

出國報告（出國類別：出席國際會議）

第二屆亞太地區數理統計學會會議
(The 2nd Institute of Mathematical
Statistics Asia Pacific Rim
Meeting)

服務機關：國立中興大學

姓名職稱：黃文瀚 教授

派赴國家：日本，茨城縣筑波市

出國期間：民國 101 年 7 月 1 日至 7 月 4 日

報告日期：民國 101 年 7 月 17 日

摘要

數理統計學會 (IMS) 是目前統計學界最重要的學會之一，2012 年之亞太地區的數理統計學會會議 (IMS Asia Pacific Rim Meetings) 係屬其分區會議，本次會議地點為日本茨城縣筑波市，舉行日期是 7 月 1 日至 7 月 4 日。本次會議由日本、印度、韓國、台灣統計學會協同合力舉辦，目的即在於推展與提升亞太地區之統計學相關的研究。此次會議主題含括統計之各大領域 (數理統計、生物統計、工業統計、社會及商業統計)，另外更有當前熱門議題的探討，如超大型資料的分析、高維度資料分析、生物資訊等。本次會議也邀請到許多國際著名統計人士到場演說，藉由參加此會議以獲取更多統計知識與經驗，瞭解統計未來其可能的發展趨勢，以提升研究之水準。

目次

一、目的	4
二、行程及研討會報告內容摘述	4
三、心得及建議	8
四、考察參觀活動與見聞	8
五、攜回資料名稱及內容	9
六、活動照片	10

一、目的：

IMS(Institute of Mathematical Statistics) 數理統計學會是國際著名之統計學會，每四年定期舉辦的世界統計年會，此外也會在各地區舉辦分區會議。由於學會所發行的期刊佔有舉足輕重的地位，例如 Annals of Statistics, Annals of Applied Statistics, Annals of Probability, Annals of Applied Probability, Statistical Science 等，因此所舉辦的會議也具有指標性的作用。

此次 2012 年之亞太地區的數理統計學會會議(IMS Asia Pacific Rim Meetings)為此分區會議之第二次舉行，地點為日本茨城縣筑波市(第一次為 2009 年於南韓首爾舉行)。本次會議由日本、印度、韓國、台灣統計學會協同合力舉辦，目的即在於推展與提升亞太地區之統計學相關的研究。內容涵蓋從理論證明到實務應用等不同類型的研究主題，藉由此機會相互交流，提升國際統計學術的研究。本次會議有 500 餘位學者參加，超過 100 個場次的演講，其中更邀請 16 位特別傑出的學者發表演講，如 Peter Hall (U. of Melbourne), S. R. S. Varadhan(New York U.), Adrian Baddeley (U. of Western Australia), Arup Bose (Indian Statistical Institute), Tony Cai (U. of Pennsylvania), Jianqing Fan (Princeton U.)發表關於 Functional data analysis, Nonparametric method, Bayesian MCMC, Complex models, Spatial statistics 之題目,這些都是目前在統計上非常重要的研究主題，也是本人十分感興趣的方向。

二、行程及研討會報告內容摘述

我於 6 月 30 日下午飛抵東京成田機場，隨後轉電聯車入住於東京市暮野里區的簡易商務旅館，隔天再搭乘火車由東京前往會議地點筑波市，並赴會場辦理報到事宜。基本上，此會議約有 500 位參予者，算是一個中小型的會議，而參加會

者大多來自中日美韓等地區，會議地點集中於會議中心且所有會議廳都蠻集中的，因此對於選擇不同場次的演講主題很方便。此外，由於會議中心距離入住旅館只有約 1.5 公里，因此開會期間的交通相當便利。7 月 2-4 日之間屬會議正式日期，除少數大會主要演講外，同時間安排有 12 個會議廳討論不同主題的演講，者個會意涵蓋約 300 多個研究發表，所以只能儘量參與個人較感興趣的研究議題。7 月 4 日中午會議結束，我赴抵東京市短暫參訪兩天，並於 7 月 6 日搭機返台。

參加這次會議的主要目的是發表論文，但是除此之外，會議期間，我也聆聽到其他與會學者的研究成果與心得。我聽的主題以函數形資料分析方法與存活分析方法為主，印象較深刻的演講是墨爾本大學 Peter Hall 教授的利用偏最小平方方法處理函數型資料，首爾大學 Yoon-Jae Whang 教授談關於函數型資料分析中的檢定問題，耶魯大學 Xiaohong Chen 教授的資料間具有若相關性時的推論，東京大學 Yasuhiro Omori 教授談論高頻時間序列之隨機波動模型之應用，及密西根大學 Naisyin Wang 教授利用懲罰函數的架構估計微分方程式的參數。藉著這個機會吸收他人的研究成果與經驗，透過不同的激盪啟發個人的一些靈感。

我的演講被安排在 7 月 3 日下午 1:30~3:15，場次是 IP-25，此場次為無母數及半母數統計之方法與應用。本場次係由台灣大學數學系的鄭明燕教授負責規劃，除了本人以外還有清華大學統計所徐南蓉教授，中央研究院統計所副所長邱政民教授，以及鄭明燕教授發表論文，每人有 25 分鐘的時間。我報告的題目是 A semiparametric model for a functional behavioural response to capture in capture-recapture experiments。報告之重點如下：

利用重複捕取模式估計母體總數是生態統計中非常重要的方法，因此關於重複捕取模式的研究也相當豐富。在重複捕取實驗中，捕取機率會因捕取時間 (time effect)、動物個體異質性 (individual heterogeneity) 及行為反應 (behavior response) 等三種因素有相關性。Otis et al. (1978) 以 $M_{t,b,h}$ 代表一模式考慮考慮上述三種因素，其中下標 t, b, h 分別代表時間變化、行為反應及個體差異等影響。若實驗之捕取機率僅與其中某一因素有關，例如捕取時間，則記

為 M_i 。在這些模式中，最難處理也是實際上最重要的因素為個體異質性的影響。Huggins(1989) 假設捕取機率为輔助變數(如體重，年齡，羽翼長或氣候資料)之函數，並利用迴歸模式估計捕取機率及動物總數。二十幾年來，此方法已被廣泛的應用於估計母體總數。相關的研究或推廣也一直持續著，例如 Yip et al. (1996) 推廣至連續型模式，Pledger (2000) 整合潛在變數於模式中，處理觀察到的輔助變數無法解釋的變異，Borchers *et al.* (1998) 將其應用至線上穿越(line transect)模式，Pollock et al. (1995) 應用至無線電追蹤模式，Yip et al. (2002) 推廣應用到開放模式。先前的研究工作多著重於模式推廣的層次。最近幾年來，陸續有研究針對 Huggins 模式的基本假設提出檢驗及修正之方式，這其中包含 Hwang & Huang (2003)，Hwang et al. (2007) 討論數據含有測量誤差時對母體總數估計的影響，Hwang & Huggins (2005) 也証明了忽略輔助變數會導致的影響，另外，Huggins & Hwang (2007) 檢驗參數化迴歸模式指定錯誤的影響，也會影響母體總數估計，並提出採用無母數迴歸方法來加以解決，本研究計劃的主要目的即是在此一方向上作一更加深入的探討。

令 t 表捕取次數，令 p_{ij} 為第 i 隻動物於第 j 次捕取時之被捕機率， Y_{ij} 為 1 或 0 代表其有無被捕獲的情形。我們提出下列模式

$$p_{ij} = \begin{cases} H(\alpha(x_i) + \beta z_i + e_j), & \text{for first captured,} \\ H(\alpha^*(x_i) + \beta z_i + e_j), & \text{for recaptured,} \end{cases}$$

其中 $H(u) = 1/(1 + e^{-u})$ ， e_j 表第 j 次捕取之效應且 $e_1=0$ ， α 與 α^* 為兩未訂形式之平滑函數。這個模式的定義相當廣泛且包含了 Huggins(1989) 之傳統模式，例如取 $\alpha(x) = \alpha_0 + \alpha_1 x$ 及 $\alpha^*(x) = \alpha_0 + \alpha_1 x + \phi$ 便是傳統之 M_{tth} 模式。Huggins and Hwang (2007) 則是僅考慮 $\beta = 0, e_j = 0$ 且 $\alpha = \alpha^*$ 的特殊情形，也就是假設時間變異及行為反應不會影響捕取機率。除此之外，這一模式有一新的特點：其動物行為反應不再限定於常數或等倍率增加/減少，而是會因個體特質有所差異。

我們假設 α 與 α^* 為局部多項式，假設 x 在 x_0 的鄰域時，

$$\alpha(x) = b_0 + b_1 \left(\frac{x-x_0}{h}\right) + \dots + b_k \left(\frac{x-x_0}{h}\right)^k,$$

$$\alpha^*(x) = b_0^* + b_1^* \left(\frac{x-x_0}{h}\right) + \dots + b_k^* \left(\frac{x-x_0}{h}\right)^k,$$

其中 k 為多項式階數且 $h>0$ 為帶寬。令

$$p_{0,ij} = \begin{cases} H(\alpha(x_i) + \beta z_i + e_j), & \text{for first captured,} \\ H(\alpha^*(x_i) + \beta z_i + e_j), & \text{for recaptured,} \end{cases}$$

及

$$L_{0,i} = \sum_{j=1}^l (Y_{ij} \log(p_{0,ij}) + (1 - Y_{ij})(1 - p_{0,ij})) - \log(1 - \prod_{j=1}^l (1 - p_{0,ij})).$$

再令 K 為一核函數， $w_0(x) = K((x - x_0)/h)/h$ ，則可導出局部對數概似函數為

$$L_0 = \sum_{j=1}^l w_0(x_i) L_{0,i} \quad (1)$$

假設 β 及 e_j 已知，即可由(1)求得 α 與 α^* 的估計。另一方面假設 α 與 α^* 已知

我們也可導出 β 及 e_j 的剖面對數概似函數(profile log-likelihood function)

$$L_G = \sum_{j=1}^l L_{G,i}, \quad (2)$$

其中 $L_{G,i} = \sum_{j=1}^l (Y_{ij} \log(\tilde{p}_{ij}) + (1 - Y_{ij})(1 - \tilde{p}_{ij})) - \log(1 - \prod_{j=1}^l (1 - \tilde{p}_{ij}))$ 且 \tilde{p}_{ij} 為將 α 與

α^* 於 x_i 的估計代入 p_{ij} 的值。反覆疊代及運用(1)(2)即可求出所有參數之估計，最後再利用Horvitz-Thompson估計式來估計母體總數。此外，我們也計算估計式之漸近偏差及漸近變異數，並採用Ruppert(1997) 之empirical bias bandwidth selection (EBBS)的方法來選取帶寬。

本研究中除了理論方法外，最後還透過 2 組實際的重複捕取實驗的資料來闡述所提出來的的方法與其應用，在其中一組澳洲袋貂資料分析中，我們發現優勢族群與劣勢族群會有截然不同的行為反應，此結果已為生物學家認同並提出生物學上

所代表之意義，利用新方法可經由數據分析方法而將此現象加以量化。會後，有兩位學者提出一些相關問題以及建議，我也很榮幸能有此機會與國外學者探討這方面的研究。

三、心得及建議

這一場演講的報告內容，雖事前準備且反覆練習多次以期能清楚表達，然臨場時還是有點緊張，尤其在有限時間中，一些細節無法作較完整的陳述。所以，在口頭發表上，如何去蕪存菁當是個人應該努力琢磨的地方。

參與此會議有相當多台灣統計學界人士，大家參加不同議題的演講與討論，而且台灣學界也負責籌畫數場的演講，此代表臺灣統計學術研究在亞太地區已具舉足輕重的地位，其發展已邁向國際化及多元化，這是非常可喜的現象。

本研討會讓我及許多與會認知更多的統計領域及其應用，在此感謝學校的補助。期望學校應多加鼓勵老師與博士班研究生參與此類型國際會議，例如：提高補助金額或簡化申請流程，致力透過推廣國際活動和促進合作，以發展及改善統計方法及其應用，以達學術交流，培育出統計專業人員之間共同價值理念，並進一步培養學生之國際觀。

四、考察參觀活動與見聞

由於第一次訪問日本，本人特別利用會後兩天停留於東京市區參觀訪問，行程包括東京都廳、早稻田大學、東京大學等。個人對於東京地區備有廣大的人口數但地上交通並不會覺得壅擠，感到十分讚嘆，其原因在於東京都公共運輸之效能、民眾習慣、與交通守法觀念等。期待我們也能朝此傑出公共運輸模式方向發展，達到節能減碳目標。

此次會議於筑波市舉行，筑波市又稱筑波研究學園城市，位於茨城縣的南部，

是日本的科學研究重鎮之一，許多重要的氣象，宇宙航空研究開發，及產業技術綜合研究所等研究機構均設於此。由於此市為一個計畫城市，因此區內的公路或鐵路網、下水道、公園及其他建築皆以棋盤狀方式建構。所以市容相當整齊優雅，在此地點開會令人感到心曠神怡也帶有幾分恬淡。這樣的研究環境及計畫都市或許也是台灣可思考的方向。

在此會議期間，聽了許多統計學家提出的新看法與突破，感受到學術研究的快速與蓬勃進展。此外，我也和參與之學者進行討論交流，從中獲得許多新的想法，此行收穫良多、不虛此行。如果沒有這樣的刺激與交流，而只是關起門來做研究，很有可能已是井底之蛙而不自知，因此參予相關的國際學術交流會議是非常重要的活動。

五、攜回資料名稱及內容

本次與會攜回帶回大會所發的會議議程書，當中提供演講者的論文摘要，內容有記載每位作者的服務地區及電子郵件等，以利我們會議結束後能夠繼續交流。

六、活動照片

編號 1. 攝於 101 年 7 月 3 日, 照片內容簡述: 我的演說 1



編號 2. 攝於 101 年 7 月 3 日, 照片內容簡述: 我的演說 2



References

- Borchers, D. L., Zucchini, W., and Fewster, R. M. (1998). Mark-recapture models for line transect surveys. *Biometrics* **54**, 1207–1220.
- Huggins, R. M. (1989). On the statistical analysis of capture experiments. *Biometrika* **47**, 725-732. 26
- Huggins, R.M. and Yip, P. (1999). Estimation of the size of an open population from capture-recapture using weighted martingale methods. *Biometrics* **55**, 387–395.
- Huggins, R.M. and Hwang, W-H. (2007). Nonparametric estimation of population size from capture-recapture data when capture probabilities depend on a covariate. *JRSS (C)*.
- Hwang W-H., and Huang S.Y.H. (2003). Estimation in Capture-recapture Models when Covariates are Subject to Measurement Errors . *Biometrics* **59**, 1115-1124.
- Hwang W-H., Huang S.Y.H., and Wang C. Y. (2007). Effects of Measurement Error and Conditional Score Estimation in Capture-Recapture Models. *Statistica Sinica*.
- Hwang W-H and Huggins R. (2005). An examination of the effect of heterogeneity on the estimation of population size using capture-recapture data. *Biometrika* **92**, 229-233.
- Pledger, S. (2000). Unified maximum likelihood estimates for closed capture-recapture models using mixtures, *Biometrics* **56**, 443–450.
- Pollock, K. H., Bunck, C. M., Winterstein, S. R. & Chen, C. L. (1995) A capture-recapture survival analysis model for radio-tagged animals, *Journal of Applied Statistics*, **22**, pp. 661 - 672.
- Ruppert, D. (1997). Empirical-bias bandwidths for local polynomial nonparametric regression and density estimation. *Journal of the American Statistical Association* **92**, 1049-1062.

Yip, P., Huggins, R. M., and Lin, D. Y. (1996), "Inference for Capture–Recapture Experiments in Continuous Time With Variable Capture Rates," *Biometrika*, 83, 477–483.

Yip, P. S. F., Wan, E. C. Y. & Chan, K. S. (2002) Robust design and use of auxiliary variables in estimating population size for an open population, *Biometrics*, 58, 852-861.