行政院國家科學委員會專題研究計畫成果報告

動態族群數量模式考慮觀測誤差的參數估計

（Parametric Estimation for Dynamic Population Model with Observation Errors）

計畫類別：個別型計畫

計畫編號：NSC90-2118-M032-006

執行期間：90 年 8 月 1 日至 91 年 7 月 31 日

計畫主持人：陳主智

計畫參與人員：陳主智、施銘權、黃彥龍、范漢君

執行單位：淡江大學數學系

中華民國　91　年　12　月　25　日

一、中文摘要

本研究討論動態族群數量模式在有觀測誤差時的參數估計方法，並比較不同的資料填補的方法在處理此問題的結果的異同。我們主要比較馬可夫鏈模擬（MCMC）法以及 EM 程序法，並提出綜合兩者的 MCM 程序法在模擬結果上的差異，並加以說明可能的原因。主要的發現如下：EM 法由於單調收斂的性質不成立，以致估計的結果非常不穩定；MCMC 法一般而言表現不錯；MCM 法在估計過程誤差參數有很良好的表現。

關鍵詞：動態族群模式、資料填補、馬可夫鍊模擬、EM 程序、MCM 程序

ABSTRACT. Statistical imputation methods such as the Markov chain Monte Carlo (MCMC) simulation technique and the EM algorithm have been popular in recent years when a portion of the observations are missing due to experimental constraints or the data structure. In this paper, we compare differences in these approaches through studying an ecological parameter estimation problem when observational data consist of both process error and observation error. Assuming that, with initial population size N_0, the logarithm of the population size $N(t)$ at time t approximately follows a Wiener process. Also, the exact as well as estimated total population sizes from the sampling are available for the first few observations. The MCMC procedure and the *EM* algorithm are applied to estimate the population parameters. An algorithm called Markov chain maximization (MCM), which is a hybrid of the MCMC and the EM algorithm, is also proposed. The monotonic convergence of the EM algorithm failed in this particular setting, and may yield very unstable estimates as a result. The MCMC method generally performed well in simulation studies. The MCM algorithm was substantially better than the EM algorithm, and performed exceptionally well in estimating the process error parameter. But it had similarly poor performance to that of the EM algorithm in the observation error estimate. It should be noted for ecological studies on population growth, that long periods of observations with errors

might have an accumulative effect on parameter estimates.

Keywords: Deterministic model; EM algorithm; Gibbs sampling; Markov chain maximization; Markov chain Monte Carlo simulation; Process error

二、緣由與目的

在探討估計某特定族群數量時，一種常被使用的方法為重複捕取（capture-recapture）的方式。此方法可用來估計在特定時間的正確族群數量，然而，若考慮歷年的族群數量，則在生態學研究中常用的方法為建立族群數量成長的模式，並藉由已知數據估計模式參數，以達到預測數量消長的目的。然而由於族群數量的歷年資料為一時間序列，而除了考慮數量的隨機過程模式的誤差外，由於每一時間點的數據都無法精確估計，因此，會有觀測誤差的存在，並且此誤差是有加乘性的。這是在處理此類問題時，估計模式參數所必須注意的地方。

本研究主要目的即是基於存在觀測誤差的考量，假設數量成長模式為對數常態分布，討論不同的處理方式的影響，以提供生態學者在處理相關問題時的因應之道。

三、結果與討論

我們的模擬研究結果發現，不同的參數估計方法的確會造成參數估計正確性有很大的出入。尤其是眾所周知的 EM 程序法，在處理本問題時，由於單調收斂性質並不成立，可能會得到完全錯誤的答案而不自知。本研究也發現，我們所提出的 EM 程序法與 MCMC 程序法混合的 MCM 程序法，在處理此問題時有不錯的表現。

相關問題在動態時間序列的數據分析與建模，經常會面臨到，例如主持人今年度的研究計畫有關人體藥物動力模式的參數估計也面臨同樣但更複雜的問題。因此，有待於進一步的探討，以求得完善的解決之道與研究的一貫性。

四、計畫成果自評

本研究成果已寫成研究論文送審，目前得到 referee 的回應是這一類的問題事實上在時間序列的有關於 dynamic linear model 利用 Kalman filter 的方法解決已有詳盡的發展，然而本文並未提及，是為一大缺失。由於相關研究領域對主持人本身為一全新的領域，未來我們也將朝此方向努力補充改進。

另外，以上的研究成果局限於模擬結果，較深入的理論性質也有待繼續研究。

五、參考文獻

Carpenter, S. R., Cottingham, K. L., and Stow, C. A. (1994). Fitting predator-prey models to time series with observation errors. *Ecology* **75**(5), 1254-1264.

Celeux, G. and Diebolt, J. (1985). The SEM algorithm: a probabilistic teacher algorithm derived from the EM algorithm for the mixture problem. *Comput. Stat. Quart.*, **2**, 73-82.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *J. R. Statist. Soc.* B, **39**, 1-38.

Dennis, B., Munholland, P. L., and Rubin, D. B. (1991). Estimation of growth and extinction parameters for endangered species. *Ecol. Monogr.*, **61**(2), 115-143.

Diebolt J. and Celeux, G. (1993). Asymptotic properties of a stochastic EM algorithm for estimating mixing proportions. *Comm. Stat. –Stoch. Mod.*, **9**(4), 599-613.

Gause, G. F. (1934). *The struggle for existence*. Williams and Wilkins, Baltimore, MD, USA. (Reprinted in 1964 by Hafner Publishing Company, New York.)

Gilks, W. R., Richardson, S., and Spiegelhalter, D. J. (1996). *Markov Chain Monte Carlo in Practice*. Eds. Chapman & Hall, London, UK.

Gelman, A., and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences (with discussion). *Statist. Sci.*, **7**, 457-511.

Heyde, C. C., and Cohen, J. E. (1985). Confidence intervals for demographic projections based on products of random matrices. *Theor. Popul. Biol.*, **27**, 120-153.

Higgins, K., Hastings, A. Sarvela, J. N., and Botsford, L. W. (1997). Stochastic dynamics and deterministic skeletons: population behavior of Dungeness crab.

Science **276**, 1431-1435.

Ludwig, D., and Walters, C. J. (1989). A robust method for parameter estimation from catch and effort data. *Can. J. Fish. Aquat. Sci.*, **46**, 137-144.

Pascual, M. A., and Kareiva, P. (1996). Predicting the outcome of competition using experimental data: maximum likelihood and Bayesian approaches. *Ecology*, **77**(2), 337-349.

Pielou, E. C. (1977). *Mathematical Ecology*. 2nd edition, New York: Wiley.

Roberts, G. O. (1996). Markov chain concepts related to sampling algorithms. In *Markov Chain Monte Carlo in Practice*, Gilks, Richardson, and Spiegelhalter (eds), 45-54. Chapman & Hall, London, UK.

Tuljapurkar, S. D., and Orzack, S. H. (1980). Population dynamics in variable environments. I. Long-run growth rates and extinction. *Theor. Popul. Biol.*, **18**, 314-342.

Wei, G. C. G., and Tanner, M. A. (1990). A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithms. *J. Am. Statist. Ass.*, **85**, 699-704.

West, M., and Harrison, J. (1997). *Bayesian Forecasting and Dynamic Models*. Springer.