知識天地

馬可夫鏈蒙地卡羅法（Markov chain Monte Carlo）

陳定立助研究員 (統計科學研究所)

馬可夫鏈蒙地卡羅法（Markov chain Monte Carlo）是一種抽樣方法。它主要是針對狀態空間（state space）巨大而不容易抽樣的狀況。我們先看個簡單的例子：一個一百乘一百像素的照片，每個像素都只有黑跟白兩種可能。即使這麼簡單的情況，可能的樣本有二的一萬次方個，這大約是一後面加三十個零。即使我們知道每個狀態的機率是正比於某個已知的函數值，由於狀態空間太大，使得要計算一些我們有興趣的統計期望值變得不可能。馬可夫鏈蒙地卡羅法，就是針對類似的情況，設計出一個馬可夫鏈來隨機抽樣出有代表性的樣本。

一個馬可夫鏈是由一個傳遞矩陣（transition matrix）所決定。所謂的傳遞矩陣，是由所有從一個狀態到另一個狀態的機率而構成。在數學上，只要這傳遞矩陣有一些好的性質，就可以證明經由這馬可夫鏈抽樣的分佈會收斂到想要的機率分佈。對於一個目標分佈，可以找到非常多個馬可夫鏈，最後會收斂到該分佈。問題是：哪一個馬可夫鏈，收斂的速度比較快，比較有效率？目前最被廣泛使用的兩種馬可夫鏈蒙地卡羅法，一個是 Metropolis-Hasting algorithm (Metropolis et al. 1953; Hasting 1970)，另一個是 Gibbs Sampler (Geman and Geman 1984)。對於馬可夫鏈蒙地卡羅法的收斂速率問題，有許多的學術討論，但能實際以理論比較這兩種方法的研究並不多。本院數學所研究員黃啟瑞和許順吉及其他合作者第二大特徵值（second largest eigenvalue）來比較各方法（Frigessi et al. 1993），另外也以最小漸近變異（asymptotic variance）的觀點，在最壞狀況分析（worst-case analysis）下，找出最佳的傳遞矩陣（Frigessi et al. 1992）。

必須一提的是，上述的最佳傳遞矩陣是限定在可逆（reversible）的條件下，而所有常用的馬可夫鏈蒙地卡羅法都是可逆的。但是，可逆是必要的嗎？其實不然。那究竟什麼原因使得所有這方面的研究都限制在可逆的條件下？當一個馬可夫鏈最終能收斂到想要的分佈，一個必要條件是在目標分佈下，每個狀態流出的機率要等於流入的機率。但由於處理的狀態空間巨大，要加總所有流出和流入的機率再判斷是否相等，是一件困難而通常不可行的事。而所謂的可逆是指任選兩個狀態，由狀態甲流到狀態乙的機率等於由狀態乙流到狀態甲的機率。因為對任一狀態，從它流到某狀態的機率跟流過來的機率相等，則從它流出的總機率自然會等於流入的總機率。所以，可逆這個條件，其實只是為了方便，讓大家容易去判定是否滿足，進而確保收斂的正確性。

接下來要問的是：最佳的傳遞矩陣是否其實剛好都會是可逆的？答案是否定的。我們先從一些簡單的例子來觀察。如果狀態空間中只有兩個狀態，兩者機率相等，而我們想知道的是某個函數在這機率分佈下的期望值。當然，我們其實只要把這兩個狀態的函數值加總平均就好了。不過，我們現在是要藉由這個例子觀察，若用馬可夫鏈蒙地卡法來解決這問題，我們應該選用怎樣的傳遞矩陣。在這個簡單問題上，很明顯地，最佳的傳遞矩陣應該是：每個狀態總是以百分之百的機率流到另一個狀態。也就是，這樣的馬可夫鏈會交替選擇兩個狀態。類似的例子，如果狀態空間中有 n 個狀態且機率相等，則最佳的傳遞矩陣是每一個狀態都百分之百流到某狀態，全部狀態輪過一輪後再輪回來。當 n 大於 2 時，這樣的傳遞矩陣明顯不是可逆的，因為每個狀態流入與流出的對象一定不同。

在我們最近的研究中（Chen et al. 2011a），對任何有限狀態空間下的任何機率分佈，我們以平均狀況分析（average-case analysis），得到了對應的漸近變異最小值。我們並且證明，總共有 2 的 n - 2 次方個傳遞矩陣可以達到這個最小值。我們找出了所有最佳傳遞矩陣的一般形式，其中有一個傳遞矩陣是：最小機率的狀態百分之百跑到第二小機率的狀態，第二小機率的狀態百分之百跑到第三小機率的狀態，依此類推，第二大機率的狀態百分之百跑到最大機率的狀態。目前，除了最大機率的狀態，其它狀態流入的機率都小於流出的機率。而這之間的不足，就剛好由最大機率狀態流出來達到平衡。

這樣的數學結果雖然漂亮，但卻無法直接使用。為什麼？因為像是上述的最佳傳遞矩陣，機率的流動幾乎都是從小流到大。而要達成這樣，表示我們必須知道所有狀態的機率排序。在最一開始已經提過，馬可夫鏈蒙地卡羅法所要處理的問題，狀態空間通常非常巨大，因此要將所有狀態依機率大小做排序，在實用上是件不可能的事。雖然理論的結果無法直接用上，但我們可以在局部上應用它。一般碰到的巨大狀態空間，常

常是維度很高（像本文第一段所舉的例子，維度是一萬維），而每個維度的狀態則可能不高。如常用的 Gibbs Sampler，其傳遞矩陣就是讓馬可夫鏈移動時，每次都在一個維度上動，它的動法是在選定維度下，根據條件機率去抽樣。我們模仿這樣的想法，在選定維度下，套上我們得到的結果，將條件機率排序，然後從小流到大，最大的再分配到各地方。我們將這方法稱為局部最佳抽樣法（Locally Optimal Sampler）。

這樣的方法，就可以實際使用了，程式撰寫上也容易。只要將 Gibbs Sampler 中依據條件機率抽樣的地方，多加進一個排序的動作。而排序後，除非剛好目前的狀態是機率最大的，否則不需要經過隨機抽樣，因此執行這方法不需要比較多的時間。我們無法理論上證明我們提出的局部最佳抽樣法總是比 Gibbs Sampler 或 Metropolis-Hasting algorithm 好。但在一些狀態個數不太大（小於十萬）的例子，經過實際的計算，我們的方法的確平均來說有較小的漸近變異。我們也同時從模擬實驗著手，實驗結果顯示我們的方法通常表現較佳（Chen et al. 2011b）。在許多模擬的例子中，我們的方法只需要不到一半的執行迭代次數，收斂情況已經比其它方法來得好。

參考文獻：

- Chen, T.-L., Chen, W.-K., Hwang, C.-R., and Pai, H.-M. (2011a). *On the Optimal Transition Matrix of MCMC Sampling* (submitted).
- Chen, T.-L., Hwang, C.-R., Shiu, S.-Y. (2011b). *Locally Optimal Sampler* (in preparation).
- Frigessi, A., Hwang, C.-R. and Younes, L. (1992). *Optimal Spectral Structure of Reversible Stochastic Matrices, Monte Carlo Methods and the Simulation of Markov Random Fields*, Ann. Appl. Probab., vol. 2, pp. 610-628.
- Frigessi, A., Hwang, C.-R., Sheu, S.-J., and di Stefano, P. (1993). *Convergence rate of the Gibbs sampler, the Metropolis algorithm, and other single site update dynamics*, J. Royal Statist. Soc., vol. B55, pp. 205-219.
- Geman, S. and Geman, D. (1984). *Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images*. IEEE Trans. Pat. Anal. Mach. Intel. vol. 6, pp. 721-741.
- Hastings, W. K. (1970), *Monte Carlo Sampling Methods Using Markov Chains and Their Applications*, Biometrika, vol. 57, pp. 97-109.
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953), *Equations of State Calculations by Fast Computing Machines*, Journal of Chemical Physics, vol. 21, pp. 1087-1092.