

## 數 理 組

---



姓名：王維菁

學歷：

台灣大學工商管理系畢業 (1988)

馬里蘭大學企管碩士 (1991)

康乃爾大學統計學博士 (1995)

現職及經歷：

北卡州立大學統計系助理教授 (1995-1996)

中研院統計所助研究員 (1996-1999)

轉任交通大學擔任教職 (1999- 迄今)



著作名稱：

1. Wang, W. & Ding, A. A. (2000). "On Assessing the Association for Bivariate Current Status Data." *Biometrika*, 87, 879-893.
2. Wang, W. (2003). "Estimating the Association Parameter for Copula Models under Dependent Censoring". *Journal of the Royal Statistical Society, Series B*. 65, 257-273.

中文簡介：

### 一、存活分析背景簡介

統計領域的存活分析(survival analysis)主要探討存活變數的推論方法，這個變數所量測的是一段時間的長度 – 由起點到某個“事件”所發生時間長(time to an event of interest)，例如“壽命”代表由出生到死亡、“發病年齡”代表由出生到發病…。一般統計方法的基本假設是資料為變數的隨機樣

本，存活資料最大的特色是當事件發生的時間無法確切得知時，資料中會包含不完整的觀測值，換言無法取得感興趣變數的隨機樣本。

統計文獻自1980起就經常出現多變量存活分析的論文，我的研究亦屬此範疇。多維度的存活分析同時探討多個存活變數，主要研究是變數間的關聯性。例如遺傳學家可以藉著探討雙生子壽命間的關聯性強弱以了解遺傳（或是共享之環境因子）對壽命的影響程度。量化的分析可以幫助科學家有系統的討論感興趣的問題，統計學家對這些有意義但是未知的量（稱之為母體參數）提出推論方法。以雙生子的壽命為例，兩者關聯性強代表人類壽命受到遺傳因子的影響程度高。有個遺傳學的假說是“高壽”可能深受遺傳的影響，這個假說可藉著對雙胞胎的壽命長度做關聯性分析，比較一般壽命的群體和高壽的群體是否在關聯性強度上有所不同予以間接的驗證。當變數的維度增加時，不完整的資料型態變化更為多樣，分析也更加複雜。我研究的角度是針對不同資料型態的特色，建構統計方法以探討和相關性分析有關的問題。

## 二、早期研究成果

我的博士論文和剛畢業後的研究題材主要針對二維度右設限的資料，研究內容涵蓋了無母數分析、模式選取、相關參數估計…等問題，請見參考資料(1)-(3)。在這個相對比較單純的資料型態下，我所涉獵的推論問題

涵蓋面甚廣，也藉此得以拓展自己對此領域的了解。在論文(4)我針對一種特殊的設限結構（常發生於紀錄的存活變數在測量時有先後順序時，例如由HIV病毒感染到發生愛滋病的“潛伏期長度”與由發生愛滋病到死亡的“罹病期長度”），提出修正偏誤的統計方法。

## 三、代表作簡介

### (1)型一區間設限資料的推論問題

我於中研院服務期間(1996-1999)曾參與生醫所潘文涵教授主持的心臟血管疾病流行病研究計劃。原本構想是將現有的推論方法直接套用於資料的分析，卻發現這個計劃的資料型態並不適用於任何現成的方法。舉例來說我們希望探討心臟病和糖尿病的發病年齡是否具有關聯性，資料卻只包含受測者“當時的年齡”與“受測時帶病與否”的資訊。這種資料所含的資訊非常少，稱之為型一區間設限資料，常見於大型流行病學的調查研究—因為規模龐大，受限於經費或是變數本身的特質，無法追蹤到事件確實發生的時間。舊有文獻的研究以單維度的結果為主，特別之處是因為資料本身太過粗造，所以無母數估計量的收斂速率比一般統計量要來得慢（非 $n^{1/2}$ 的收斂速率而是 $n^{1/3}$ ）。我因為過去的研究背景，想嚐試文獻上最早處理二維度此種型態資料的問題。我和過去在康乃爾大學就讀的同學Ding針對這個方向合作了兩篇論文（論文5,6），討論兩個發病時間變數關聯性參數的估計問題和獨立性的檢定，在證明過程中我們克服了前述非典型收斂速度的問題。

## (2) 半競爭風險資料的推論問題

許多醫學的問題病程發展往往有多重路徑。例如接受骨髓移植後可能經歷 transplant → recurrence → death 或是 transplant → death 兩條路徑，文獻稱為“illness-death model”。早期的文獻對此類資料的推論方法多是以隨機過程的 Markov 或是 semi-Markov 模式為基礎，然而近年來多變量存活分析的發展亦帶來新的思維，所以這個方向又重新受到學者注意。我個人亦以此角度出發，認為有兩個特質值得深入探討。第一是“死亡”是“復發”的競爭風險但是“復發”卻非“死亡”的競爭風險，我在論文(7)針對這種非對等的競爭關係下提出估計兩個變數的關聯性參數的方法。我提出的估計函數的建構方法，一是利用計數函數(counting process)與 martingale 等式的關係做為估計函數的基礎。另一方法是利用二項列聯表(two-by-two table)的性質，兩個概念都具一般性可以推廣到其它的問題。第二個想法是考慮“受到死亡所設限的個體”(即死前未經復發者)和“免疫”的關聯性。例如在死亡前未罹患某個疾病的人可否說他們是“免疫”？若否，需要怎樣的假設以區分二者？此時發現存活分析已存在一個次領域稱之為 cure model，容許存在永不會發生事件的免疫者(或稱 long-term survivor、non-susceptibility)。我所採取的角度有別於傳統的 cure model 在於我將來自死亡的競爭風險納入考量，論文(8)包含部份結果。

## 四、結 論

我目前的研究興趣在於以雙維度或是多維度的 cure model 出發，藉此探討複雜遺傳疾病(如乳癌、精神分裂…)的家族群聚性，目前做法是將“罹病與否”和“發病年齡”的群聚性以 mixture model 型式分開討論，希望對科學發展有所助益。

## 評審簡評：

王維菁女士的研究主題是存活分析與競爭風險模型。她與Ding合作的論文(Ding是她的同學)把Current State data 拓展到高維度的情形。他們估計參數估計量仍具有 $\sqrt{n}$ 的收斂速度，即使在Likelihood function 中代入marginal functions的NPMLE時只有 $\sqrt[3]{n}$ 的收斂速率。在科學上這是極有貢獻的結果，一位評審人說：王維菁近年已成為多變量統計存活分析的領導人之一。在實際應用上，中研院生醫所潘文涵已使用她們的結果來分析高血壓與糖尿病發病時間是否有關聯。

在競爭風險研究方面，她提出一個新的統計推論方法，以取代傳統以概似函數為基礎的方法。她在競爭風險模型的研究，為這領域提供一種統一且具有突破性的方法，這種方法將廣為公衛學者所採用。