

行政院國家科學委員會補助專題研究計畫成果報告

計畫名稱：族群動態考慮遺傳與環境變異的統計問題探討 II

計畫類別：個別型計畫

計畫編號：NSC 89-2118-M-032-020

執行期間：89 年 8 月 1 日至 90 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、魏珊妮

執行單位：淡江大學數學系

中華民國 91 年 1 月 14 日

一、中文摘要

本研究計畫成果主要考慮當一連串觀測記錄而得的生態族群數量同時具有觀測誤差及過程誤差時的參數估計。若在時間 t 時族群數量為 $N(t)$ ，當時間夠長時，根據中央極限定理，Tuljapurkar and Orzack (1980)以及 Heyde and Cohen (1985)得到以下結論： $X(t) = \log N(t)$ 大致為一常態分布 $N(\log N_0 + \sim t, t^2 t)$ 。根據此結果，假設 $X(t) = \log N(t)$ 為一 Wiener-drift process，且時間序列的前幾項可準確觀測，本計畫比較 EM algorithm，與馬可夫鏈模擬 (MCMC)，半參數馬可夫鏈模擬 (semi-parametric MCMC)。以及所提出的馬可夫鏈最大值(Markov Chain Maximization, MCM)的方法在估計數量成長模式的參數方面的異同。結果發現，一般而言，MCMC 法的表現最穩定，而由於 EM algorithm 的單調收斂性質在此處並不成立，其所得的估計量也最差，至於所提出的馬可夫鏈最大值法，則可以有有效的改進 EM algorithm 在此處的缺陷。

關鍵詞：既定模式、過程誤差、觀測誤差、馬可夫鏈模擬、Em 程序、馬可夫鏈最大值

ABSTRACT

This project considers parameter estimation for dynamic population growth in the presence of both process error and observation error. Based on the results of Tuljapurkar and Orzack (1980) and Heyde and Cohen (1985), the quantity $X(t) = \log N(t)$ will have, as time t becomes large, an approximate normal distribution with a mean of $\log N(0) + \sim t$ and a variance of $t^2 t$, where $N(t)$ is the population size at time t . Based on the above assumption, and that for the first few observations, both the real population size $N(t)$ and the estimated population size $O(t)$ are available, this project compared the population parameters estimates by the EM algorithm, MCMC and semi-parametric MCMC methods, and a proposed Markov Chain Maximization (MCM) method. Simulation results showed that, in general, the estimates by the MCMC methods were most efficient. And since the monotone convergence property does not hold here, its estimates were unstable. The proposed MCM estimates could efficiently improve the EM estimates under the setting.

Keywords : process error, observation error, MCMC, semi-parametric MCMC, EM algorithm, Markov Chain maximization

二、緣由與目的

動態物種族群數量的估計在生態學研究中扮演著極重要的角色，舉凡食物的供應來源與保存生態環境中物種的多樣化等，均需要對某些動物或昆蟲的數量做正確的推估與預測，以做進一步的防範與措施。

一般數量生態學家(population biologist)最常用到的方式是根據生物學的知識，以及觀測記錄或實驗所得的一系列時間序列的數據，利用數學與統計為工具，建立一既定的數學模式(deterministic model)，用以描述族群數量隨時間消長的情行。因此，既定模式的參數估計與可能的範圍便會影響到對整個動態族群數量所做的推論與決策。

在建立既定模式的過程，最常見的方式是利用在時間點 t 的數量 N_t 去預測下一個時間點 $t+1$ 的數量 N_{t+1} ，並與實際觀測所得的數量 O_{t+1} 做比較，兩者之間的差異稱之為過程誤差(process error)。至於過程誤差的原因，則有許多數量生態學者歸諸於環境變異(environmental variation) (Higgins, et al, 1997)或來自於個體的存活率與繁衍率的差異(demographic variation)，或來自於不可抗拒的自然因素(catastrophic)，甚至因應於自然環境的變遷，物種本身基因(genetic)所做的改變。至於其間的差異，則眾說紛紜(Dennis, Munholland, and Scott, 1991)，此處我們統稱之為過程誤差。

由於既定模式中 N_{t+1} 取決於 N_t ，但是 N_t 在實際觀測時因為取樣誤差，所得的數量為 O_t ，因此便有觀測誤差(observation error)的現象，需要在統計推估上作一調整。但是就以現有文獻發現，由於單一時間序列的數據過程誤差與觀測誤差的不可區別，只能取一誤差假設而捨另一誤差，因此有必要以更精確的統計方法同時考慮兩種誤差加以調整，這也是本計劃的主要研究內容與目的。

在統計方法上，上述問題可視為不完整數據的一個特例。比較特別的是，此處不完整數據是一連串彼此相關的時間序列，因此與傳統的資料補齊方法有很明顯的差異。

三、結果與討論

本計畫主要結果發現如下：

- 1、在統計方法上，由於 EM algorithm 的單調收斂性質對此資料型態並不成立，未小心審視即套用有其危險性。MCMC 法對此資料型態是可行的方案。而所提出的 MCM 法可以有效改進 EM algorithm 此處的缺失，其相關的性質值得進一步的探討。
- 2、若未區分過程誤差與觀測誤差的不同，所得的估計值會有一定程度的偏差。利用幾次的精密觀測(validation data)對象族群真正數量，並在隨後的觀測中，將無法正確觀測而得的實際數量視為缺失值，並以 MCMC 等方法將缺失值補齊，再進而估計參數值是一個相當有效也是在生態學研究非常值得嘗試的方法。

由於人力與物力的限制，精密觀測的次數一般極為有限，因此在觀測誤差的

分布上，必須做一些假設（通常為 log-normal 分布），而不易以半參數的方式估計，這是本研究方法一個自然的限制。

在進一步的應用上，由於許多生態族群數量的成長牽涉到非線性且有不同成長階段的既定模式，因此統計方法將更為複雜，也值得更進一步深入的探討。

一、計畫成果自評

本計畫成果雖然與原定名稱有些出入，但可視為原定計畫內容的先期結果，並進一步考慮在有遺傳與環境變異的情形下，如何以統計方法處理相關問題的參數估計。預期這將會是生態學研究上相當重要的課題。

二、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. *Computational Statistics Quarterly*, **2**, 73-82.
- Chen, C. C., and Hsieh, F. (2002). Parameter estimate for dynamic population growth in the presence of observation errors. *Submitted*.
- Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society, Series B*, **39**, 1-38.
- Dennis, B., Munholland, P. L., and Rubin, D. B. (1991). Estimation of growth and extinction parameters for endangered species. *Ecological Monographs*, **61**(2), 115-143.
- Diebolt J. and Celeux, G. (1993). Asymptotic properties of a stochastic EM algorithm for estimating mixing proportions. *Communications in Statistics- Stochastic Models*, **9**(4), 599-613.
- Gause, G. F. (1934). *The struggle for existence*. Williams and Wilkins, Baltimore, Maryland, 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*. Chapman & Hall, London, UK.
- Gilman, A., and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences (with discussion). *Statistical Science*, **7**, 457-511.
- Heyde, C. C., and Cohen, J. E. (1985). Confidence intervals for demographic projections based on products of random matrices. *Theoretical Population Biology*, **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. *Canadian Journal of Fisheries and Aquatic Sciences*, **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*. John, Wiley & Sons, Inc.
- Tuljapurkar, S. D., and Orzack, S. H. (1980). Population dynamics in variable environments. I. Long-run growth rates and extinction. *Theoretical Population Biology*, **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. *Journal of the American Statistical Association*, **85**, 699-704.