本院1997年年輕研究人員著作獎得獎人研究成果簡介

長遠相關與自似性——從“獨立宣言”到創世紀

何淮中
統計科學研究所副研究員*

統計學發展至今，已經脫離傳統的窠臼，不再僅僅滿足於把數據包裝成花俏圖表的敘述性角色了。從機率論的基礎中吸取而得的豐富內涵以及多樣化應用的觸角，加之與電腦快速運算的結合，統計所呈現的，是一番嶄新活潑的面貌和充滿著生機無限。近十幾年來，許多分支領域都在相互激盪中取得革命性的進展，其中長遠相關過程的研究，由涓滴而成巨流，適足引爲代表性的見証。

獨立性的假設，長久以來，已是統計與機率學家的開門第一件事，所造就的成果，斐然可觀，只是面對許多實際的時間序列資料和問題，此一“獨立宣言"即使加上一些增修條款如m-dependence, Markov process, ARMA等模型，還是免不了時有難以爲繼的窘況，尤其是很多水文學及經濟指標資料，顯示出上下振盪當中，往往起或伏的時程會拉的很長，此一特性，前述的幾個模型一直無法提供滿意的解釋。長遠相關的想法，適時的因應而生，反覆推敲，幾經錘煉，乃逐漸浮現成形，B. Mandelbrot深感其重要性，特將長遠相關稱之爲約瑟效應(Joseph effect)（名稱是取自於舊約創世紀，希伯萊人約瑟爲埃及法老王解夢，預見埃及受到尼羅河水量影響將有連續七年豐收而後七年饑荒）。

晚近甚受學界各方注目的碎形幾何學(fractal geometry)其立論主要基石之一，即是所謂自似性(self-similarity)的概念。循此概念而衍生的研究題材，不可勝數，機率與統計領域中的自似過程，正是對自似性概念從隨機的角度所做的一項註解。與自似過程互爲一體兩面的長遠相關過程，不但跳出傳統平穩過程的限制，開展出理論研究一片新而豐富的領域，在許多應用的學門如計量經濟與財務金融、水文學、環境科學、及電腦網路等，更因提供了另一種簡便有效的建模(modeling)手段，而得到很大的正面迴響。

長遠相關過程由於其結構的特殊，造成研究上很高的挑戰性，也是因爲幾個核心的問題沒有了解清楚，很多比較應用取向的探討就顯得枝枝節節。在面對一組時間序列，基本要務之一，應是其分佈函數的估計和準確的掌握估計誤差，此一問題，在長遠相關及非常態分佈架構下，一直是個瓶頸。許多困難，何以傳統的方法都無法克服，大致情形是這樣的。我們考慮線性非常態長遠相關過程$X_n=\Sigma_{i=1}^{\infty} a_i\varepsilon_{n-i}$，$n$=1,2,...,其中，$a_i$=c.i^{-d}，1/2<d<1，$\varepsilon_i$爲獨立同分佈（亦即i.i.d.），二階矩有限。由於a_i趨近0的速度很慢，而造成X_n的長遠相關性。假設X_n的分佈函數爲$F(x)$，通常用來估計$F(x)$是所謂的經驗分佈函數：$F_N(x)=(1/N)\Sigma_{n=1}^{N} I(X_n\leq x)$，其中函數$I(X_n\leq x)$的值是1，若$X_n\leq x$，

*作者申請本著作獎時職稱爲副研究員，現職爲研究員。

値是0，若X_n>x。當a_i趨近0的速度夠快再配合著ε_i的分佈具備很好的條件，就可以造成X_n很近似獨立的性質，套上一些現成的方法，得到的結果，與在獨立性假設情形下，幾無二致。可是在長遠相關，有些特例已經告訴我們，會得到與在獨立條件下截然不同的結果，現有的方法顯不適用，只好另闢途徑，轉而借助X_n是常態分佈的假設，惟有如此，才可以把$I(X_n \le x)$展開成多項式的和，進而對$F_N(x)$做些分析，可是一旦跳開常態分佈，此路又是不通，整個問題再度陷入困境得不到解決。

我們發展一套新的方法，成功的超越X_n須爲常態以及函數$I(\bullet \le x)$本身極度非線性（甚至不連續）的束縛，將$F_N(x)$的極限行爲與$F_N(x)$估計$F(x)$的誤差給了一個非常完整的答案。方法的主要構想，是充分的利用X_n的線性結構，在函數$I(\bullet \le x)$與X_n同步進行工作。先把X_n拆成前後兩段$X_n = X_{n,1} + X_{n,2}$，後段$X_{n,2}$很重要保留不動，前段$X_{n,1}$則借來經由條件期望値積分的手段，將函數$I(\bullet \le x)$予以平滑化。利用這個方法，我們得到$F_N(x)$-$F(x)$一個類似Martingale Decomposition的表示式，藉著這個表示式，無須知道X_n的確實分佈，就能直接進行一些精密的估算，很準確的描述估計誤差$F_N(x)$-$F(x)$，導出的結果，是如下的公式：

$$F_N(x)-F(x)$$
$$=(-1)F'(x)[S_N]+(-1)^2F''(x)[S_N]^2+\dots, \text{ a. s.},$$

其中F'，F''分別是F的第一階與第二階導數，$[S_N]=(1/N)\Sigma_{n=1}^{N} X_n$，$[S_N]^2=(1/N)\Sigma_{n=1}^{N}(X_n^2)$*，這裡，$(X_n^2)$*是近似於$X_n^2$但期望値爲0的隨機變數。有了這樣一個公式，許許多多與經驗過程相關的重要統計推論工作，如適合性測試(goodness of fit)，密度函數估計，無母數統計量，以及百分數估計等，其應用的理論基礎，才得以確立。

延續前述的方法並做些修正，我們更進一步的從$I(\bullet \le x)$推廣至一般的非線性轉換$K(x_1,\dots, x_p)$，把部份和$A_N \Sigma_{n=1}^{N} K(X_{n+t_1}, \dots, X_{n+t_p})$的極限分佈很清楚的刻畫出來。不僅如此，因爲方法本身甚具一般性，同時適用於近程相關，許多古典模型中從前望之遙不可及的問題，如二次型的收斂速度及重對數法則，也都一併澄清。

長遠相關過程就宏觀而言，本身是平穩的，可是局部來看，它又呈現不平穩性的走勢，正是此一序亂交互與許多實際觀察現象極相契合的特性，引來各方的重視，爲長遠相關研究灌注了豐沛的活水源頭。進入這個領域，你可以思古悠情的倘佯尼羅河上，意氣風發的縱横華爾街，熱帶雨林，溫室效應，電腦網路，紊流迷蹤，無不盡收眼底，淺嚐深吮，隨君之興。志同道合的朋友，歡迎你的加入打拼。

何淮中
學經歷：
東吳大學數學系(1974)
美國韋恩州立大學數學博士(1984)
國立中山大學應數所副教授、教授(1988-91)
本院統計所副研究員(1984)、研究員(1997)