筆調查“抗氧化劑補充的使用與 Alzheimer 症的風險之間的關係”的資料，這個研究計劃的成員來自美國數個大學與醫療中心。研究從 1995 年開始，針對美國猶他州 Cache 郡 65 歲以上的長久居民抽取樣本 5097 人，並紀錄當時是否得到 Dementia 以及補充劑的使用情形。三年後(在 1998 到 2000 之間)再對這些參加者進行第二次診斷，並紀錄此時是否罹患 Dementia 以及服用補充劑的情況。資料成員包含被抽中的居民及其一等親，包含父母和兄弟姐妹在內的所有成員。

論文最初的動機是分析這筆資料，但是因為抽樣結構十分複雜，我們先定義符號以建構資料型態。對於這個研究我們主要感興趣的變數是疾病的發病年齡，以 X 表示。令壽命為 Y ，個體進入研究的年齡為 A ，研究結束(資料分析時)的年齡 C ，令 B 表示個體進入研究期間的長度，所以 $C = A + B$ 。下圖討論哪些情形個體有可能進入樣本中，其中“ x ”代表出生，“ $\#$ ”代表死亡。



個體①：研究開始前已死亡： $Y_1 < A_1$ ，且 $Y_1 < C_1$ 。

個體②：研究開始前已出生，研究結束前已死亡： $Y_2 > A_2$ ，且 $Y_2 < C_2$ 。

個體③：研究開始前已出生，研究結束尚未死亡： $Y_3 > A_3$ ， $Y_3 > C_3$ 。

個體④：研究開始尚未出生，在研究期間出生，但研究結束前已死亡： $A_4 < 0$ ，
 $Y_4 > A_4$ ， $Y_4 < C_4$ 。

個體⑤：研究開始尚未出生，在研究期間出生，研究結束尚未死亡： $A_5 < 0$ ，
 $Y_5 > A_5$ ， $Y_5 > C_5$ 。

我們發現所蒐集的樣本可能有以下的限制：

第一種可能的限制：研究開始時存活的人($A < Y$)才有可能進入樣本中。這包含兩種情形：若是 $A > 0$ ，代表研究開始時已出生並仍存活着才有機會進入樣本，可以用" $A < Y$ "表示取樣條件，此時①④⑤會被排除在樣本之外。若是 $A < 0$ ，但是 $A + B = C > 0$ ，代表研究開始時未出生，但是研究期間出生者才有機會進入樣本，仍可用" $A < Y$ "表示取樣條件，此時只有①會被排除。

第二種可能的限制：只有 $0 < C < Y$ 的人可進入樣本中，代表只有分析時是存活狀態才能進入樣本中，此時①②④會被排除在樣本之外。但是若又加上 $A > 0$ 的條件，①②④⑤都會被排除在樣本之外。

以上兩種資料都顯示，所觀察到的樣本經過“截切”(truncation)，第一個例子的截切條件為 $A < Y$ ，第二個例子的截切條件為 $0 < C < Y$ 。而發病時間 X 又受到壽命 Y 和研究結束時年齡 C 的右設限，也就是只有當分析時已經發病的個體才有可能觀察到 X 的確切值。基於以上的截切的條件，可觀察到的變數可表示為 $(\tilde{X}, \tilde{Y}, \delta_x, \eta_y)$ ，其中 $\tilde{X} = X \wedge Y \wedge C$ ， $\tilde{Y} = Y \wedge C$ ， $\delta_x = I(X \leq (Y \wedge C))$ ， $\eta_y = I(Y \leq C)$ 。

1.2 Cache 資料

在原始資料中研究個體可以分為兩種：一為(遺傳疾病研究中家系的)淵源者 (proband)，另一個體為親屬 (relatives)。在美國猶他州 Cache 郡由 1995 年開始進行的研究中，針對 65 歲以上的居民進行抽樣，這些被抽中者即為淵源者；而淵源者的父母和兄弟姐妹就是親屬。

先敘述我們取得資料的原始碼，原資料中，” alive” 代表 still living，” lastage” 代表 current age or age of death，” dementia” 為 prevalent dementia status 的指標，” onsetage” 則是 age of dementia。當我們把問題量化後，就把 Cache 資料轉為所定義之符號。以下是變數的定義：

(1) 令 η_y 代表 “是否仍存活” 的指標函數，若 alive = 1，則 $\eta_y = 0$ ；反之，

當 alive = 0，則 $\eta_y = 1$ 。

(2) 令 $\tilde{Y} = \text{lastage}$ 。

(3) 令 δ_x 代表 “是否得 dementia” 的指標函數，因此若 dementia = 0，則 $\delta_x = 0$ ；相反，當 dementia = 1，則 $\delta_x = 1$ 。

(4) \tilde{X} ：若 $\delta_x = 0$ ，則 $\tilde{X} = \text{lastage}$ ；反之，當 $\delta_x = 1$ ， $\tilde{X} = \text{onsetage}$ 。

資料中並沒有紀錄個別個體進入研究的年齡紀錄，因為整個實驗追蹤時間為三年，我們假設所有的個體都是在分析時的前三年即進入研究，並令 $B = 3$ ， $A = C - 3$ 。以下是變數轉換的實例，我們挑出兩個個體詳述其變數的登錄。

原始碼：

id	Sex	Alive	Last age	dementia	onset age
1	2	1	90	1	89
3	2	1	82	0	.

轉化新符號的登錄方式：

\tilde{X}	δ_x	\tilde{Y}	η_y	A	B	C
89	1	90	0	87	3	90
82	0	82	0	79	3	82

對上述兩個個體做簡單說明：個體 “id = 1”，性別為女性，分析時是存活狀態，當時年齡 90 歲，在 89 歲得到 dementia。可以推得 $\tilde{X} = \text{onsetage} = 89$ ， $\delta_x = 1$ ， $\tilde{Y} = C = 90$ ， $\eta_y = 0$ ；假設研究一開始就進入，則 $A = 90 - 3 = 87$ 。個體 “id = 3”，性別為女性，分析時是存活狀態，當時年齡 82 歲，但是尚未得到 dementia。可以推得 $\tilde{X} = \text{lastage} = 82$ ， $\delta_x = 0$ ， $\tilde{Y} = C = 82$ ， $\eta_y = 0$ ；若假設研究一開始就進入，則 $A = 82 - 3 = 79$ 。

在我們所獲得的資料中，淵源者並不包含“在分析前已死亡者”。換言之所有的淵源者都必須符合 $0 < C < Y$ 的截切條件。這個限制使得對淵源者來說

$\tilde{Y} = \text{lastage} = C$ ，而且實際觀察到的 $\Pr(\eta_y = 1)$ 為 0。前述 $(\tilde{X}, \delta_x, \tilde{Y}, \eta_y, C)$ 的資料結構，在 $0 < C < Y$ 的條件下，只能觀察到 (\tilde{X}, C, δ_x) 。在 5092 個淵源者中，

$\sum_{i=1}^{5092} I(\delta_{xi} = 1) = 355$ ，換言之，只有 7% 左右的觀測值已觀察到 dementia 的發生。

這個比例如此小，部份原因可能來自於研究時間太短以致設限情形嚴重。

將研究問題量化後我們發現要直接分析這筆資料有相當的困難，因為所牽涉到的統計推論問題，相當具有挑戰性且可以參考的現有文獻十分有限。我們感興趣的變數為發病時間 X ，但是其受到截切和設限的條件又都與死亡時間 Y 有關，而同一人身上的發病與死亡又是相互關聯的變數。這麼複雜的取樣方式實難以套用現有文獻分析之。然而這個挑戰也提供我們新的研究方向，我們希望自行發展出合適的推論方法以分析類似結構的資料。

1.3 論文方向

我們決定把論文方向由資料分析調整為方法論的探討。希望針對 $(\tilde{X}, \delta_x, \tilde{Y}, \eta_y, C)$ 的資料型態提出可行的推論方法，藉此可以看出 Y 的資訊在其中所扮演的角色，如果可能，未來請資料提供者再分享更完整的資料。

第二章的內容主要為文獻回顧，內容包含“競爭風險資料”與“半競爭風險資料”與“截切資料”的介紹。在第三章中我們把具體的研究目標鎖定為 $F_1(t) = \Pr(X \leq t, X < Y)$ 的估計，這個函數描述了母體發病年齡的分配函數。我們回顧了現有文獻的推論方法。早期的研究主要針對右設限資料，Chen, Chang and Wang (2006) 的論文中探討三種推論技巧（分解法 decomposition、加權法 weighting 以及補差法 imputation），並證明這三種不同方法可得相同的答案。我們在第三章中並摘錄了 Peng & Fine (2006) 的結果，他們亦是針對半競爭風險與截切資料探討累積發生函數的估計。我們在網路上獲得 Peng & Fine 文章的檔案，並發現這篇文章已被 *Biometrika* 期刊接受並會在近期發表。雖然我們感興趣的問題已經有學者提出答案，我們在第四章中仍然以另一種角度建構新的估計量。在 Chen, Chang and Wang (2006) 的架構下，Peng & Fine (2006) 的做法比較像是利用分解法 (decomposition)，而我們的方法則是用加權 (weighting) 的概念。在第五章我們透過模擬實驗比較不同估計量的表現。第六章為結論與未來值得探討的研究方向。

第二章文獻研讀—不完整資料型態簡介

在本章中，我們回顧幾種常見的不完整資料結構（截切、競爭風險、半競爭風險）。我們所欲探討的問題，則是結合數種架構更為複雜的資料型態。

2.1 截切資料

令 Y 為感興趣事件的發生時間，當 Y 的值唯有介於 (A_L, A_R) 中才有可能被觀察到時，我們稱其受到了截切（truncation）。左截切是當 $A_R = \infty$ 的情形，右截切則是當 $A_L = 0$ 的情形。截切資料和一般設限（censoring）資料之不同處在於被設限的變數仍包含部份的資訊，然而被截切的資料卻無法由所觀測到的變數中獲取任何資訊。欲對 Y 的分配做進一步推論，通常需要假設 Y 與截切變數間彼此獨立，這是避免所謂“不可辨識性”（un-identifiability）的常用假設。

以下針對最常見的左截切資料（ $A_L = A$ ， $A_R = \infty$ ），討論存活函數 $\Pr(Y > t)$ 的估計方法。在不失一般性下先排除設限情形，可觀察的資料包含 $\{(Y_i, A_i) (i = 1, \dots, n)\}$ ，且需滿足截切條件 $Y_i > A_i$ 。存活函數可以表示為風險函數的連乘積：

$$\Pr(Y > t) = \prod_{u \leq t} \left\{ 1 - \frac{\Pr(Y \in [u, u + du])}{\Pr(Y \geq u)} \right\}, \quad (2.1)$$

其中風險函數 $\lambda_2(u)$ 滿足

$$\lambda_2(u)du = \frac{\Pr(Y \in [u, u + du])}{\Pr(Y \geq u)}. \quad (2.2)$$

當資料為完整時，風險函數機率 $\lambda_2(u)du$ 的估計量為 $\sum_{i=1}^n I(Y_i = u) / \sum_{i=1}^n I(Y_i \geq u)$ 。然

而若是資料為左截切， $\sum_{i=1}^n I(Y_i = u) / \sum_{i=1}^n I(Y_i \geq u)$ 所估計的非 $\lambda_2(u)du$ ，而是

$$\frac{\Pr(Y \in [u, u + du), Y > A)}{\Pr(Y \geq u, Y > A)}. \quad (2.3)$$

圖 2.1 提供了更詳細的說明。在未截切下，時間 u 的風險集合為 $\{Y \geq u\}$ 。圖 2.1 中橫虛線標示的部份為風險集合中因截切而無法觀察到的區域，橫實線所標示為

截切後可觀察到的區域即 $\{Y \geq u, Y > A\}$ 。在此截切下的風險集合中，實際在 u 點發生感興趣的事件的觀測值滿足 $\{Y = u, Y > A\}$ 。以上分析顯示若未調整風險集合的定義，會發現 Y 值越大，越容易出現在集合 $\{Y \geq u, Y > A\}$ 中。即使 A 與 Y 彼此獨立，以 $\sum_{i=1}^n I(Y_i = u) / \sum_{i=1}^n I(Y_i \geq u)$ 估計 $\lambda_2(u)du$ 仍有低估的問題，因為分子和分母被截切的比例不同。

針對左截切資料，Lynden-Bell (1971) 修正了風險集合的定義，以去除截切所造成的偏誤。修正的方向為針對風險集合做更進一步的截切，也就是把風險集合修正為 $\{Y \geq u, A < u\}$ 。在圖 2.1 中，修正後的風險集合為橫直線方格所標示的區域。此時在 A 與 Y 彼此獨立的情形下，因為 $\{Y = u, A < u\}$ 與 $\{Y \geq u, A < u\}$ 兩個集合受截切的比例是相同的，所以 $\sum_{i=1}^n I(Y_i = u, A_i < u) / \sum_{i=1}^n I(Y_i \geq u, A_i < u)$ 能夠合理的估計 $\lambda_2(u)du$ 。此時 $\Pr(Y > t)$ 的估計量可表示為

$$\bar{S}_Y(t) = \prod_{u \leq t} \left\{ 1 - \frac{\sum_{i=1}^n I(Y_i = u, A_i < u)}{\sum_{i=1}^n I(Y_i \geq u, A_i < u)} \right\} \quad (2.4)$$

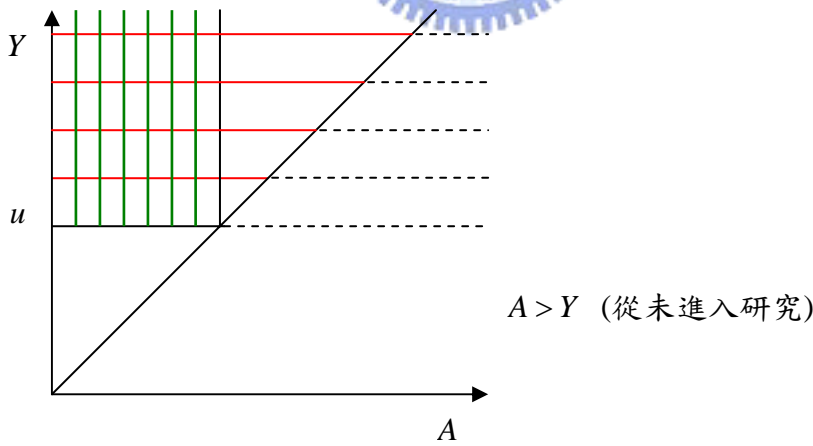


圖 2.1：截切資料觀測範圍

以上利用“進一步截切”的技巧雖巧妙的修正偏誤，卻也顯現另一個問題：當 $u < \min\{A_i, i = 1, \dots, n\}$ 時，我們無法估計 $\lambda_2(u)du$ 。圖 2.2 顯示在 $u < \min\{A_i, i = 1, \dots, n\}$ 時，沒有任何觀測值會落在所對應的風險集合中(以斜線表

示)。這個問題對存活函數的估計所產生的影響可由以下的分解得見。令

$a \geq \min\{A_i, i = 1, \dots, n\}$ ，當 $t > a$ 時

$$\begin{aligned} \Pr(Y > t) &= \Pr(Y > a) \Pr(Y > t | Y > a) \\ &= \prod_{u \leq t} \left[1 - \frac{\Pr(Y \in [u, u + du))}{\Pr(Y \geq u)} \right] \\ &= \prod_{u \leq a} \left[1 - \frac{\Pr(Y \in [u, u + du))}{\Pr(Y \geq u)} \right] \prod_{a \leq u \leq t} \left[1 - \frac{\Pr(Y \in [u, u + du))}{\Pr(Y \geq u)} \right]. \end{aligned}$$

為避免估計 $\Pr(Y > a)$ 可能會出現前述之問題，退而求其次可得

$S_Y^a(t) = \Pr(Y > t | Y \geq a)$ 的估計量為

$$\bar{S}_Y^a(t) = \prod_{a \leq u \leq t} \left\{ 1 - \frac{\sum_{i=1}^n I(Y_i = u, A_i < u)}{\sum_{i=1}^n I(Y_i \geq u, A_i < u)} \right\}. \quad (2.5)$$

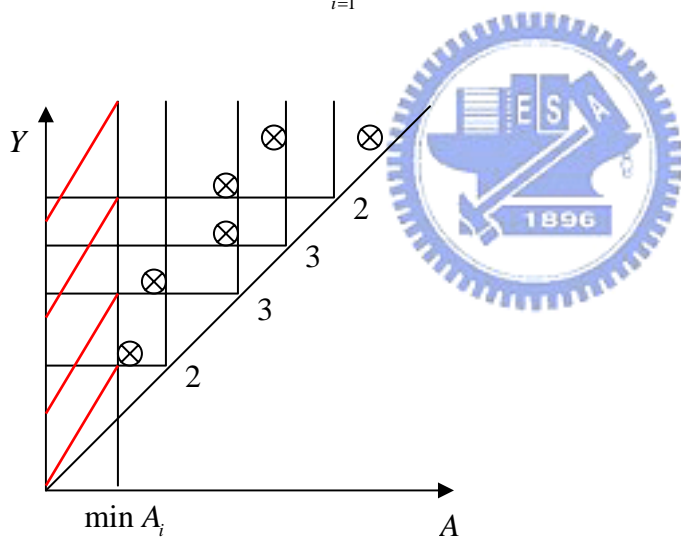


圖 2.2：截切資料風險集合

在考慮右設限下，假設設限變數為 C ，其隨機樣本為 (C_1, \dots, C_n) 。可觀察的資料為 $\{(\tilde{Y}_i, A_i, \delta_{yi}) (i = 1, \dots, n)\}$ ，其中 $\tilde{Y}_i = Y_i \wedge C_i$ ， $\delta_{yi} = I(Y_i \leq C_i)$ 。此樣本受制於截切條件 $\tilde{Y}_i > A_i$ ，也就是當 $Y_i > A_i$ 且 $C_i > A_i$ ，才可能被觀察到。當 C_i 與 Y_i 相互獨立時， $\sum_{i=1}^n I(\tilde{Y}_i = u, \delta_{yi} = 1, A_i < u) / n$ 可以合理的估計

$$\Pr(Y \in [u, u + du)) \Pr(C > u, A < u) ,$$

$\sum_{i=1}^n I(\tilde{Y}_i \geq u, A_i < u) / n$ 可以合理的估計

$$\Pr(Y \in [u, u + du)) \Pr(C > u, A < u) .$$

也因此當 $a \geq \min\{A_i, i = 1, \dots, n\}$, $S_Y^a(t)$ 的估計量由(2.5)修正為

$$\hat{S}_Y^a(t) = \prod_{a \leq u \leq t} \left\{ 1 - \frac{\sum_{i=1}^n I(\tilde{Y}_i = u, \delta_{yi} = 1, A_i < u)}{\sum_{i=1}^n I(\tilde{Y}_i \geq u, A_i < u)} \right\} . \quad (2.6)$$

Klein and Moeschberger 一書 (2002) 整理了估計量 $\hat{S}_Y^a(t)$ 的理論性質。當

Y 與截切變數 A 間彼此獨立時, Y 的風險機率可表示為

$$\frac{\Pr(Y = y | A = a, Y > A)}{\Pr(Y \geq y | A = a, Y > A)} = \frac{\Pr(Y = y)}{\Pr(Y \geq y)} .$$

此時 $\hat{S}_Y^a(t)$ 為 $S_Y(t)$ 的 nonparametric MLE。雖然 $\hat{S}_Y^a(t)$ 具有理想的大樣本性質, 在有限樣本時靠近邊界時間的估計卻會出現不穩定的現象。問題出在風險集合個

數的計算。沒有經人工截切的風險集合個數為 $R(u) = \sum_{i=1}^n I(\tilde{Y}_i \geq u)$, 當 u 遞增時

$R(u)$ 會遞減, 因此估計的不穩定出現在大的時間點。經人工截切的風險集合個

數為 $\bar{R}(u) = \sum_{i=1}^n I(\tilde{Y}_i \geq u, A_i \leq u)$, 當 u 遞增時, $R(u)$ 會遞減, 但 $\bar{R}(u)$ 卻不是呈現

單調變化, 並且在 u 很小或是很大時 $\bar{R}(u) \rightarrow 0$, 此時估計量 $\hat{S}_Y^a(t)$ 在頭尾兩端階

會有不穩定的問題。Lai & Ying (1991) 建議做以下的修正:

$$\tilde{S}_Y(t) = \prod_{u \leq t} \left\{ 1 - \frac{I(\tilde{Y}_i = u, \delta_{yi} = 1, A_i < u)}{I(\tilde{Y}_i \geq u, A_i < u)} I \left[\sum_{i=1}^n I(\tilde{Y}_i \geq u, A_i < u) \geq cn^\alpha \right] \right\} ,$$

其中 $c > 0$, $0 < \alpha < 1$ 。修正後的估計量排除了分母太小的可能性, 在大樣本時與 $\bar{S}_Y(t)$ 有相同的近似分配。 $S_Y^a(t) = \Pr(Y > t | Y \geq a)$ 的估計亦可做相同的調整。

2.2 競爭風險資料

許多醫學研究, 個體往往歷經多次或是多種事件。在我們考慮的例子中研究

個體可能經歷“發病”，或是“死亡”兩種事件。令 X 代表發病時間， Y 代表死亡時間，可知 Y 是 X 的競爭風險，但 X 卻非 Y 的競爭風險。以發病的例子而言，發病時間 X 的定義是模糊的，因為部份研究個體可能至死都不會發病。這些終生未發病的個體，我們通常假設 $X > Y$ ，但是避免定義 X 的值。

針對多重事件的資料，在右設限下若只記錄第一個事件的資訊(包含“事件發生的時間”與“事件的種類”)，文獻稱之為競爭風險資料 (competing risks data)。令設限變數為 C ，競爭風險資料所記錄到的變數為 $(\tilde{X}, \delta_x, \delta_y)$ ，其中 $\tilde{X} = X \wedge Y \wedge C$ 、 $\delta_x = I(X \leq Y \wedge C)$ 、 $\delta_y = I(Y \leq X \wedge C)$ 。當 $\delta_x = 1$ 時， $\tilde{X} = X$ ；當 $\delta_y = 1$ 時， $\tilde{X} = Y$ ；當 $\delta_x = 0$ 且 $\delta_y = 0$ 時， $\tilde{X} = C$ 。

以下兩個函數經常被用於描述競爭風險資料：第一個是“特定原因之風險函數” (the cause-specific hazard function)：

$$\lambda_1(t) = \lim_{dt \rightarrow 0} \frac{\Pr(X \in [t, t+dt] | X \geq t, Y \geq t)}{dt},$$

與“累積發生函數” (the cumulative incidence function)：

$$F_1(t) = \Pr(X \leq t, X < Y)。$$

之後的論文方向會以 $F_1(t)$ 的估計為重點，以我們的例子來說 $F_1(t)$ 代表母體中到 t 歲前會發生症的比例，其中 $X < Y$ 生前發病的條件必然成立。在第三章中我們會回顧以無母數方法估計 $F_1(t) = \Pr(X \leq t, X < Y)$ 的相關文獻。

2.3 半競爭風險資料

在我們感興趣的例子中，“死亡”並非發病的競爭風險，換言之兩個事件的競爭關係並非雙向的。此時研究者可記錄的事件不只限於“第一個”發生的事件，還能包含死亡這類具有“終結” (terminal) 特質的事件。當所記錄的變數包含 $(\tilde{X}, \tilde{Y}, \delta_x, \eta_y)$ ，其中 $\tilde{Y} = Y \wedge C$ ， $\eta_y = I(Y \leq C)$ ，Fine et al.(2001)稱之為半競爭風險資料 (semi-competing risks data)。

這種資料型態的特色是只能觀察到 $\{(X, Y): 0 \leq X < Y < \infty\}$ 的情況下，文獻中探討半競爭風險資料的論文包含探討 X 之雙樣本檢定問題。Lin, Robins and Wei, 1996; Chang, 2000。探討 (X, Y) 的關聯性的論文則包含 Day et al. (1997)、Fine et al. (2001) 以及 Wang (2003)。由於競爭風險資料只紀錄第一個發生的事件，半競爭風險資料比競爭風險資料多了關於終結事件的資訊，我們希望瞭解這樣的訊息是否對累積發生函數 $F_1(t)$ 的估計更有幫助。在第三章將回顧 Peng and Fine (2006) 的文章，他們發現在截切的條件下，半競爭風險資料確實比競爭風險資料提供更多訊息，對估計 $F_1(t)$ 也更加精準。



第三章 文獻研讀—累積發生函數之估計

3.1 競爭風險架構

在論文中我們假設 X 代表發病時間， Y 代表死亡時間。在 2.2 小節中我們以競爭風險架構 (competing risks framework) 描述發病和死亡的關係，並提及論文的目標會鎖定於“累積發生函數” $F_1(t) = \Pr(X \leq t, X < Y)$ 的估計。在本章中，我們針對不同資料型態，回顧文獻所提出估計 $F_1(t)$ 的方法。在右設限下，令設限變數為 C 。競爭風險資料所記錄的變數為

$$\tilde{X} = X \wedge Y \wedge C, \delta_x = I(X \leq Y \wedge C), \delta_c = I(C \leq X \wedge Y), \delta_y = 1 - \delta_x - \delta_c.$$

時間變數 \tilde{X} 代表第一個發生事件的發生時間， $(\delta_x, \delta_y, \delta_c)$ 為第一個事件發生型態 (cause of failure) 的指標， $\delta_x = 1$ 代表最先發病、 $\delta_y = 1$ 代表死亡時還未發病且未被設限、 $\delta_c = 1$ 代表還未觀察到發病或是死亡就被設限了。競爭風險架構下所記錄的隨機樣本可以表示為 $\{(\tilde{X}_i, \delta_{xi}, \delta_{yi})(i = 1, \dots, n)\}$ 。



3.2 競爭風險資料下的估計方法

當 $C = \infty$ 時，資料可以簡化為 $\{X_i \wedge Y_i, \Delta_i = I(X_i \leq Y_i) (i = 1, \dots, n)\}$ 。此時 $F_1(t)$ 可用動差法估計 $\sum_{i=1}^n I(X_i \wedge Y_i \leq t, \Delta_i = 1) / n$ 。然而當外來設限發生時，無法得知部份個體的 Δ_i 值。對於 $F_1(t)$ 的估計，現有文獻多數考慮右設限的資料 $\{(\tilde{X}_i, \delta_{xi}, \delta_{yi})(i = 1, \dots, n)\}$ 。Chen, Chang and Wang (2006) 曾把文獻常用的方法區分為三類，並證明這三個方法會得到一致的答案。我們在此摘要他們的結果。

文獻的第一個方向是先將 $F_1(t)$ 分解，個別估計後再代回。可將 $F_1(t)$ 表示為

$$F_1(t) = \Pr(X \leq t, X < Y) = \int_0^t S_{X \wedge Y}(u-) \lambda_1(u) du, \quad (3.1)$$

其中 $S_{X \wedge Y}(u) = \Pr(X > u, Y > u) = \Pr(X \wedge Y > u)$ 。可以採用 Kaplan-Meier 方法估計 $S_{X \wedge Y}(t)$ ，公式如下

$$\hat{S}_{X \wedge Y}(t) = \prod_{u \leq t} \left\{ 1 - \frac{\sum_{i=1}^n I(\tilde{X}_i = u, \delta_{xi} + \delta_{yi} = 1)}{\sum_{i=1}^n I(\tilde{X}_i \geq u)} \right\}. \quad (3.2)$$

特定風險函數 $\lambda_1(u)du$ 的估計量為 $\sum_{i=1}^n I(\tilde{X}_i = u, \delta_{xi} = 1) / \sum_{i=1}^n I(\tilde{X}_i \geq u)$ 。利用“代入” (plug-in) 的原則， $F_1(t)$ 的估計量可以表示為

$$\sum_{u \leq t} \left\{ \prod_{k < u} \left[1 - \frac{\sum_{i=1}^n I(\tilde{X}_i = k, \delta_{xi} + \delta_{yi} = 1)}{\sum_{i=1}^n I(\tilde{X}_i \geq k)} \right] \frac{\sum_{i=1}^n I(\tilde{X}_i = u, \delta_{xi} = 1)}{\sum_{i=1}^n I(\tilde{X}_i \geq u)} \right\}. \quad (3.3)$$

第二個方向是利用所謂“加權” (weighting) 的概念以修正因設限所造成的偏誤。令感興趣的事件為 $I(X = u, X < Y)$ ，但設限下只能觀察到

$I(\tilde{X} = u, \delta_x = 1)$ 。然而以 $I(\tilde{X} = u, \delta_x = 1)$ 做為 $I(X = u, X < Y)$ 的代理者

(proxy) 是有偏誤的。在假設 C 與 (X, Y) 獨立的情形下，可推得

$$E[I(\tilde{X} = u, \delta_x = 1) | X, Y] = E[I(X = u, Y \geq u, C \geq u) | X, Y] = I(X = u, Y \geq u) S_C(u^-),$$

其中 $S_C(t) = \Pr(C > t)$ ，且 $S_C(t^-) = \Pr(C \geq t)$ 。我們可以把 $S_C(u^-)$ 視為偏誤的比例，透過加權的調整可發現

$$\begin{aligned} E \left[\frac{I(\tilde{X} = u, \delta_x = 1)}{S_C(u^-)} \right] &= E_{X,Y} \left[E_C \left[\frac{I(\tilde{X} = u, \delta_x = 1)}{S_C(u^-)} \mid X, Y \right] \right] \\ &= E_{X,Y} \left[I(X = u, Y \geq u) \frac{S_C(u^-)}{S_C(u^-)} \mid X, Y \right] \\ &= \Pr(X = u, Y \geq u). \end{aligned}$$

以上推導顯示 $\frac{I(\tilde{X} = u, \delta_x = 1)}{S_C(u^-)}$ 與 $I(X = u, X < Y)$ 有相同的一階動差，均為

$\Pr(X = u, Y > u)$ ，前者仍可以利用動差法估計之。因為 $S_C(t)$ 未知，可以用

Kaplan-Meier 方法估計如下

$$\hat{S}_C(t) = \prod_{u \leq t} \left\{ 1 - \frac{\sum_{i=1}^n I(\tilde{X}_i = u, \delta_{xi} = 0, \delta_{yi} = 0)}{\sum_{i=1}^n I(\tilde{X}_i \geq u)} \right\}. \quad (3.4)$$

以加權法所得 $F_1(t)$ 的估計量為

$$\int_{u \leq t} \sum_{i=1}^n \frac{I(\tilde{X}_i = u, \delta_{xi} = 1)}{n \hat{S}_C(u-)} = \frac{1}{n} \sum_{i=1}^n \frac{I(\tilde{X}_i \leq t, \delta_{xi} = 1)}{\hat{S}_C(\tilde{X}_i-)}。 \quad (3.5)$$

第三個方向則是以“補差” (imputation) 的方式來填補因為設限所造成的訊息缺失。在給定觀察值 $(\tilde{X}_i, \delta_{xi}, \delta_{yi})$ 下，感興趣的變數 $I(X_i = u, Y_i > u)$ 的條件期望值可表示為

$$\begin{aligned} & E[I(X_i = u, Y_i > u) | \tilde{X}_i, \delta_{xi}, \delta_{yi}] \\ &= I(\tilde{X}_i = u, \delta_{xi} = 1) \\ &+ I(\tilde{X}_i < u, \delta_{xi} = 0, \delta_{yi} = 0) \times \Pr(X_i = u, Y_i > u | \tilde{X}_i, \tilde{X}_i < u, \delta_{xi} = 0, \delta_{yi} = 0), \end{aligned}$$

其中

$$\Pr(X_i = u, Y_i > u | \tilde{X}_i, \tilde{X}_i < u, \delta_{xi} = 0, \delta_{yi} = 0) = \frac{\Pr(X_i = u, Y_i > u)}{S_{X \wedge Y}(\tilde{X}_i)}。$$

透過期望值的操作可發現

$$\begin{aligned} \Pr(X_i = u, Y_i > u) &= E\left[E\left[I(X_i = u, Y_i > u) | \tilde{X}_i, \delta_{xi}, \delta_{yi}\right]\right] \\ &= E\left[I(\tilde{X}_i = u, \delta_{xi} = 1)\right] \\ &+ E\left[I(\tilde{X}_i < u, \delta_{xi} = 0, \delta_{yi} = 0) \frac{\Pr(X_i = u, Y_i > u)}{S_{X \wedge Y}(\tilde{X}_i)}\right]。 \end{aligned}$$

以上等式左右兩邊都牽涉到 $\Pr(X_i = u, Y_i > u)$ ，形成所謂“自我一致” (self-consistent) 方程式，以此建構估計量需要用進一步用數值方法求解。經過一番整理可得以下更易操作的關係式：

$$\Pr(X_i = u, Y_i > u) = \frac{E\left[I(\tilde{X}_i = u, \delta_{xi} = 1)\right]}{1 - E\left[I(\tilde{X}_i < u, \delta_{xi} = 0, \delta_{yi} = 0) / S_{X \wedge Y}(\tilde{X}_i)\right]}。$$

代入 (3.2) $S_{X \wedge Y}(t)$ 的估計量後，以補差法求得之 $F_1(t)$ 估計量如下：

$$\int_{u \leq t} \frac{\sum_{i=1}^n I(\tilde{X}_i = u, \delta_{xi} = 1)}{n - \sum_{i=1}^n I(\tilde{X}_i < u, \delta_{xi} = 0, \delta_{yi} = 0) / \hat{S}_{X,Y}(\tilde{X}_i)} \quad (3.6)$$

Chen, Chang and Wang (2006) 證明 (3.3) (3.5) (3.6) 的公式相等，也就是三種方法可以得到一致的答案。

3.3 截切後的競爭風險架構

在本節中，我們所討論的競爭風險資料為截切後的樣本。這個問題始出現在 Peng & Fine (2006) 的論文中。我們依自己的架構陳述問題，摘錄 Peng & Fine 已有的結果，後續會提出新的方法做為比較。

為了簡化問題，我們先暫時不考慮設限的影響，也就是令 $C = \infty$ 。令 X 為發病年齡， Y 為死亡年齡， A 為進入研究的年齡。未截切時可觀察到 $(X \wedge Y, \Delta)$ ，若截切條件為 $X \wedge Y \geq A$ ，代表進入樣本時個體必須是存活且尚未發病的狀態。資料可表示為

$$\{X_i \wedge Y_i, \Delta_i = I(X_i \leq Y_i), A_i\} \quad (i = 1, \dots, n),$$

但是受限於 $X_i \wedge Y_i \geq A_i$ 。當 $(X, Y) \perp A$ 時，可得

$$\begin{aligned} \lambda_1(t)dt &= \Pr(X \wedge Y \in [t, t+dt), \Delta = 1 \mid X \wedge Y \geq t) \\ &= \Pr(X \wedge Y \in [t, t+dt), \Delta = 1 \mid X \wedge Y \geq t, X \wedge Y \geq A, A < t) ; \end{aligned}$$

$$\begin{aligned} S_{X \wedge Y}(t) &= \Pr(X \wedge Y > t) = \prod_{u \leq t} \left\{ 1 - \frac{\Pr(X \wedge Y = u)}{\Pr(X \wedge Y \geq u)} \right\} \\ &= \prod_{u \leq t} \left\{ 1 - \frac{\Pr(X \wedge Y = u \mid X \wedge Y \geq A, A < u)}{\Pr(X \wedge Y \geq u \mid X \wedge Y \geq A, A < u)} \right\} . \end{aligned}$$

因此 $\lambda_1(t)dt$ 與 $S_{X \wedge Y}(t)$ 的估計量分別調整為

$$\frac{\sum_{i=1}^n I(X_i \wedge Y_i = t, \Delta_i = 1, A_i < t)}{\sum_{i=1}^n I(X_i \wedge Y_i \geq t, A_i < t)},$$

$$\prod_{u \leq t} \left\{ 1 - \frac{\sum_{i=1}^n I(X_i \wedge Y_i = u)}{\sum_{i=1}^n I(X_i \wedge Y_i \geq u)} \right\}.$$

再考慮設限的影響，截切的條件可表示為 $X \wedge Y \wedge C = \tilde{X} \geq A$ 。也就是個體必須是存活、尚未發病、且突破潛在的設限因素影響才得以進入樣本。所觀察到的資料可表示為

$$\{(\tilde{X}_i, \delta_{xi}, \delta_{yi}, A_i)(i = 1, \dots, n)\},$$

其中 $\tilde{X}_i \geq A_i$ 。風險函數 $\lambda_1(t)dt$ 的估計量可以表示成

$$\frac{\sum_{i=1}^n I(\tilde{X}_i = t, \delta_{xi} = i, A_i \leq t)}{\sum_{i=1}^n I(\tilde{X}_i \geq t, A_i \leq t)};$$

$S_{X \wedge Y}(t)$ 則可以下式 Kaplan-Meier 的方法估計：

$$\hat{S}_{X \wedge Y}^T(t) = \prod_{u \leq t} \left[1 - \frac{\sum_{i=1}^n I(\tilde{X}_i = u, \delta_{xi} + \delta_{yi} = 1, A_i \leq u)}{\sum_{i=1}^n I(\tilde{X}_i \geq u, A_i \leq u)} \right]. \quad (3.7)$$

將以上兩估計量分別代入(3.1)式，可以得到 $F_1(t)$ 的估計量為

$$\hat{F}_1^{PN}(t) = \sum_{u \leq t} \left\{ \prod_{s < u} \left[1 - \frac{\sum_{i=1}^n I(\tilde{X}_i = s, \delta_{xi} + \delta_{yi} = 1, A_i \leq s)}{\sum_{i=1}^n I(\tilde{X}_i \geq s, A_i \leq s)} \right] \left[\frac{\sum_{i=1}^n I(\tilde{X}_i = u, \delta_{xi} = 1, A_i \leq u)}{\sum_{i=1}^n I(\tilde{X}_i \geq u, A_i \leq u)} \right] \right\} \quad (3.8)$$

Peng & Fine (2006) 文章將 (3.8) 的估計量稱為 “naïve estimator”，他們認為此估計方法雖然是合理的，但是卻未能利用到更多的資訊。“進一步的截切” (further truncation) 的技巧是用來修正截切造成的偏誤，但這個方法會捨棄一些可用的資訊。在 $\lambda_1(t)dt$ 與 $S_{X \wedge Y}(t)$ 的估計中經過進一步的截切後所考慮的風險集合包含滿足 $X > u, Y > u, A < u, Y > A$ 的觀測值，而將資料中未符合 $X > u, Y > u, A > u, Y > A$ 者捨棄。假設我們要探討在年齡 $u = 70$ 的發病風險，直覺上樣本中在 70 歲時尚未發病和仍存活的人應該進入風險集合裏，但是透

過“人為截切” (artificially truncation) 的機制，只會用到那些“在 70 歲以前進入樣本者”的資訊 (註：如果有個人在 72 歲進入樣本時尚未發病且存活，他的資訊不會被使用)。在下一節中 Peng & Fine (2006) 利用半競爭風險資料中存活變數 Y 的資訊以估計 $F_1(t)$ ，過程中捨棄人為截切的技巧，因此可避免所資料流失的缺點。

3.4 截切後的半競爭風險架構

Peng & Fine (2006) 仍利用分解的方法得到以下的等式

$$\begin{aligned} F_1(t) &= \Pr(X \leq t, X < Y) \\ &= \int \int_{u \leq t, v > u} \Pr(X \wedge Y = u, Y = v) dv du, \end{aligned} \quad (3.9)$$

其中 $\Pr(X \wedge Y = x, Y = y) = E[I(X \wedge Y = x, Y = y)]$ 。他們發現 $F_1(t)$ 的估計可以利用代入 $\Pr(X \wedge Y = x, Y = y)$ 的無母數估計量，而後者的估計所需要的是半競爭風險的資料。考慮設限下的半競爭風險資料為

$$\{(\tilde{X}_i, \tilde{Y}_i, \delta_{xi}, \eta_{yi}, A_i, C_i)(i=1, \dots, n)\},$$

其中 $\tilde{X}_i = X_i \wedge Y_i \wedge C_i$ 、 $\tilde{Y}_i = Y_i \wedge C_i$ 、 $\delta_{xi} = I(X_i \leq Y_i \wedge C_i)$ 、 $\eta_{yi} = I(Y_i \leq C_i)$ ，截切的條件為 $\tilde{Y}_i \geq A_i$ 。

第一目標為估計 $\Pr(X \wedge Y = x, Y = y)$ 。在截切下可得

$$\begin{aligned} \Pr(\tilde{X} = x, \tilde{Y} = y, \eta_y = 1 | \tilde{Y} \geq A) &= \Pr(X \wedge Y = x, Y = y, Y \leq C | \tilde{Y} \geq A) \\ &= \frac{\Pr(X \wedge Y = x, Y = y) \Pr(A \leq y \leq C)}{\Pr(\tilde{Y} \geq A)}. \end{aligned}$$

調整以上等式可得

$$\Pr(X \wedge Y = x, Y = y) = \frac{\Pr(\tilde{X} = x, \tilde{Y} = y, \eta_y = 1 | \tilde{Y} \geq A) \Pr(\tilde{Y} \geq A)}{\Pr(A \leq y \leq C)}. \quad (3.10)$$

下一目標是估計上式等號右邊的每個成員。欲估計 $\Pr(A \leq y \leq C)$ 可以先探討以下

關係:

$$\begin{aligned}\Pr(A \leq y \leq \tilde{Y} | \tilde{Y} \geq A) &= \frac{\Pr(A \leq y \leq C, A \leq y \leq Y)}{\Pr(\tilde{Y} \geq A)} \\ &= \frac{\Pr(A \leq y \leq C, y \leq Y)}{\Pr(\tilde{Y} \geq A)} \\ &= \frac{\Pr(A \leq y \leq C)}{\Pr(\tilde{Y} \geq A)} \Pr(Y \geq y),\end{aligned}\quad (3.11)$$

其中

$$\frac{\Pr(A \leq y \leq C)}{\Pr(\tilde{Y} \geq A)} = \frac{\Pr(A \leq y \leq \tilde{Y} | \tilde{Y} \geq A)}{\Pr(Y \geq y)} = \frac{\Pr(A \leq y \leq \tilde{Y} | \tilde{Y} \geq A)}{S_Y(y-)}.\quad (3.12)$$

依據(3.11)(3.12)的推導，(3.10)可以進一步表示為

$$\begin{aligned}\Pr(X \wedge Y = x, Y = y) &= \frac{\Pr(\tilde{X} = x, \tilde{Y} = y, \eta_y = 1 | \tilde{Y} \geq A) \Pr(\tilde{Y} \geq A)}{\Pr(A \leq y \leq C)} \\ &= \Pr(\tilde{X} = x, \tilde{Y} = y, \eta_y = 1 | \tilde{Y} \geq A) \frac{S_Y(y-)}{\Pr(A \leq y \leq \tilde{Y} | \tilde{Y} \geq A)}.\end{aligned}$$

上式右邊的每個成員皆可套用現成方法估計之，分述如下。可以用 Lynden-Bell 方法估計 $S_Y(t) = \Pr(Y > t)$:

$$\tilde{S}_Y^T(y) = \prod_{\tilde{Y}_i \leq y} \left\{ 1 - \frac{\sum_{j=1}^n I(\tilde{Y}_j = \tilde{Y}_i, \eta_{y_i} = 1)}{\sum_{j=1}^n I(A_j \leq \tilde{Y}_i \leq \tilde{Y}_j)} \right\};\quad (3.13)$$

$\Pr(A \leq y \leq \tilde{Y} | A \leq \tilde{Y})$ 的估計量為

$$C_n(y) = \frac{1}{n} \sum_{i=1}^n I(A_i \leq y \leq \tilde{Y}_i); \quad (3.14)$$

$\Pr(\tilde{X} = x, \tilde{Y} = y, \eta_y = 1 | \tilde{Y} \geq A)$ 的估計量為

$$F_n^*(dx, dy) = \frac{1}{n} \sum_{i=1}^n I(\tilde{X}_i = x, \tilde{Y}_i = y, \eta_{y_i} = 1). \quad (3.15)$$

綜合以上(3.13)-(3.15)的推導依代入的原則可得 $\Pr(X \leq t, X < Y)$ 的估計量

$$\int_0^t \int_x^\infty \frac{\tilde{S}_Y^T(y)}{C_n(y)} F_n^*(dx, dy) = \frac{1}{n} \sum_{i=1}^n \frac{\tilde{S}_Y^T(\tilde{Y}_i^-)}{C_n(\tilde{Y}_i)} I(\tilde{X}_i \leq t, \tilde{X}_i < \tilde{Y}_i, \eta_{y_i} = 1). \quad (3.16)$$

估計式 (3.16) 中牽涉積分到無窮大的運算，也就是對 $\Pr(X \wedge Y = x, Y = y)$

的估計必須要在 y 很大時亦為合理。然而在當研究時間有限時，往往會發生較晚的發病時間且長壽的病人會受到設限。當變數 C 的上界比變數 $X \wedge Y$ 來得短時，因為資訊不足使得估計 $\Pr(X \wedge Y = x, Y = y)$ 的尾端有實質上的困難。對此問題 Peng & Fine (2006) 提出修正的方法。他們先選取一常數 τ 滿足

$$t \leq \tau \leq \sup\{t : \Pr(\tilde{X} > t) > 0\},$$

並將 $\Pr(X \leq t, X < Y)$ 分成 $\Pr(X \leq t, X < Y \leq \tau)$ 和 $\Pr(X \leq t, Y > \tau)$ 兩部份。第一部份 $\Pr(X \leq t, X < Y \leq \tau)$ 內層積分範圍只到 τ ，可用以下估計量

$$\frac{1}{n} \sum_{i=1}^n \frac{\tilde{S}_Y^T(\tilde{Y}_i^-)}{C_n(\tilde{Y}_i)} I(\tilde{X}_i \leq t, \tilde{X}_i < \tilde{Y}_i \leq \tau, \eta_{yi} = 1)。$$

第二部份可以繼續拆解為

$$\Pr(X \leq t, Y > \tau) = \Pr(Y > \tau) - \Pr(X > t, Y > \tau)。 \quad (3.17)$$

他們利用以下關係

$$\Pr(X > x | Y > y) = \frac{\Pr(\tilde{X} > x, A \leq y < \tilde{Y} | A \leq \tilde{Y})}{\Pr(A \leq y < \tilde{Y} | A \leq \tilde{Y})},$$

推導出

$$\Pr(X > x, Y > y) = S_Y(y) \frac{\Pr(X > x, A \leq y < \tilde{Y} | A \leq \tilde{Y})}{\Pr(A \leq y < \tilde{Y} | A \leq \tilde{Y})}。 \quad (3.18)$$

將 (3.18) 代入 (3.17) 中，可得

$$\begin{aligned} \Pr(X \leq t, Y > \tau) &= \Pr(Y > \tau) - \Pr(X > t, Y > \tau) \\ &= S_Y(\tau) \left\{ 1 - \frac{\Pr(X > t, A \leq \tau < \tilde{Y} | A \leq \tilde{Y})}{\Pr(A \leq \tau < \tilde{Y} | A \leq \tilde{Y})} \right\}, \end{aligned} \quad (3.19)$$

其對應的估計量為

$$\tilde{S}_Y^T(\tau) \left\{ 1 - \frac{C_{2,n}(t, \tau^+)}{C_n(\tau^+)} \right\},$$

其中 $C_{2,n}(x, y) = \frac{1}{n} \sum_{i=1}^n I(\tilde{X}_i > x, A_i \leq y \leq \tilde{Y}_i)$ 。綜合以上結果，他們建議將

$\Pr(X \leq t, X < Y)$ 的估計量調整為

$$\tilde{F}_1^{PF}(t) = \frac{1}{n} \sum_{i=1}^n \frac{\tilde{S}_Y^T(\tilde{Y}_i^-)}{C_n(\tilde{Y}_i)} I(\tilde{X}_i \leq t, \tilde{X}_i < \tilde{Y}_i \leq \tau, \eta_{y_i} = 1) + \tilde{S}_Y^T(\tau) \left\{ 1 - \frac{C_{2,n}(t, \tau^+)}{C_n(\tau^+)} \right\}. \quad (3.20)$$

以上依據半競爭風險所架構的估計量 $\tilde{F}_1^{PF}(t)$ 與(3.8)式依據競爭風險資料所架構的 naïve 估計量 $\hat{F}_1^{PF}(t)$ 相比較，會發現 $\hat{F}_1^{PF}(t)$ 利用了人為截切，但是 $\tilde{F}_1^{PF}(t)$ 卻不需要這個條件。換言之， $\tilde{F}_1^{PF}(t)$ 可以更充分的善用每一筆觀察值。此外，根據模擬試驗的結果， $\tilde{F}_1^{PF}(t)$ 估計的比 $\hat{F}_1^{PF}(t)$ 更為準確。



第四章 我們提出的估計量

在第三章中我們回顧了以無母數方法估計累積發生函數的現有文獻，其中比較多的研究針對競爭風險資料，Peng and Fine (2006) 則是針對截切下的半競爭風險資料，也是本論文的主要研究方向。我們希望把 Chen, Chang and Wang (2006) 論文探討的推論技巧（“分解法”、“加權法”與“補差法”），應用到受制於截切的半競爭風險資料結構，所提出的方法將與 Peng and Fine (2006) 提出的估計量做比較。

4.1 未截切資料

在 Peng & Fine (2006) 的論文中強調半競爭風險資料含有較多有用資訊。我們第一個想探討的是此多餘資訊是如何對估計產生幫助？如前所述累積發生函數做以下的分解：

$$F_1(t) = \Pr(X \leq t, X < Y) = \int_{u \leq t} \Pr(X = u, Y > u) = \int_{u \leq t} \Pr(X \wedge Y = u, \Delta = 1),$$

其中 $\Delta = I(X < Y)$ 。由此拆解可知 $F_1(t)$ 可以表示為 $(X \wedge Y, \Delta)$ 的積分函數，這是未設限下的競爭風險資料型態。另一種分解方法為

$$\begin{aligned} F_1(t) &= \Pr(X \leq t, X < Y) \\ &= \int_{u \leq t} \int_{v > u} \Pr(X = u, Y = v) dv du \\ &= \int_{u \leq t} \int_{v > u} \Pr(X \wedge Y = u, Y = v) dv du \circ \end{aligned}$$

以上的表示法顯示 $F_1(t)$ 亦可以表示為 $(X \wedge Y, Y)$ 的積分函數，這是未設限下半競爭風險資料可觀察到的變數。

先不考慮設限的情形，令競爭風險的樣本為 $\{(X_i \wedge Y_i, \Delta_i)(i = 1, \dots, n)\}$ ；半競爭風險的樣本為 $\{(X_i \wedge Y_i, Y_i)(i = 1, \dots, n)\}$ 。可得

$$\bar{F}_1(t) = \sum_{i=1}^n I(X_i \leq t, X_i < Y_i) / n = \sum_{i=1}^n I(X_i \wedge Y_i \leq t, \Delta_i = 1) / n .$$

因為

$$\int_{v>u} I(X_i \wedge Y_i = u, Y_i = v) = I(X_i \wedge Y_i = u, Y_i > u) = I(X_i \wedge Y_i = u, \Delta_i = 1) , \quad (4.1)$$

可發現

$$\bar{F}_1(t) = \int \int \sum_{u<t, v>u} I(X_i \wedge Y_i = u, Y_i = v) / n .$$

也就是以估計 $F_1(t)$ 而言，兩種資料型態在沒有設限時是無差異的。由 (4.1) 發現對累積發生函數的估計死亡的確切時間 Y 並不重要，而是只要知道 $\Delta = I(X \leq Y)$ 的值就足夠。

當外生設限發生時競爭風險資料可表示為 $(\tilde{X}_i, \delta_{x_i}, \delta_{y_i}) (i=1, \dots, n)$ 。當 $\delta_{x_i} + \delta_{y_i} = 0$ ， Δ_i 的值未知。3.2 節中已討論三種技巧的估計均可以得到同樣的估計量，其加權版的估計式如下：

$$\tilde{F}_1^C(t) = \int_{u \leq t} \sum_{i=1}^n \frac{I(\tilde{X}_i = u, \delta_{x_i} = 1)}{n \hat{S}_C(u-)} = \frac{1}{n} \sum_{i=1}^n \frac{I(\tilde{X}_i \leq t, \delta_{x_i} = 1)}{\hat{S}_C(\tilde{X}_i-)} . \quad (4.2)$$

設限下的半競爭風險資料可以表示為 $\{(\tilde{X}_i, \delta_{x_i}, \tilde{Y}_i, \eta_{y_i}) (i=1, \dots, n)\}$ ，其中 $\eta_{y_i} = I(Y_i \leq C_i)$ 。經過簡單分解運算後可發現

$$\Pr(X \wedge Y = x, Y = y) = \frac{\Pr(\tilde{X} = x, \tilde{Y} = y, \eta_y = 1)}{S_C(y-)} ,$$

其中分子與分母都可以分別估計。我們依加權的概念估計 $F_1(t)$ 得到以下的估計量：

$$\tilde{F}_1^{SC}(t) = \frac{1}{n} \sum_{i=1}^n \frac{I(\tilde{X}_i \leq t, \tilde{Y}_i > \tilde{X}_i, \eta_{y_i} = 1)}{\tilde{S}_C(\tilde{Y}_i-)} , \quad (4.3)$$

其中

$$\tilde{S}_C(y) = \prod_{\tilde{Y}_j \leq y} \left\{ 1 - \frac{\sum_{j=1}^n I(\tilde{Y}_j = \tilde{Y}_i, \eta_{yi} = 0)}{\sum_{j=1}^n I(\tilde{Y}_j \geq \tilde{Y}_i)} \right\}.$$

在模擬中我們比較兩個估計量 $\tilde{F}_1^C(t)$ (4.2) 與 $\tilde{F}_1^{SC}(t)$ (4.3)，發現 $\tilde{F}_1^C(t)$ 和 $\tilde{F}_1^{SC}(t)$ 的表現不分軒輊各有好壞， (\tilde{Y}_i, η_{yi}) 提供的多餘資訊在未截切時可能對猜測 Δ_i 的值有幫助。

4.2 競爭風險下的截切資料

截切的條件和取樣的限制有關。我們先考慮競爭風險下的截切，令 A 代表進入樣本的年齡， $X \wedge Y \geq A$ 的截切條件代表指進入研究的人是尚未得病的活人。考慮設限的變數，則截切條件修訂為 $\tilde{X} \geq A$ ，代表進入研究的人除了是尚未得病且存活外，還需要突破潛在的設限變因 ($C > A$)。我們先考慮以下的資料結構： $\{(\tilde{X}_i, \delta_{Xi}, \delta_{Yi}, A_i, C_i)(i=1, \dots, n)\}$ ，受制於 $\tilde{X}_i \geq A_i$ 的截切條件。可發現

$$\sum_{i=1}^n I(\tilde{X}_i = u, \delta_{Xi} = 1) / n \text{ 估計}$$

$$E[I(\tilde{X} = u, \delta_x = 1 | \tilde{X} \geq A)] = \frac{1}{\Pr(\tilde{X} \geq A)} \Pr(X = u, Y \geq u) \Pr(A \leq u, C \geq u)。$$

上式可以表示為

$$\Pr(X = u, Y \geq u) = \frac{E[I(\tilde{X} = u, \delta_x = 1 | \tilde{X} \geq A)] \Pr(\tilde{X} \geq A)}{\Pr(A \leq u, C \leq u)}。$$

令 $w(t) = \Pr(A \leq t, C \geq t)$ ，以加權的方法可得 $\Pr(X = t, Y > t)$ 的估計量為

$$\hat{\Pr}(\tilde{X} \geq A) \cdot \sum_{i=1}^n \frac{I(\tilde{X}_i = t, \delta_{Xi} = 1)}{n \hat{w}(t)}, \quad (4.4)$$

其中 $\hat{w}(t)$ 與 $\hat{\Pr}(\tilde{X} \geq A)$ 分別為 $w(t)$ 和 $\Pr(\tilde{X} \geq A)$ 的估計量。在截切條件下

$\sum_{i=1}^n I(A_i \leq t, C_i \geq t) / n$ 所估計的是

$$\frac{\Pr(A \leq t, C \geq t, X \geq A, Y \geq A)}{\Pr(\tilde{X} \geq A)},$$

所以加權的方法會得到

$$\hat{w}(t) = \hat{\Pr}(\tilde{X} \geq A) \cdot \sum_{i=1}^n \frac{I(A_i \leq t, C_i \geq t)}{n \hat{S}_{x \wedge y}^T(A_i^-)},$$

其中 $\hat{S}_{x \wedge y}^T(t)$ 是 $\Pr(X > t, Y > t)$ 的估計量，公式記錄在(3.7)中。綜合以上結果

$\Pr(X = t, Y \geq t)$ 的估計量可表示為

$$\sum_{i=1}^n \frac{I(\tilde{X}_i = t, \delta_{X_i} = 1)}{n} \left(\sum_{i=1}^n \frac{I(A_i \leq t, C_i \geq t)}{n \hat{S}_{x \wedge y}^T(A_i^-)} \right)^{-1}.$$

可得 $\Pr(X \leq t, X < Y)$ 的估計量為

$$\hat{F}_1^W(t) = \sum_{u \leq t} \left[\sum_{i=1}^n \frac{I(\tilde{X}_i = u, \delta_{X_i} = 1)}{n} \left(\sum_{i=1}^n \frac{I(A_i \leq u, C_i \geq u)}{n \hat{S}_{x \wedge y}^T(A_i^-)} \right)^{-1} \right]. \quad (4.5)$$

以上估計量可與以下 Peng & Fine (2006) 所提出的 naïve 估計量(原(3.8)式)比較

$$\hat{F}_1^{PN}(t) = \sum_{u \leq t} \left\{ \prod_{s < u} \left[1 - \frac{\sum_{i=1}^n I(\tilde{X}_i = s, \delta_{X_i} + \delta_{Y_i} = 1, A_i \leq s)}{\sum_{i=1}^n I(\tilde{X}_i \geq s, A_i \leq s)} \right] \left[\frac{\sum_{i=1}^n I(\tilde{X}_i = u, \delta_{X_i} = 1, A_i \leq u)}{\sum_{i=1}^n I(\tilde{X}_i \geq u, A_i \leq u)} \right] \right\}.$$

我們可發現(4.5)式中多用了 C 的資訊，在模擬中表現也確實比較準確。

4.3 半競爭風險下的截切資料

若資料為半競爭風險資料時，在設限並且截切條件放寬為 $\tilde{Y} \geq A$ (進入樣本時，個體必須是存活的状态，但是可容許進入研究前已得病，即 $X \leq A$)。可觀察到的資料為 $\{(\tilde{X}_i, \delta_{X_i}, \tilde{Y}_i, \eta_{Y_i}, A_i) (i=1, \dots, n)\}$ 。我們可得

$$\begin{aligned} & E[I(\tilde{X} = t, \delta_x = 1) | \tilde{Y} \geq A] \\ &= \frac{1}{\Pr(\tilde{Y} \geq A)} E[I(X = t, Y \geq t, C \geq t, \tilde{Y} \geq A)] \\ &= \frac{1}{\Pr(\tilde{Y} \geq A)} E_{X,Y,C} [E_A[I(X = t, Y \geq t, C \geq t, Y \geq A, C \geq A | X, Y, C)]] \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{\Pr(\tilde{Y} \geq A)} E_{X,Y,C} [I(X = t, Y \geq t, C \geq t) \Pr(A \leq Y, A \leq C | X, Y, C)] \\
&= \frac{1}{\Pr(\tilde{Y} \geq A)} E_{X,Y,C} [I(X = t, Y \geq t, C \geq t) \Pr(A \leq \tilde{Y} | \tilde{Y})]
\end{aligned}$$

進一步整理可得

$$E \left[\frac{I(\tilde{X} = t, \delta_x = 1)}{F_A(\tilde{Y} | \tilde{Y})} \mid \tilde{Y} \geq A \right] = \frac{\Pr(X = t, Y \geq t) S_C(t-)}{\Pr(\tilde{Y} \geq A)},$$

其中 $F_A(t) = \Pr(A \leq t)$ ，且 $S_C(t) = \Pr(C > t)$ ， $S_C(t-) = \Pr(C \geq t)$ 。對於 $F_A(t)$ 的估計，我們令 $C = A + B$ 且 $B > 0$ ，並假設 A （代表進入研究時的年齡）與 B （代表研究時間長度）獨立，並利用 product-limit decomposition 和 artificial truncation 的概念，獲知

$$\begin{aligned}
F_A(t) &= \Pr(A \leq t) = \prod_{u>t} \left\{ 1 - \frac{\Pr(A = u)}{\Pr(A \leq u)} \right\} \\
&= \prod_{u>t} \left\{ 1 - \frac{\Pr(A = u, Y \geq u)}{\Pr(A \leq u, Y \geq u)} \right\}
\end{aligned}$$

發現

$$\begin{aligned}
\Pr(A = u, \tilde{Y} \geq u \mid \tilde{Y} \geq A) &= \frac{\Pr(A = u, Y \geq u, C \geq u)}{\Pr(\tilde{Y} \geq A)} \\
&= \frac{\Pr(A = u, Y \geq u)}{\Pr(\tilde{Y} \geq A)}
\end{aligned}$$

並利用加權法，

$$\begin{aligned}
E \left[\frac{I(A \leq u, \tilde{Y} \geq u)}{S_B((u-A)^-)} \mid \tilde{Y} \geq A \right] &= \frac{1}{\Pr(\tilde{Y} \geq A)} E \left\{ E \left[\frac{I(A \leq u, \tilde{Y} \geq u)}{S_B((u-A)^-)} \mid A, Y \right] \right\} \\
&= \frac{1}{\Pr(\tilde{Y} \geq A)} E \left\{ E \left[\frac{I(A \leq u, Y \geq u, B \geq u-A)}{S_B((u-A)^-)} \mid A, Y \right] \right\} \\
&= \frac{1}{\Pr(\tilde{Y} \geq A)} E \left\{ I(A \leq u, Y \geq u) E \left[\frac{I(B \geq u-A)}{S_B((u-A)^-)} \mid A, Y \right] \right\} \\
&= \frac{\Pr(A \leq u, Y \geq u)}{\Pr(\tilde{Y} \geq A)}
\end{aligned}$$

因此我們可以 $\frac{\sum_{i=1}^n I(A_i = u, \tilde{Y}_i \geq u)}{\sum_{i=1}^n I(A_i \leq u, \tilde{Y}_i \geq u) / \hat{S}_B((u - A_i)^-)}$ 估計 $\frac{\Pr(A = u)}{\Pr(A \leq u)}$ 的，而 $F_A(t)$ 的

估計量則可以表示成

$$\tilde{F}_A(t) = \prod_{u>t} \left\{ 1 - \frac{\sum_{i=1}^n I(A_i = u, \tilde{Y}_i \geq u)}{\sum_{i=1}^n I(A_i \leq u, \tilde{Y}_i \geq u) / \hat{S}_B((u - A_i)^-)} \right\} \quad (4.6)$$

那麼我們可利用 $\frac{1}{n} \sum_{i=1}^n \frac{I(\tilde{X}_i = t, \delta_{xi} = 1)}{\tilde{F}_A(\tilde{Y}_i)}$ 做為 $\frac{\Pr(X = t, Y \geq t) S_C(t-)}{\Pr(\tilde{Y} \geq A)}$ 的估計量。考慮

$\Pr(\tilde{Y} \geq A)$ 的估計，可得

$$\begin{aligned} E \left[\frac{1}{F_A(\tilde{Y} | \tilde{Y})} \mid \tilde{Y} \geq A \right] &= \frac{1}{\Pr(\tilde{Y} \geq A)} E \left[\frac{I(\tilde{Y} \geq A)}{F_A(\tilde{Y} | \tilde{Y})} \right] \\ &= \frac{1}{\Pr(\tilde{Y} \geq A)} E \left[E \left[\frac{I(\tilde{Y} \geq A)}{F_A(\tilde{Y} | \tilde{Y})} \mid Y, C \right] \right] \\ &= \frac{1}{\Pr(\tilde{Y} \geq A)} E \left[\frac{1}{F_A(\tilde{Y} | \tilde{Y})} F_A(\tilde{Y} | \tilde{Y}) \right] \\ &= \frac{1}{\Pr(\tilde{Y} \geq A)} \circ \end{aligned}$$

因此 $\frac{1}{n} \sum_{i=1}^n \frac{1}{\tilde{F}_A(\tilde{Y}_i)}$ 可以做為 $\frac{1}{\Pr(\tilde{Y} \geq A)}$ 的估計量。有關 $S_C(t-)$ 的估計，我們利用

條件期望值的操作得

$$\begin{aligned} S_C(t^-) &= \Pr(A + B \geq t) \\ &= E[I(A + B \geq t)] \\ &= E_A[E_B[I(A + B \geq t) \mid A]] \\ &= E_A[E_B[I(B \geq t - A) \mid A]] \\ &= E_A[S_B((t - A)^- \mid A)] \end{aligned}$$

$$= \int_A S_B((t-u)^-) dF_A(u)。$$

根據 $\{(\tilde{Y}_i - A_i, 1 - \eta_{Y_i}) (i = 1, \dots, n)\}$ 的資料，可得 $S_B(t) = \Pr(B > t)$ 的 product-limit 估

計量為：

$$\tilde{S}_B^T(t) = \prod_{u \leq t} \left\{ 1 - \frac{\sum_{i=1}^n I(\tilde{Y}_i - A_i = u, \eta_{Y_i} = 0)}{\sum_{i=1}^n I(\tilde{Y}_i - A_i \geq u)} \right\}。 \quad (4.7)$$

將 $\tilde{S}_B^T(t)$ 代入，可獲得估計量 $\tilde{S}_C^T(C \geq t) = \int_A \tilde{S}_B^T((t-u)^-) d\tilde{F}_A(u)$ 如下：

$$\sum_{j=1}^n \left[\prod_{k < (t-A_j)} \left[1 - \frac{\sum_{i=1}^n I(\tilde{Y}_i - A_i = k, \eta_{Y_i} = 0)}{\sum_{i=1}^n I(\tilde{Y}_i - A_i \geq k)} \right] \left[\tilde{F}_A(A_j) - \tilde{F}_A(A_{j-1}) \right] \right] \quad (4.8)$$

其中 $A_0 = 0$ 。綜合以上分析可得 $\Pr(X = t, X < Y)$ 的估計量如下：

$$\sum_{i=1}^n \frac{I(\tilde{X}_i = t, \delta_{xi} = 1)}{\tilde{F}_A(\tilde{Y}_i)} \left[\sum_{i=1}^n \frac{1}{\tilde{F}_A(\tilde{Y}_i)} \right]^{-1} \left[\tilde{S}_C^T(t-) \right]^{-1}。$$

基於 $\Pr(X \leq t, X < Y) = \int_{u=0}^t \Pr(X = u, Y > u) du$ ，我們提出適用於截切下的半競爭

風險資料 $F_1(t)$ 估計量為：

$$\tilde{F}_1^W(t) = \sum_{u \leq t} \left\{ \sum_{i=1}^n \frac{I(\tilde{X}_i = u, \delta_{xi} = 1)}{\tilde{F}_A(\tilde{Y}_i)} \left[\sum_{i=1}^n \frac{1}{\tilde{F}_A(\tilde{Y}_i)} \right]^{-1} \left[\tilde{S}_C^T(u-) \right]^{-1} \right\}。 \quad (4.9)$$

將上式估計量與 Peng & Fine (2006) 所提出的半競爭風險估計量(原(3.20)式)

比較，Peng & Fine 在估計 A 和 C 的函數時，是估計兩個變數的聯合函數，而我

們的估計量則是將 A 和 C 的函數分開估計，但必須有較強的假設 $C = A + B$ 且 A

與 B 獨立。為了更清楚估計量的準確性，由 CIF 的模擬試驗發現 $\tilde{F}_1^W(t)$ 的表現比

$\tilde{F}_1^{PF}(t)$ 為好。

第五章 模擬實驗

5.1 資料生成方法

在本章我們利用電腦模擬的方式檢驗推論方法的相對表現。之前的討論中，我們假設發病時間與競爭風險發生時間的聯合分配是 copula 模式分配。模擬程式假設發病時間與競爭風險發生時間服從 Clayton 分配，其聯合分配函數可以表示為

$$F(x, y) = \begin{cases} \left[F_1(x)^{1-\theta} + F_2(y)^{1-\theta} - 1 \right]^{-1} & \text{if } \theta > 1, \\ F_1(x)F_2(y) & \text{if } \theta = 1 \end{cases}$$

其中參數 θ 控制關聯性的強度，與 Kendall's tau 有以下關係：

$$\tau = \frac{\theta - 1}{\theta + 1}。$$

模擬實驗的資料生成過程整理如下：

第一步：生成具有關聯性的連續變數 $(X, Y) \sim \text{Clayton}(\tau)$ 如下：

- 給定 Kendall's tau，則可以計算出 $\theta = \frac{1+\tau}{1-\tau}$ ；
- 生成彼此獨立的變數 (U_1, U_2) ，其中 $U_i \sim \text{Uniform}(0,1)$ ($i=1,2$)；
- $a = (1-U_2)^{1-\theta}$ ，
 $Y = -\log(1-U_2)$ ，

$$X = \frac{1}{\theta-1} \log \left\{ (1-a) + a(1-U_1)^{\frac{1-\theta}{\theta}} \right\}。$$

重覆(a)至(c)步驟，可得 (X_i, Y_i) 。

第二步：生成截切變數 A 、研究期間長度之變數 B ，與設限變數 C 如下：

- $A \sim U(-a, k_1)$ 當 $a=0$ 時，代表排除研究開始時尚未出生者；
- $B \sim U(0, k_2)$ 代表研究期間長度，亦可為定值；
- $C = A + B$ 。

重覆上敘(d)至(f)步驟，可得 (A_i, C_i) 。

第三步：生成截切資料

只允許滿足截切條件的 (X_i, Y_i) 進入研究。有兩種截切條件一為

$\tilde{X}_i > A_i$ ，另一為 $\tilde{Y}_i > A_i$ 。

第四步：在競爭風險架構下，製造可觀察的時間變數 \tilde{X} 和保留原有指標指標

(δ_x, δ_y) ：

(1) 當 $X_i \leq (Y_i \wedge C_i)$ 時，則 $\tilde{X}_i = X_i$ 和 $(\delta_{xi}, \delta_{yi}) = (1,0)$ ；

(2) 當 $Y_i \leq (X_i \wedge C_i)$ 時，則 $\tilde{X}_i = Y_i$ 和 $(\delta_{xi}, \delta_{yi}) = (0,1)$ ；

(3) 當 $C_i \leq (X_i \wedge Y_i)$ 時，則 $\tilde{X}_i = C_i$ 和 $(\delta_{xi}, \delta_{yi}) = (0,0)$ 。

可得到資料 $(\tilde{X}_i, \delta_{xi}, \delta_{yi})$ ($i = 1, 2, \dots, n$)。

第五步：在半競爭風險架構下，製造可觀察的時間變數 (\tilde{X}, \tilde{Y}) 和生前是否發病與

研究結束前是否死亡之指標 (δ_x, η_y) ：

(1) 當 $X_i \leq (Y_i \wedge C_i)$ 且 $Y_i \leq C_i$ 時，則 $(\tilde{X}_i, \tilde{Y}_i) = (X_i, Y_i)$ 和 $(\delta_{xi}, \eta_{yi}) = (1,1)$ ；

(2) 當 $X_i \leq (Y_i \wedge C_i)$ 且 $Y_i > C_i$ 時，則 $(\tilde{X}_i, \tilde{Y}_i) = (X_i, C_i)$ 和

$(\delta_{xi}, \eta_{yi}) = (1,0)$ ；

(3) 當 $X_i > (Y_i \wedge C_i)$ 且 $Y_i \leq C_i$ 時，則 $(\tilde{X}_i, \tilde{Y}_i) = (Y_i, Y_i)$ 和 $(\delta_{xi}, \eta_{yi}) = (0,1)$ ；

(4) 當 $X_i > (Y_i \wedge C_i)$ 且 $Y_i > C_i$ 時，則 $(\tilde{X}_i, \tilde{Y}_i) = (C_i, C_i)$ 和

$(\delta_{xi}, \eta_{yi}) = (0,0)$ 。

可得到資料 $(\tilde{X}_i, \tilde{Y}_i, \delta_{xi}, \eta_{yi})$ ($i = 1, 2, \dots, n$)。

當資料生成後，我們即可比較同種資料型態下所提出不同的估計量。比較的標準為重覆實驗 1000 次後的“平均誤差” (average bias) 與“標準差” (standard deviation)。

5.2 競爭風險資料之估計量比較

我們比較以下兩個 $F_1(t)$ 的估計量(未截切樣本數 1000，重複 1000 次)：

$$\hat{F}_1^{PN}(t) = \sum_{u \leq t} \left\{ \prod_{s < u} \left[1 - \frac{\sum_{i=1}^n I(\tilde{X}_i = s, \delta_{xi} + \delta_{yi} = 1, A_i \leq s)}{\sum_{i=1}^n I(\tilde{X}_i \geq s, A_i \leq s)} \right] \left[\frac{\sum_{i=1}^n I(\tilde{X}_i = u, \delta_{xi} = i, A_i \leq u)}{\sum_{i=1}^n I(\tilde{X}_i \geq u, A_i \leq u)} \right] \right\},$$

$$\hat{F}_1^W(t) = \sum_{u \leq t} \left[\sum_{i=1}^n \frac{I(\tilde{X}_i = u, \delta_{xi} = 1)}{n} \left(\sum_{i=1}^n \frac{I(A_i < u, C_i > u)}{n \hat{S}_{X,Y}^T(A_i)} \right)^{-1} \right].$$

會影響的變因有 τ 、 $\Pr(X \wedge Y \geq A)$ 、 $\Pr(C > X | X \wedge Y \geq A)$ 、 $\Pr(C > Y | X \wedge Y \geq A)$ 。

模擬包含四個分析，分別是分析一：未截切未設限；分析二：未截切右設限；分析三：左截切未設限，與分析四：左截切右設限。結果記錄於表 5-1 到表 5-20。

(分析一)中兩個估計量一致，對於(分析二)我們原本預期亦會得到一致的答案，卻發現兩個估計量會卻有些許的出入。更進一步分析，當設 $A_i = 0$ 時，Naïve 估計量 $\hat{F}_1^{PN}(t)$ 的確和 [3.3] 的估計量是完全相同。然而權重估計量會變成

$$\hat{F}_1^W(t) = \sum_{u \leq t} \left[\sum_{i=1}^n \frac{I(\tilde{X}_i = u, \delta_{xi} = 1)}{n} \left(\sum_{i=1}^n \frac{I(C_i > u)}{n} \right)^{-1} \right],$$

和(3.5)比較會發現 $\hat{F}_1^W(t)$ 以 $\sum_{i=1}^n \frac{I(C_i > u)}{n}$ 估計 $\Pr(C > u)$ ，然而(3.5)式卻是用

Kaplan-Meier 估計量 $\hat{S}_C(u)$ 估計。若估計 $\Pr(C > u)$ 為主要目標，則 $\sum_{i=1}^n \frac{I(C_i > u)}{n}$

顯然是最好的估計量。但是 $\Pr(C > u)$ 卻是估計 $F_1(t)$ 的 nuisance parameter，文獻曾討論過代入一個較差的估計量有時候反而會得到更穩定的估計量。

(分析三)與(分析四)加入了左截切條件，發現權重估計量的標準差較小。若是 C 的完整資訊不可得，則有必要調整 $\hat{F}_1^W(t)$ 的權數估計方法，也就是應該尋求另一種方法以估計 $\Pr(A \leq u, C \geq u | \tilde{X} \geq A)$ 。

5.3 比較 Peng and Fine 估計量與半競爭風險權重估計量

我們比較以下兩個 $F_1(t)$ 的估計量(未截切樣本數 1000，重複 1000 次)：

Peng & Fine estimator :

$$\tilde{F}_1^{PF}(t) = \frac{1}{n} \sum_{i=1}^n \frac{\tilde{S}_Y^T(\tilde{Y}_i^-)}{C_n(\tilde{Y}_i)} I(\tilde{X}_i \leq t, \tilde{X}_i < \tilde{Y}_i \leq \tau, \eta_{yi} = 1) + \tilde{S}_T^Y(\tau) \left\{ 1 - \frac{C_{2,n}(t, \tau^+)}{C_n(\tau^+)} \right\};$$

半競爭風險權重估計量：

$$\tilde{F}_1^W(t) = \sum_{u \leq t} \left\{ \sum_{i=1}^n \frac{I(\tilde{X}_i = u, \delta_{xi} = 1)}{\tilde{F}_A(\tilde{Y}_i)} \left[\sum_{i=1}^n \frac{1}{\tilde{F}_A(\tilde{Y}_i)} \right]^{-1} \left[\tilde{S}_C^T(u^-) \right]^{-1} \right\}.$$

模擬包含四個分析，分別是分析五：未截切未設限，分析六：未截切右設限，分析七：左截切未設限，與分析八：左截切右設限。模擬結果記錄於表 5-21 至表 5-40。

模擬結果發現，(分析五)中兩估計量得到相同結果。(分析六)加入右設限條件，兩者的偏差與標準差均增加，以 $\tilde{F}_1^{PF}(t)$ 增加較多；此時兩估計量的相對表現是權重估計量 $\tilde{F}_1^W(t)$ 略好，也就是其偏差與標準差普遍均比 $\tilde{F}_1^{PF}(t)$ 小。(分析七)考慮資料只受制於左截切， $\tilde{F}_1^W(t)$ 的標準差較 $\tilde{F}_1^{PF}(t)$ 為小，偏誤則以 $\tilde{F}_1^{PF}(t)$ 較小。(分析八) 多考慮設限的情形，兩個估計量亦呈現同樣的的相對表現。

綜合以上分析發現為截切對估計的影響大於設限，原因可能是因為截切是資料完全的遺失。我們提出的加權方法在很多情形下表現比 Peng and Fine (2006) 提出的方法更好。然而我們的方法在競爭風險架構下利用 C 的完整資訊，在半競爭風險架構下假設 A 與 B 獨立，所以評比未必公平。然而因論文時間有限，無法進一步討論放寬現有假設下的權數估計方法。

比較 Peng and Fine 的競爭風險與半競爭風險兩種估計方法，於未截切架構下，標準差比值與，均方差(Mean square error)比值分別為：

$$\frac{SD_{CR}}{SD_{SCR}} \in [0.85, 1.02] \quad \frac{MSE_{CR}}{MSE_{SCR}} \in [0.71, 1.04],$$

但是當資料受限於截切時，則是

$$\frac{SD_{CR}}{SD_{SCR}} \in [1.26, 3.57] \quad \frac{MSE_{CR}}{MSE_{SCR}} \in [1.62, 6.33],$$

表示半競爭風險資料在截切架構下對函數估計有明顯幫助。接著，觀察加權法之兩種方法，資料未受限截切時，標準差比值與均方差比值如下：

$$\frac{SD_{CR}}{SD_{SCR}} \in [0.92, 1.02] \quad \frac{MSE_{CR}}{MSE_{SCR}} \in [0.84, 1.05] ,$$

然而受截切條件下，各比值為：

$$\frac{SD_{CR}}{SD_{SCR}} \in [1.28, 3.39] \quad \frac{MSE_{CR}}{MSE_{SCR}} \in [1.25, 7.37] ,$$

這與 Peng and Fine 兩估計量所呈現的結果一致，半競爭風險資料於資料受截切時可有效的提供資訊，使得函數估計更為準確。



分析一：競爭風險（未截切，未設限）

$F_1(t)$		Weighting	Naïve
0.1	Average bias	-0.000465	-0.000465
	(Std)	(0.009189)	(0.009189)
	MSE	0.000085	0.000085
0.2	Average bias	-0.000451	-0.000451
	(Std)	(0.012180)	(0.012180)
	MSE	0.000149	0.000149
0.3	Average bias	-0.000959	-0.000959
	(Std)	(0.014373)	(0.014373)
	MSE	0.000208	0.000208
0.4	Average bias	-0.000464	-0.000464
	(Std)	(0.015416)	(0.015416)
	MSE	0.000138	0.000138

表 5-1： $\hat{F}_1^W(t)$ 與 $\hat{F}_1^{PF}(t)$ 之比較

$\tau = 0.9$, $\Pr(\tilde{X} \geq A) = 1.0$, $\Pr(X < C | \tilde{X} \geq A) = 1.0$, $\Pr(Y < C | \tilde{X} \geq A) = 1.0$

$F_1(t)$		Weighting	Naïve
0.1	Average bias	-0.000349	-0.000349
	(Std)	(0.008961)	(0.008961)
	MSE	0.000080	0.000080
0.2	Average bias	-0.000138	-0.000138
	(Std)	(0.012483)	(0.012483)
	MSE	0.000156	0.000156
0.3	Average bias	-0.000553	-0.000553
	(Std)	(0.014476)	(0.014476)
	MSE	0.000210	0.000210
0.4	Average bias	-0.000012	-0.000012
	(Std)	(0.015584)	(0.015584)
	MSE	0.000243	0.000243

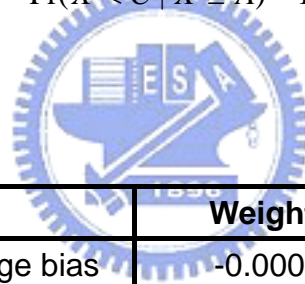
表 5-2： $\hat{F}_1^W(t)$ 與 $\hat{F}_1^{PF}(t)$ 之比較

$\tau = 0.7$, $\Pr(\tilde{X} \geq A) = 1.0$, $\Pr(X < C | \tilde{X} \geq A) = 1.0$, $\Pr(Y < C | \tilde{X} \geq A) = 1.0$

$F_1(t)$		Weighting	Naïve
0.1	Average bias	-0.000279	-0.000279
	(Std)	(0.008981)	(0.008981)
	MSE	0.000081	0.000081
0.2	Average bias	-0.000088	-0.000088
	(Std)	(0.012718)	(0.012718)
	MSE	0.000162	0.000162
0.3	Average bias	-0.000600	-0.000600
	(Std)	(0.014765)	(0.014765)
	MSE	0.000218	0.000218
0.4	Average bias	-0.000074	-0.000074
	(Std)	(0.015631)	(0.015631)
	MSE	0.000244	0.000244

表 5-3： $\hat{F}_1^W(t)$ 與 $\hat{F}_1^{PF}(t)$ 之比較

$\tau = 0.5$, $\Pr(\tilde{X} \geq A) = 1.0$, $\Pr(X < C | \tilde{X} \geq A) = 1.0$, $\Pr(Y < C | \tilde{X} \geq A) = 1.0$



$F_1(t)$		Weighting	Naïve
0.1	Average bias	-0.000098	-0.000098
	(Std)	(0.009112)	(0.009112)
	MSE	0.000083	0.000083
0.2	Average bias	0.000035	0.000035
	(Std)	(0.012778)	(0.012778)
	MSE	0.000163	0.000163
0.3	Average bias	-0.000163	-0.000163
	(Std)	(0.014380)	(0.014380)
	MSE	0.000207	0.000207
0.4	Average bias	0.000261	0.000261
	(Std)	(0.015303)	(0.015303)
	MSE	0.000234	0.000234

表 5-4： $\hat{F}_1^W(t)$ 與 $\hat{F}_1^{PF}(t)$ 之比較

$\tau = 0.3$, $\Pr(\tilde{X} \geq A) = 1.0$, $\Pr(X < C | \tilde{X} \geq A) = 1.0$, $\Pr(Y < C | \tilde{X} \geq A) = 1.0$

$F_1(t)$		Weighting	Naïve
0.1	Average bias	-0.000154	-0.000154
	(Std)	(0.009415)	(0.009415)
	MSE	0.000089	0.000089
0.2	Average bias	0.000053	0.000053
	(Std)	(0.012545)	(0.012545)
	MSE	0.000157	0.000157
0.3	Average bias	0.000102	0.000102
	(Std)	(0.014504)	(0.014504)
	MSE	0.000210	0.000210
0.4	Average bias	0.000485	0.000485
	(Std)	(0.015135)	(0.015135)
	MSE	0.000229	0.000229

表 5-5： $\hat{F}_1^W(t)$ 與 $\hat{F}_1^{PF}(t)$ 之比較

$$\tau = 0.1, \Pr(\tilde{X} \geq A) = 1.0, \Pr(X < C | \tilde{X} \geq A) = 1.0, \Pr(Y < C | \tilde{X} \geq A) = 1.0$$

分析二：競爭風險（未截切，右設限）

$F_1(t)$		Weighting	Naïve
0.1	Average bias	-0.000417	-0.000417
	(Std)	(0.009320)	(0.009323)
	MSE	0.000087	0.000087
0.2	Average bias	-0.000312	-0.000312
	(Std)	(0.012550)	(0.012556)
	MSE	0.000158	0.000158
0.3	Average bias	-0.000839	-0.000848
	(Std)	(0.014960)	(0.014929)
	MSE	0.000225	0.000224
0.4	Average bias	-0.000538	-0.000525
	(Std)	(0.016872)	(0.016703)
	MSE	0.000225	0.000279

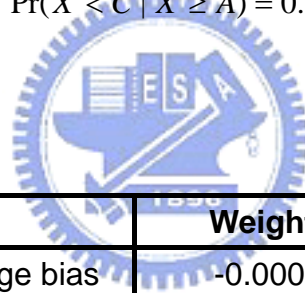
表 5-6： $\hat{F}_1^W(t)$ 與 $\hat{F}_1^{PF}(t)$ 之比較

$$\tau = 0.9, \Pr(\tilde{X} \geq A) = 1.0, \Pr(X < C | \tilde{X} \geq A) = 0.80, \Pr(Y < C | \tilde{X} \geq A) = 0.80$$

$F_1(t)$		Weighting	Naïve
0.1	Average bias	-0.000344	-0.000345
	(Std)	(0.009039)	(0.009042)
	MSE	0.000082	0.000082
0.2	Average bias	-0.000073	-0.000078
	(Std)	(0.012690)	(0.012689)
	MSE	0.000161	0.000161
0.3	Average bias	-0.000457	-0.000464
	(Std)	(0.015000)	(0.014960)
	MSE	0.000225	0.000224
0.4	Average bias	-0.000203	-0.000195
	(Std)	(0.016841)	(0.016694)
	MSE	0.000284	0.000279

表 5-7： $\hat{F}_1^W(t)$ 與 $\hat{F}_1^{PF}(t)$ 之比較

$\tau = 0.7$, $\Pr(\tilde{X} \geq A) = 1.0$, $\Pr(X < C | \tilde{X} \geq A) = 0.80$, $\Pr(Y < C | \tilde{X} \geq A) = 0.80$



$F_1(t)$		Weighting	Naïve
0.1	Average bias	-0.000300	-0.000301
	(Std)	(0.009034)	(0.009034)
	MSE	0.000082	0.000082
0.2	Average bias	-0.000078	-0.000806
	(Std)	(0.012880)	(0.012882)
	MSE	0.000166	0.000167
0.3	Average bias	-0.000467	-0.000481
	(Std)	(0.015064)	(0.015049)
	MSE	0.000227	0.000227
0.4	Average bias	-0.000124	-0.000149
	(Std)	(0.016382)	(0.016327)
	MSE	0.000268	0.000267

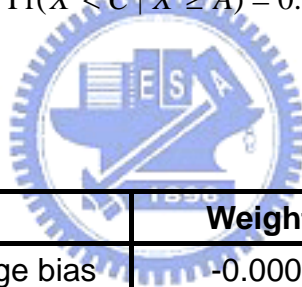
表 5-8： $\hat{F}_1^W(t)$ 與 $\hat{F}_1^{PF}(t)$ 之比較

$\tau = 0.5$, $\Pr(\tilde{X} \geq A) = 1.0$, $\Pr(X < C | \tilde{X} \geq A) = 0.80$, $\Pr(Y < C | \tilde{X} \geq A) = 0.80$

$F_1(t)$		Weighting	Naïve
0.1	Average bias	-0.000110	-0.000112
	(Std)	(0.009215)	(0.009212)
	MSE	0.000085	0.000085
0.2	Average bias	0.000030	0.000019
	(Std)	(0.012980)	(0.012978)
	MSE	0.000168	0.000168
0.3	Average bias	0.000033	0.000031
	(Std)	(0.014839)	(0.014817)
	MSE	0.000220	0.000220
0.4	Average bias	0.000277	0.000258
	(Std)	(0.015937)	(0.015865)
	MSE	0.000254	0.000252

表 5-9： $\hat{F}_1^W(t)$ 與 $\hat{F}_1^{PF}(t)$ 之比較

$\tau = 0.3$, $\Pr(\tilde{X} \geq A) = 1.0$, $\Pr(X < C | \tilde{X} \geq A) = 0.80$, $\Pr(Y < C | \tilde{X} \geq A) = 0.80$



$F_1(t)$		Weighting	Naïve
0.1	Average bias	-0.000174	-0.000174
	(Std)	(0.009483)	(0.009483)
	MSE	0.000090	0.000090
0.2	Average bias	0.000094	0.000074
	(Std)	(0.012698)	(0.012704)
	MSE	0.000161	0.000161
0.3	Average bias	0.000274	0.000244
	(Std)	(0.014804)	(0.014797)
	MSE	0.000219	0.000219
0.4	Average bias	0.000516	0.000466
	(Std)	(0.015595)	(0.015496)
	MSE	0.000243	0.000240

表 5-10： $\hat{F}_1^W(t)$ 與 $\hat{F}_1^{PF}(t)$ 之比較

$\tau = 0.1$, $\Pr(\tilde{X} \geq A) = 1.0$, $\Pr(X < C | \tilde{X} \geq A) = 0.80$, $\Pr(Y < C | \tilde{X} \geq A) = 0.80$

分析三：競爭風險（左截切，未設限）

$F_1(t)$		Weighting	Naïve
0.1	Average bias	-0.000040	0.001688
	(Std)	(0.056008)	(0.068942)
	MSE	0.003137	0.004756
0.2	Average bias	0.001241	0.002763
	(Std)	(0.053213)	(0.064060)
	MSE	0.002833	0.004111
0.3	Average bias	0.001315	0.002642
	(Std)	(0.049128)	(0.058195)
	MSE	0.002415	0.003394
0.4	Average bias	0.002392	0.003506
	(Std)	(0.046183)	(0.053321)
	MSE	0.002139	0.002855

表 5-11： $\hat{F}_1^W(t)$ 與 $\hat{F}_1^{PF}(t)$ 之比較

$\tau = 0.9$, $\Pr(\tilde{X} \geq A) = 0.61$, $\Pr(X < C | \tilde{X} \geq A) = 1.0$, $\Pr(Y < C | \tilde{X} \geq A) = 1.0$

$F_1(t)$		Weighting	Naïve
0.1	Average bias	-0.001512	-0.000569
	(Std)	(0.058636)	(0.065185)
	MSE	0.003440	0.004249
0.2	Average bias	-0.001167	-0.000338
	(Std)	(0.056316)	(0.061749)
	MSE	0.003173	0.003813
0.3	Average bias	-0.000525	0.000192
	(Std)	(0.052616)	(0.057087)
	MSE	0.002769	0.003259
0.4	Average bias	0.000907	0.001521
	(Std)	(0.048661)	(0.052216)
	MSE	0.002369	0.002729

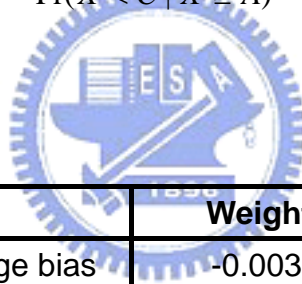
表 5-12： $\hat{F}_1^W(t)$ 與 $\hat{F}_1^{PF}(t)$ 之比較

$\tau = 0.7$, $\Pr(\tilde{X} \geq A) = 0.56$, $\Pr(X < C | \tilde{X} \geq A) = 1.0$, $\Pr(Y < C | \tilde{X} \geq A) = 1.0$

$F_1(t)$		Weighting	Naïve
0.1	Average bias	-0.002920	-0.001960
	(Std)	(0.059188)	(0.065690)
	MSE	0.003512	0.004319
0.2	Average bias	-0.002378	-0.001557
	(Std)	(0.056733)	(0.062146)
	MSE	0.003224	0.003865
0.3	Average bias	-0.001848	-0.001130
	(Std)	(0.052950)	(0.057411)
	MSE	0.002807	0.003297
0.4	Average bias	-0.000399	0.000223
	(Std)	(0.049392)	(0.052911)
	MSE	0.002440	0.002800

表 5-13 : $\hat{F}_1^W(t)$ 與 $\hat{F}_1^{PF}(t)$ 之比較

$\tau = 0.5$, $\Pr(\tilde{X} \geq A) = 0.52$, $\Pr(X < C | \tilde{X} \geq A) = 1.0$, $\Pr(Y < C | \tilde{X} \geq A) = 1.0$



$F_1(t)$		Weighting	Naïve
0.1	Average bias	-0.003423	-0.001550
	(Std)	(0.059756)	(0.072105)
	MSE	0.003582	0.005202
0.2	Average bias	-0.001568	0.000085
	(Std)	(0.058270)	(0.068393)
	MSE	0.003398	0.004678
0.3	Average bias	-0.001267	0.000190
	(Std)	(0.054822)	(0.063137)
	MSE	0.003007	0.003986
0.4	Average bias	0.000071	0.001313
	(Std)	(0.050962)	(0.057575)
	MSE	0.002597	0.003317

表 5-14 : $\hat{F}_1^W(t)$ 與 $\hat{F}_1^{PF}(t)$ 之比較

$\tau = 0.3$, $\Pr(\tilde{X} \geq A) = 0.48$, $\Pr(X < C | \tilde{X} \geq A) = 1.0$, $\Pr(Y < C | \tilde{X} \geq A) = 1.0$

$F_1(t)$		Weighting	Naïve
0.1	Average bias	-0.004830	-0.003565
	(Std)	(0.059042)	(0.065582)
	MSE	0.003509	0.004314
0.2	Average bias	-0.002931	-0.002066
	(Std)	(0.058407)	(0.063656)
	MSE	0.003420	0.004056
0.3	Average bias	-0.001671	-0.000944
	(Std)	(0.055772)	(0.060018)
	MSE	0.003113	0.003603
0.4	Average bias	-0.000486	0.000146
	(Std)	(0.052454)	(0.055777)
	MSE	0.002752	0.003111

表 5-15： $\hat{F}_1^W(t)$ 與 $\hat{F}_1^{PF}(t)$ 之比較

$$\tau = 0.1, \Pr(\tilde{X} \geq A) = 0.45, \Pr(X < C | \tilde{X} \geq A) = 1.0, \Pr(Y < C | \tilde{X} \geq A) = 1.0$$

分析四：競爭風險（左截切，右設限）

$F_1(t)$		Weighting	Naïve
0.1	Average bias	-0.000111	0.001634
	(Std)	(0.056112)	(0.069038)
	MSE	0.003149	0.004769
0.2	Average bias	0.001484	0.003015
	(Std)	(0.053348)	(0.064179)
	MSE	0.002848	0.004128
0.3	Average bias	0.001606	0.002943
	(Std)	(0.049324)	(0.058384)
	MSE	0.002435	0.003417
0.4	Average bias	0.002678	0.003774
	(Std)	(0.046610)	(0.053669)
	MSE	0.002180	0.002895

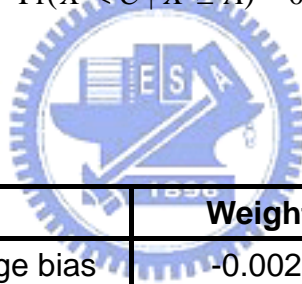
表 5-16： $\hat{F}_1^W(t)$ 與 $\hat{F}_1^{PF}(t)$ 之比較

$$\tau = 0.9, \Pr(\tilde{X} \geq A) = 0.61, \Pr(X < C | \tilde{X} \geq A) = 0.79, \Pr(Y < C | \tilde{X} \geq A) = 0.79$$

$F_1(t)$		Weighting	Naïve
0.1	Average bias	-0.001559	-0.000613
	(Std)	(0.058770)	(0.065293)
	MSE	0.003456	0.004264
0.2	Average bias	-0.001018	-0.000197
	(Std)	(0.056458)	(0.061880)
	MSE	0.003189	0.003829
0.3	Average bias	-0.000355	0.000386
	(Std)	(0.052848)	(0.057310)
	MSE	0.002793	0.003285
0.4	Average bias	0.000937	0.001581
	(Std)	(0.048912)	(0.052482)
	MSE	0.002393	0.002757

表 5-17 : $\hat{F}_1^W(t)$ 與 $\hat{F}_1^{PF}(t)$ 之比較

$\tau = 0.7$, $\Pr(\tilde{X} \geq A) = 0.56$, $\Pr(X < C | \tilde{X} \geq A) = 0.78$, $\Pr(Y < C | \tilde{X} \geq A) = 0.78$



$F_1(t)$		Weighting	Naïve
0.1	Average bias	-0.002949	-0.001993
	(Std)	(0.059097)	(0.065606)
	MSE	0.003501	0.004308
0.2	Average bias	-0.002385	-0.001582
	(Std)	(0.056709)	(0.062130)
	MSE	0.003222	0.003863
0.3	Average bias	-0.001731	-0.001002
	(Std)	(0.052895)	(0.057373)
	MSE	0.002801	0.003293
0.4	Average bias	-0.000450	0.000156
	(Std)	(0.049420)	(0.052977)
	MSE	0.002443	0.002807

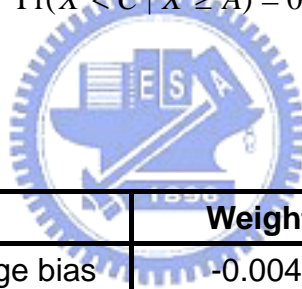
表 5-18 : $\hat{F}_1^W(t)$ 與 $\hat{F}_1^{PF}(t)$ 之比較

$\tau = 0.5$, $\Pr(\tilde{X} \geq A) = 0.52$, $\Pr(X < C | \tilde{X} \geq A) = 0.78$, $\Pr(Y < C | \tilde{X} \geq A) = 0.78$

$F_1(t)$		Weighting	Naïve
0.1	Average bias	-0.003441	-0.001577
	(Std)	(0.059650)	(0.072017)
	MSE	0.003570	0.005189
0.2	Average bias	-0.001542	0.000086
	(Std)	(0.058268)	(0.068402)
	MSE	0.003398	0.004679
0.3	Average bias	-0.001070	0.000368
	(Std)	(0.054886)	(0.063207)
	MSE	0.003014	0.003995
0.4	Average bias	0.000204	0.001425
	(Std)	(0.051001)	(0.057615)
	MSE	0.002601	0.003322

表 5-19 : $\hat{F}_1^W(t)$ 與 $\hat{F}_1^{PF}(t)$ 之比較

$\tau = 0.3$, $\Pr(\tilde{X} \geq A) = 0.48$, $\Pr(X < C | \tilde{X} \geq A) = 0.78$, $\Pr(Y < C | \tilde{X} \geq A) = 0.78$



$F_1(t)$		Weighting	Naïve
0.1	Average bias	-0.004892	-0.003929
	(Std)	(0.058869)	(0.065432)
	MSE	0.003489	0.004297
0.2	Average bias	-0.003043	-0.002193
	(Std)	(0.058363)	(0.063627)
	MSE	0.003415	0.004053
0.3	Average bias	-0.001663	-0.000931
	(Std)	(0.055822)	(0.060076)
	MSE	0.003119	0.003610
0.4	Average bias	-0.000475	0.000201
	(Std)	(0.052331)	(0.055692)
	MSE	0.002739	0.003102

表 5-20 : $\hat{F}_1^W(t)$ 與 $\hat{F}_1^{PF}(t)$ 之比較

$\tau = 0.1$, $\Pr(\tilde{X} \geq A) = 0.45$, $\Pr(X < C | \tilde{X} \geq A) = 0.79$, $\Pr(Y < C | \tilde{X} \geq A) = 0.79$

分析五：半競爭風險（未截切，未設限）

$F_1(t)$		Weighting	Peng & Fine
0.1	Average bias	0.000289	0.000289
	(Std)	(0.010012)	(0.010012)
	MSE	0.000100	0.000100
0.2	Average bias	0.000549	0.000549
	(Std)	(0.013581)	(0.013581)
	MSE	0.000185	0.000185
0.3	Average bias	0.000626	0.000626
	(Std)	(0.015108)	(0.015108)
	MSE	0.000229	0.000229
0.4	Average bias	0.000487	0.000487
	(Std)	(0.015961)	(0.015961)
	MSE	0.000255	0.000250

表 5-21： $\tilde{F}_1^W(t)$ 與 $\tilde{F}_1^{PF}(t)$ 之比較

$\tau = 0.9$, $\Pr(\tilde{Y} \geq A) = 1.0$, $\Pr(X < C | \tilde{Y} \geq A) = 1.0$, $\Pr(Y < C | \tilde{Y} \geq A) = 1.0$

$F_1(t)$		Weighting	Peng & Fine
0.1	Average bias	0.000235	0.000235
	(Std)	(0.009602)	(0.009602)
	MSE	0.000092	0.000092
0.2	Average bias	0.000235	0.000235
	(Std)	(0.012967)	(0.012967)
	MSE	0.000168	0.000168
0.3	Average bias	0.000721	0.000721
	(Std)	(0.014833)	(0.014833)
	MSE	0.000221	0.000221
0.4	Average bias	0.000334	0.000334
	(Std)	(0.015883)	(0.015883)
	MSE	0.000252	0.000252

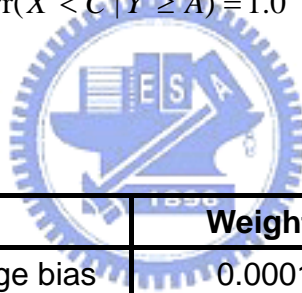
表 5-22： $\tilde{F}_1^W(t)$ 與 $\tilde{F}_1^{PF}(t)$ 之比較

$\tau = 0.7$, $\Pr(\tilde{Y} \geq A) = 1.0$, $\Pr(X < C | \tilde{Y} \geq A) = 1.0$, $\Pr(Y < C | \tilde{Y} \geq A) = 1.0$

$F_1(t)$		Weighting	Peng & Fine
0.1	Average bias	0.000246	0.000246
	(Std)	(0.009401)	(0.009401)
	MSE	0.000088	0.000088
0.2	Average bias	0.000521	0.000521
	(Std)	(0.012734)	(0.012734)
	MSE	0.000162	0.000162
0.3	Average bias	0.000852	0.000852
	(Std)	(0.014779)	(0.014779)
	MSE	0.000219	0.000219
0.4	Average bias	0.000212	0.000212
	(Std)	(0.015574)	(0.015574)
	MSE	0.000243	0.000243

表 5-23 : $\tilde{F}_1^W(t)$ 與 $\tilde{F}_1^{PF}(t)$ 之比較

$\tau = 0.5$, $\Pr(\tilde{Y} \geq A) = 1.0$, $\Pr(X < C | \tilde{Y} \geq A) = 1.0$, $\Pr(Y < C | \tilde{Y} \geq A) = 1.0$



$F_1(t)$		Weighting	Peng & Fine
0.1	Average bias	0.000148	0.000148
	(Std)	(0.009255)	(0.009255)
	MSE	0.000086	0.000086
0.2	Average bias	0.000524	0.000524
	(Std)	(0.012692)	(0.012692)
	MSE	0.000161	0.000161
0.3	Average bias	0.000409	0.000409
	(Std)	(0.014504)	(0.014504)
	MSE	0.000211	0.000211
0.4	Average bias	0.000375	0.000375
	(Std)	(0.015567)	(0.015567)
	MSE	0.000242	0.000242

表 5-24 : $\tilde{F}_1^W(t)$ 與 $\tilde{F}_1^{PF}(t)$ 之比較

$\tau = 0.3$, $\Pr(\tilde{Y} \geq A) = 1.0$, $\Pr(X < C | \tilde{Y} \geq A) = 1.0$, $\Pr(Y < C | \tilde{Y} \geq A) = 1.0$

$F_1(t)$		Weighting	Peng & Fine
0.1	Average bias	0.000134	0.000134
	(Std)	(0.009236)	(0.009236)
	MSE	0.000085	0.000085
0.2	Average bias	0.000203	0.000203
	(Std)	(0.012869)	(0.012869)
	MSE	0.000166	0.000166
0.3	Average bias	0.000275	0.000275
	(Std)	(0.014500)	(0.014500)
	MSE	0.000210	0.000210
0.4	Average bias	0.000029	0.000029
	(Std)	(0.015848)	(0.015848)
	MSE	0.000251	0.000251

表 5-25： $\tilde{F}_1^W(t)$ 與 $\tilde{F}_1^{PF}(t)$ 之比較

$\tau = 0.1$, $\Pr(\tilde{Y} \geq A) = 1.0$, $\Pr(X < C | \tilde{Y} \geq A) = 1.0$, $\Pr(Y < C | \tilde{Y} \geq A) = 1.0$

分析六：競爭風險（未截切，右設限）

$F_1(t)$		Weighting	Peng & Fine
0.1	Average bias	0.000309	0.000328
	(Std)	(0.010030)	(0.010090)
	MSE	0.000101	0.000102
0.2	Average bias	0.000645	0.000651
	(Std)	(0.013676)	(0.013790)
	MSE	0.000187	0.000191
0.3	Average bias	0.000811	0.000832
	(Std)	(0.015571)	(0.015571)
	MSE	0.000243	0.000243
0.4	Average bias	0.000779	0.000738
	(Std)	(0.017019)	(0.016976)
	MSE	0.000290	0.000289

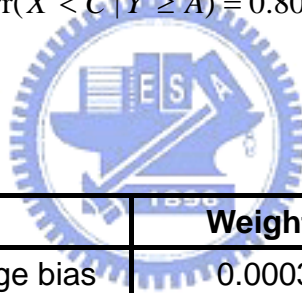
表 5-26： $\tilde{F}_1^W(t)$ 與 $\tilde{F}_1^{PF}(t)$ 之比較

$\tau = 0.9$, $\Pr(\tilde{Y} \geq A) = 1.0$, $\Pr(X < C | \tilde{Y} \geq A) = 0.80$, $\Pr(Y < C | \tilde{Y} \geq A) = 0.80$

$F_1(t)$		Weighting	Peng & Fine
0.1	Average bias	0.000257	0.000293
	(Std)	(0.009649)	(0.009855)
	MSE	0.000093	0.000097
0.2	Average bias	0.000377	0.000416
	(Std)	(0.013028)	(0.013196)
	MSE	0.000170	0.000174
0.3	Average bias	0.000918	0.000945
	(Std)	(0.015265)	(0.015314)
	MSE	0.000234	0.000235
0.4	Average bias	0.000391	0.000391
	(Std)	(0.016897)	(0.017094)
	MSE	0.000286	0.000292

表 5-27 : $\tilde{F}_1^W(t)$ 與 $\tilde{F}_1^{PF}(t)$ 之比較

$\tau = 0.7$, $\Pr(\tilde{Y} \geq A) = 1.0$, $\Pr(X < C | \tilde{Y} \geq A) = 0.80$, $\Pr(Y < C | \tilde{Y} \geq A) = 0.80$



$F_1(t)$		Weighting	Peng & Fine
0.1	Average bias	0.000307	0.000314
	(Std)	(0.009393)	(0.009750)
	MSE	0.000088	0.000095
0.2	Average bias	0.000631	0.000701
	(Std)	(0.012754)	(0.012920)
	MSE	0.000163	0.000167
0.3	Average bias	0.000935	0.000982
	(Std)	(0.015053)	(0.015588)
	MSE	0.000227	0.000244
0.4	Average bias	0.000234	0.000431
	(Std)	(0.016201)	(0.016925)
	MSE	0.000263	0.000287

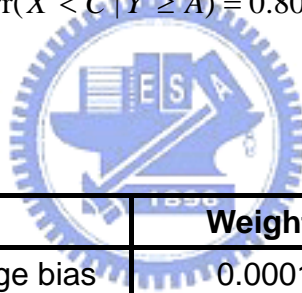
表 5-28 : $\tilde{F}_1^W(t)$ 與 $\tilde{F}_1^{PF}(t)$ 之比較

$\tau = 0.5$, $\Pr(\tilde{Y} \geq A) = 1.0$, $\Pr(X < C | \tilde{Y} \geq A) = 0.80$, $\Pr(Y < C | \tilde{Y} \geq A) = 0.80$

$F_1(t)$		Weighting	Peng & Fine
0.1	Average bias	0.000151	0.000167
	(Std)	(0.009246)	(0.010029)
	MSE	0.000086	0.000101
0.2	Average bias	0.000589	0.000559
	(Std)	(0.012683)	(0.013646)
	MSE	0.000161	0.000187
0.3	Average bias	0.000453	0.000788
	(Std)	(0.014846)	(0.016011)
	MSE	0.000221	0.000257
0.4	Average bias	0.000350	0.000707
	(Std)	(0.016588)	(0.017887)
	MSE	0.000275	0.000320

表 5-29 : $\tilde{F}_1^W(t)$ 與 $\tilde{F}_1^{PF}(t)$ 之比較

$\tau = 0.3$, $\Pr(\tilde{Y} \geq A) = 1.0$, $\Pr(X < C | \tilde{Y} \geq A) = 0.80$, $\Pr(Y < C | \tilde{Y} \geq A) = 0.80$



$F_1(t)$		Weighting	Peng & Fine
0.1	Average bias	0.000148	0.000333
	(Std)	(0.009264)	(0.010958)
	MSE	0.000086	0.000120
0.2	Average bias	0.000279	0.000821
	(Std)	(0.012947)	(0.014916)
	MSE	0.000168	0.000223
0.3	Average bias	0.000252	0.000810
	(Std)	(0.014699)	(0.017474)
	MSE	0.000216	0.000306
0.4	Average bias	0.000110	0.000351
	(Std)	(0.016584)	(0.019358)
	MSE	0.000275	0.000370

表 5-30 : $\tilde{F}_1^W(t)$ 與 $\tilde{F}_1^{PF}(t)$ 之比較

$\tau = 0.1$, $\Pr(\tilde{Y} \geq A) = 1.0$, $\Pr(X < C | \tilde{Y} \geq A) = 0.80$, $\Pr(Y < C | \tilde{Y} \geq A) = 0.80$

分析七：半競爭風險（左截切，未設限）

$F_1(t)$		Weighting	Peng & Fine
0.1	Average bias	0.000054	-0.000336
	(Std)	(0.032866)	(0.033192)
	MSE	0.001080	0.001102
0.2	Average bias	0.000238	-0.000453
	(Std)	(0.034752)	(0.036357)
	MSE	0.001208	0.001322
0.3	Average bias	0.000652	-0.000321
	(Std)	(0.035125)	(0.038727)
	MSE	0.001234	0.001500
0.4	Average bias	0.000724	-0.000514
	(Std)	(0.035919)	(0.042045)
	MSE	0.001291	0.001768

表 5-31： $\tilde{F}_1^W(t)$ 與 $\tilde{F}_1^{PF}(t)$ 之比較

$\tau = 0.9$, $\Pr(\tilde{Y} \geq A) = 0.63$, $\Pr(X < C | \tilde{Y} \geq A) = 1.0$, $\Pr(Y < C | \tilde{Y} \geq A) = 1.0$

$F_1(t)$		Weighting	Peng & Fine
0.1	Average bias	0.000407	-0.000035
	(Std)	(0.027806)	(0.028170)
	MSE	0.000773	0.000794
0.2	Average bias	0.000600	-0.000146
	(Std)	(0.030617)	(0.032386)
	MSE	0.000938	0.001049
0.3	Average bias	0.001623	0.000604
	(Std)	(0.032503)	(0.036363)
	MSE	0.001059	0.001323
0.4	Average bias	0.001461	0.000154
	(Std)	(0.035060)	(0.041316)
	MSE	0.001231	0.001707

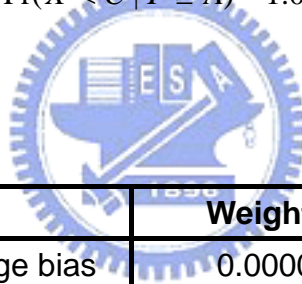
表 5-32： $\tilde{F}_1^W(t)$ 與 $\tilde{F}_1^{PF}(t)$ 之比較

$\tau = 0.7$, $\Pr(\tilde{Y} \geq A) = 0.63$, $\Pr(X < C | \tilde{Y} \geq A) = 1.0$, $\Pr(Y < C | \tilde{Y} \geq A) = 1.0$

$F_1(t)$		Weighting	Peng & Fine
0.1	Average bias	0.000723	0.000390
	(Std)	(0.022174)	(0.022767)
	MSE	0.000492	0.000518
0.2	Average bias	0.001013	0.000389
	(Std)	(0.026535)	(0.028670)
	MSE	0.000705	0.000822
0.3	Average bias	0.001829	0.000918
	(Std)	(0.029964)	(0.034196)
	MSE	0.000901	0.001170
0.4	Average bias	0.001443	0.000231
	(Std)	(0.033955)	(0.040433)
	MSE	0.001155	0.001635

表 5-33 : $\tilde{F}_1^W(t)$ 與 $\tilde{F}_1^{PF}(t)$ 之比較

$\tau = 0.5$, $\Pr(\tilde{Y} \geq A) = 0.63$, $\Pr(X < C | \tilde{Y} \geq A) = 1.0$, $\Pr(Y < C | \tilde{Y} \geq A) = 1.0$



$F_1(t)$		Weighting	Peng & Fine
0.1	Average bias	0.000073	-0.000242
	(Std)	(0.019529)	(0.020221)
	MSE	0.000381	0.000409
0.2	Average bias	0.000640	0.000038
	(Std)	(0.024962)	(0.027193)
	MSE	0.000624	0.000739
0.3	Average bias	0.000899	-0.000008
	(Std)	(0.028856)	(0.033211)
	MSE	0.000833	0.001103
0.4	Average bias	0.001110	-0.000093
	(Std)	(0.033568)	(0.040102)
	MSE	0.001128	0.001608

表 5-34 : $\tilde{F}_1^W(t)$ 與 $\tilde{F}_1^{PF}(t)$ 之比較

$\tau = 0.3$, $\Pr(\tilde{Y} \geq A) = 0.63$, $\Pr(X < C | \tilde{Y} \geq A) = 1.0$, $\Pr(Y < C | \tilde{Y} \geq A) = 1.0$

$F_1(t)$		Weighting	Peng & Fine
0.1	Average bias	-0.000730	-0.000103
	(Std)	(0.017426)	(0.018588)
	MSE	0.000304	0.000346
0.2	Average bias	0.000656	-0.000163
	(Std)	(0.022669)	(0.025828)
	MSE	0.000514	0.000667
0.3	Average bias	0.001036	-0.000161
	(Std)	(0.027717)	(0.032403)
	MSE	0.000769	0.001050
0.4	Average bias	0.001955	-0.000344
	(Std)	(0.032549)	(0.040067)
	MSE	0.001063	0.001610

表 5-35： $\tilde{F}_1^W(t)$ 與 $\tilde{F}_1^{PF}(t)$ 之比較

$\tau = 0.1$, $\Pr(\tilde{Y} \geq A) = 0.63$, $\Pr(X < C | \tilde{Y} \geq A) = 1.0$, $\Pr(Y < C | \tilde{Y} \geq A) = 1.0$

分析八：半競爭風險（左截切，右設限）

$F_1(t)$		Weighting	Peng & Fine
0.1	Average bias	-0.005828	-0.000283
	(Std)	(0.031319)	(0.033405)
	MSE	0.001015	0.001116
0.2	Average bias	-0.012955	-0.000396
	(Std)	(0.033347)	(0.036730)
	MSE	0.001280	0.001349
0.3	Average bias	-0.019889	-0.000200
	(Std)	(0.034000)	(0.039143)
	MSE	0.001552	0.001532
0.4	Average bias	-0.025923	-0.000415
	(Std)	(0.034982)	(0.042376)
	MSE	0.001896	0.001796

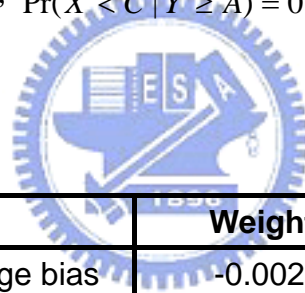
表 5-36： $\tilde{F}_1^W(t)$ 與 $\tilde{F}_1^{PF}(t)$ 之比較

$\tau = 0.9$, $\Pr(\tilde{Y} \geq A) = 0.63$, $\Pr(X < C | \tilde{Y} \geq A) = 0.80$, $\Pr(Y < C | \tilde{Y} \geq A) = 0.80$

$F_1(t)$		Weighting	Peng & Fine
0.1	Average bias	-0.004162	-0.000065
	(Std)	(0.027053)	(0.028326)
	MSE	0.000749	0.000802
0.2	Average bias	-0.009673	-0.000074
	(Std)	(0.030015)	(0.032606)
	MSE	0.000994	0.001063
0.3	Average bias	-0.015080	0.000714
	(Std)	(0.032100)	(0.036648)
	MSE	0.001258	0.001344
0.4	Average bias	-0.021368	0.000209
	(Std)	(0.034724)	(0.041676)
	MSE	0.001662	0.001737

表 5-37 : $\tilde{F}_1^W(t)$ 與 $\tilde{F}_1^{PF}(t)$ 之比較

$\tau = 0.7$, $\Pr(\tilde{Y} \geq A) = 0.63$, $\Pr(X < C | \tilde{Y} \geq A) = 0.80$, $\Pr(Y < C | \tilde{Y} \geq A) = 0.80$



$F_1(t)$		Weighting	Peng & Fine
0.1	Average bias	-0.002657	0.000404
	(Std)	(0.021624)	(0.022981)
	MSE	0.000475	0.000528
0.2	Average bias	-0.006470	0.000489
	(Std)	(0.025788)	(0.028819)
	MSE	0.000707	0.000831
0.3	Average bias	-0.011097	0.000926
	(Std)	(0.029132)	(0.034520)
	MSE	0.000850	0.001192
0.4	Average bias	-0.017411	0.000236
	(Std)	(0.033082)	(0.040899)
	MSE	0.001398	0.001673

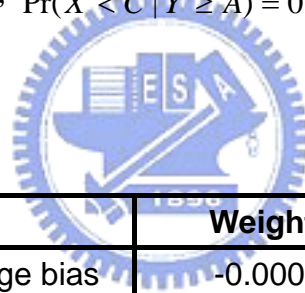
表 5-38 : $\tilde{F}_1^W(t)$ 與 $\tilde{F}_1^{PF}(t)$ 之比較

$\tau = 0.5$, $\Pr(\tilde{Y} \geq A) = 0.63$, $\Pr(X < C | \tilde{Y} \geq A) = 0.81$, $\Pr(Y < C | \tilde{Y} \geq A) = 0.80$

$F_1(t)$		Weighting	Peng & Fine
0.1	Average bias	-0.001933	-0.000277
	(Std)	(0.019367)	(0.020517)
	MSE	0.000379	0.000421
0.2	Average bias	-0.004701	0.000079
	(Std)	(0.024612)	(0.027556)
	MSE	0.000628	0.000759
0.3	Average bias	-0.008923	0.000113
	(Std)	(0.028489)	(0.033692)
	MSE	0.000891	0.001135
0.4	Average bias	-0.014449	0.000134
	(Std)	(0.033138)	(0.040767)
	MSE	0.001307	0.001662

表 5-39 : $\tilde{F}_1^W(t)$ 與 $\tilde{F}_1^{PF}(t)$ 之比較

$\tau = 0.3$, $\Pr(\tilde{Y} \geq A) = 0.63$, $\Pr(X < C | \tilde{Y} \geq A) = 0.83$, $\Pr(Y < C | \tilde{Y} \geq A) = 0.80$



$F_1(t)$		Weighting	Peng & Fine
0.1	Average bias	-0.000680	-0.000198
	(Std)	(0.018046)	(0.018998)
	MSE	0.000326	0.000361
0.2	Average bias	-0.003090	0.000120
	(Std)	(0.023614)	(0.026668)
	MSE	0.000567	0.000711
0.3	Average bias	-0.006762	-0.000105
	(Std)	(0.027940)	(0.033415)
	MSE	0.000826	0.001117
0.4	Average bias	-0.011978	-0.000329
	(Std)	(0.033187)	(0.041131)
	MSE	0.001105	0.001690

表 5-40 : $\tilde{F}_1^W(t)$ 與 $\tilde{F}_1^{PF}(t)$ 之比較

$\tau = 0.1$, $\Pr(\tilde{Y} \geq A) = 0.63$, $\Pr(X < C | \tilde{Y} \geq A) = 0.85$, $\Pr(Y < C | \tilde{Y} \geq A) = 0.80$

第六章 結論

論文的原始目標是研究痴呆症 dementia 發病年齡的機率分佈。一般而言這是一種老年才會發作的疾病。為了減少抽樣的成本，往往只截取某些歲數以上的個體進入樣本中，也造成資料的截切。此外除了外生的設限原因（個體失聯或是研究結束）會造成發病年齡未能完整被記錄，“死亡”亦是發病事件的競爭風險，因為有一部份的個體終其一生都不會發病。我們以“半競爭風險與截切”的架構分析這樣的資料型態，並把目標鎖定於“累積發生函數”的估計。

原以為我們是第一個嘗試分析“半競爭風險與截切”資料的研究團體，在論文的進行過程卻發現 Peng & Fine (2006) 亦從事類似的研究，我們在網路上先取得他們的論文的草稿進行比較，發現他們是以“分解法”的技巧做為推論的基礎。我們的方法將 Chen et al. (2006) 的論文予以延伸，以“加權法”的技巧提出新的估計量，並透過模擬實驗比較這兩種不同估計方法之優劣。值得一提的是我們所提出權重的估計法使用了比 Peng & Fine (2006) 論文更強的條件（可以完整觀測到 C 或是 $A \perp B$ ），或許因此我們的估計量在許多模擬中表現得更穩定。因為論文時間有限，無法繼續投入精力去推導更具一般性的估計方法，若未來將論文投稿勢必要對權數估計問題做進一步分析與改進。

經過基本運算推導，發現在沒有截切且沒有外來設限的情況下，半競爭風險與競爭風險資料對估計累積發生函數會得到相同的無母數估計式。當加入設限的條件後，模擬實驗的結果顯示半競爭風險估計量與競爭風險估計量表現各有好壞，不分軒輊；但在左截切右設限的條件下，半競爭風險資料的確對估計有所幫助。

雖然我們希望建立了方法論後可以藉此分析 Cache 資料，但由於所獲得之資料並未包含任何死亡的資訊，使得立即做資料分析的目標無法達成。原因之一是我們間接取得經其他學者整理過的資料，而非原始資料。可能分析這筆資料的學者當時並不需要死亡的資訊，以致我們無法繼續利用。我們打算未來由原始資

料的擁有者處再尋求獲得更完整記錄資料。此外，Cache data 仍包含親屬間的資料，在我們的研究中只限於 proband 資料的分析，未來可以往家族關聯性與遺傳的方向做更進一步的探討。



參考文獻

1. Chang, S. H.(2000). A Two-Sample Comparison for Multiple Ordered Event Data. *Biometrics*, **56**, 183-189.
2. Chang, S. H. and Tzeng, S.J.(2006) Nonparametric estimation of Sojourn time distributions for truncated serial event data – A weight-adjusted approach.
3. Chen, C.H. , Chang, W.H. and Wang, W.(2006). Estimation of the cumulative incidence function for multiple events data.
4. Day, R., Bryant, J. and Lefkopoulou, M. (1997). Adaptation of Bivariate Frailty Models for Prediction , with Application to Biological Markers as Prognostic Indicators. *Biometrika*, **84**,45-56.
5. Fine, J. P. , Jiang, H. and Chappell, R. (2001). On Semi-Competing Risks Data. *Biometrika*, **88**,4,907-919.
6. Jiang, H. Fine, J and Chappell, R. (2004). Semiparametric methods for semi-competing risks problem with censoring and truncation
7. Kaplan, E.L. and Meier, P. (1958) Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*.
53,457-481
8. Klein, J. P. and Moeschberger, M. L. (2002)Survival analysis techniques for censored and truncated data. 2nd
9. Lai, T. L. and Ying, Z.(1991) Estimating a distribution function with truncated and censored data. *Annals of Statistics* ,**19** , 417-442
10. Lin, D. Y., Robins, J. M. and Wei, L. J. (1996). Comparing Two Failure Time Distribution in The Presence of Dependent Censoring. *Biometrika*,
83,2,381-393

11. Lynden-Bell, D. (1971). A method of allowing for known observational selection in small samples applied to 3CR quasars. *Monthly Notices Roy. Astronom. Soc.* **155**, 95-188
12. Peng, L. and Fine, J. P. (2006). Nonparametric estimation with left truncated semi-competing risks data. **93**, 2, 367-383
13. Wang, W. (2003). Estimating the Association Parameter for Copula Models under Dependent Censoring. *Journal of the Royal Statistical Society, Series B.* **65**, 257-273.

