經濟學名著翻譯叢書第一四六種

計量經濟學原理

（下冊）

Henri Theil 著

湯 慎 之 譯

臺灣銀行經濟研究室編印

計量經濟學原理

Principles of Econometrics

Henri Theil 著

湯 愼 之 譯

第八章 漸近分配論

在前二章，在某些場合，曾敍述着一個估計量或測驗統計量漸近化地具有某些分配性質，當樣本是相當大時。在極大多數場合，非常困難去導出統計量之眞正分配，但當樣本很大很大時，此困難將消失。此明白地乃帶有近似誤差，當無窮大樣本之結果用來中等大之樣本時，此中等大小樣本乃計量經濟中之實際情況。

本章發展出大樣本的近似性統計推定理論。在此方面最佳的一般統計理論的原始文獻是 Cramér (1946) 及 C. R. Rao (1965a)。8.1節 及 8.2 和 8.9節中之子節是此方面介紹性的課程用於計量經濟上，餘者較高深。

8.1 極限、機率極限及一致性估計量A

數串之極限

考察一實數數串 $a_1, a_2, \cdots\cdots, a_n$, 例如，$a_n=3+1/n$ ($n=1, 2, \cdots\cdots$)。此數串被稱爲收斂於一極限 α，當對任何一個正數 δ（不論如何小）言，必存在有一個數 $N(\delta)$ 使得絶對差額值 $|a_n-\alpha|$ 小於 δ，對全體 $n>N(\delta)$。此指明如下：

$$(1.1) \qquad \lim_{n\to\infty} a_n = \alpha$$

在本場合，$a_n = 3 + 1/n$， 必有一極限，此等於3，因對任何已知 $\delta>0$ 言，絶對差額值 $|a_n-3| = 1/n$ 是小於 δ，對任何 $n>N(\delta) = 1/\delta$ 皆如此。注意並非每一數串皆有一個極限。反例爲 $\log n$ 及 $-n^2$ 也，兩者當 n 變成無限時，或發散于正方向或負方向。另一反例爲 $(-1)^n$， 此亦無極限，但至少有界限，界限之意義乃指其絶對值至多等於一個正的常數 C（獨立於 n）。

可在統計推定範圍內作一例，考慮一個常態分配變異數 σ^2 之最大概似估計量 $(1/n)\Sigma(X_i-\bar{X})^2$，此乃形成於2.7節之 (7.16) 式。我們在那裏發見此估計量之偏倚性爲

$$(1.2)\qquad \mathscr{E}\left[\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2\right]-\sigma^2=\frac{n-1}{n}\sigma^2-\sigma^2=-\frac{\sigma^2}{n}$$

當我們想像到樣本大增時，偏倚性 $-\sigma^2/n$ 是一個具有零爲極限之數串。雖然 σ^2 的最大概似估計量對任何有限樣本有偏倚性，然此偏倚性在極限中消失，當 n 無限增尤時。

數串大小的階度

考慮數串 $a_n=4+n-3n^2$。當 n 增得大大的時，4 和 n 這二項比之 $-3n^2$，其等絶對值乃小小的，所以，最後這項稱爲數串之領項。這個項決定了數串大小之階度。較形式言。我們說數串 a_n 至多是 n^k 階度，寫成 $O(n^k)$，當數串 $n^{-k}a_n$ 是有界限的。在上例取 $k=2$:

$$n^{-2}a_n=n^{-2}(4+n-3n^2)=\frac{4}{n^2}+\frac{1}{n}-3$$

當 $n\rightarrow\infty$，此收斂于 -3，故此確定爲有界；因此斯數串是 $O(n^2)$。一個相連之觀念是「較 n^k 階度小」，以 $o(n^k)$ 表之，此意義着數串 $n^{-k}a_n$ 收斂于零。在我們之例中，$a_n=o(n^3)$ 但亦爲 $a_n=o(n^{5/2})$ 及 $a_n=o(n^{10})$ 對同樣的 a_n。〔注意我們可取 $k=0$；然後 $a_n=O(n^0)=O(1)$ 此意味着此數串乃有界限的，而 $a_n=o(n^0)=o(1)$，此指有零作極限。〕數串之代數運算乃相當直接的。取 $a_n=3n$ 及 $b_n=1+n^{-1}$；然後 $a_n+b_n=3n+1+n^{-1}$ 及 $a_nb_n=3n+3$，此解明了若 $a_n=O(n^h)$ 及 $b_n=O(n^k)$，然後 $a_nb_n=O(n^{h+k})$ 而 $a_n+b_n=O(n^g)$, $g=\max(h, k)$。

一個隨機串的機率收斂

現在我們進而去考慮在純統計方面之極限觀念。取一個隨機變數串$Y_1, Y_2, \cdots\cdots, Y_n, \cdots$具有分配函數 $F_1(\ ), F_2(\ ), \cdots\cdots, F_n(\ ), \cdots\cdots$。此串被稱為機率收斂于一常數 c；若

$$(1.3) \qquad \lim_{n\to\infty} P[|Y_n - c| > \epsilon] = 0 \quad \text{對每個} \quad \epsilon > 0$$

在括弧內式子 $|Y_n - c| > \epsilon$ 是一不等式，此亦許眞的或不眞。此為眞的機率決定於 Y_n 之分配函數 $F_n()$，c 及 ϵ。故給出 c 及分配串，此機率形成一串 $a_1(\epsilon), a_2(\epsilon)\cdots\cdots, a_n(\epsilon), \cdots\cdots$此等參數性地依存於 ϵ 。故定義 (1.3) 和下列相當的。隨機變數串 $Y_1, Y_2, \cdots\cdots$機率收斂于 c，倘若此機率串之極限等於零，不論任何正值之 ϵ 。此觀念解明於 8.1圖，此乃對任何一串隨機變數，它們具有連續分配。以密度曲線下在區間外 $(c \pm \epsilon)$ 之面積作此機率之測計，此機率之極限在 (1.3) 所考慮的。清楚的，當 $n = 50$時，此面積小於$n = 10$時而 $n = 250$，則更小。

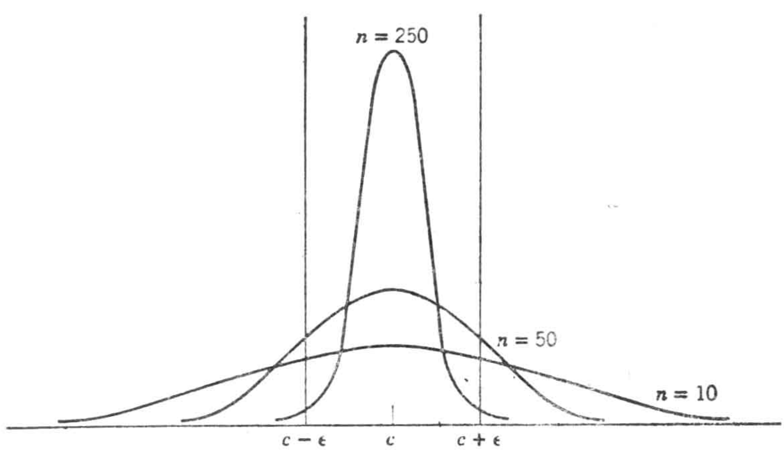

圖 8.1　機率收斂於一常數的解明

若欲知機率收斂于一常數之例，則考慮樣本平均 $\bar{X}_n=(1/n)\Sigma X_i$；此乃一隨機本（$X_1, \cdots\cdots, X_n$），由一任意具有有限平均 μ 及變異數 σ^2 之母體抽出（$\bar{X}_n$ 的添字 n 當作基本樣本之大小 n）。我們知道 $\bar{X}_n$ 的分配具有平均 μ 及變異數 σ^2/n。此充分地保證隨機變數串 $\bar{X}_1, \bar{X}_2, \cdots\cdots, \bar{X}_n, \cdots\cdots$（充分大樣本的樣本平均）機率收斂于母體平均 μ。一個簡單之證明依據于 Chebyshev 的不等式①，此陳述對任何具有有限平均 μ 及變異數 σ^2 的隨機變數 Z 言，平均離差之機率等於 k 乘標準差或至多等於 $1/k^2$：

$$\text{(1.4)} \qquad P[|Z-\mu|\geq k\sigma]\leq\frac{1}{k^2} \quad \text{對任何} \quad k>0$$

在 $\bar{X}_n$ 之場合，此平均是 μ 及變異數爲 σ^2/n，故在括弧之不等式成爲 $|\bar{X}_n-\mu|\geq k\sigma/\sqrt{n}$。命 $k=\epsilon\sqrt{n}/\sigma$ 這樣特定，然後得到

$$\text{(1.5)} \qquad P[|X_n-\mu|\geq\epsilon]\leq\frac{\sigma^2}{n\epsilon^2} \quad \text{對任何} \quad \epsilon>0$$

因 $\sigma^2/n\epsilon^2$ 收斂于零，當 $n\to\infty$ 時，只要 $\epsilon>0$，從（1.3）定義，我們可結論着從任意一具有有限變異數之母體中拿出隨機樣本（$X_1, \cdots\cdots, X_n$）之樣本平均 $\bar{X}_n$ 機率收斂于母體平均 μ。

① 閱2.1節末之問題1.1。

Khintchine 定理

前段之結果是 Khintchine 定理之特例，它可作下列形式。命 $X_1, \cdots\cdots, X_n$ 是獨立之隨機變數，皆具有相同之分配，而此相同分配有有限平均 μ；然後它們的平均 $\bar{X}_n$ 機率收斂于 μ 當 $n\to\infty$ 時。注意此定理並不須要變異數有限的。閱問題2.2之下，乃其證明，此須要在下節所考慮之一定理。

隨機串之機率密度

（1.3）符號指明機率收斂嫌麻煩。較簡潔之符號而廣用者是

$$\operatorname*{plim}_{n\to\infty} Y_n = c \tag{1.6}$$

或用文字說：隨機串的機率極限（縮寫為 plim）是等於 c。此全等於：〔串 $Y_1, Y_2,\cdots\cdots$ 機率收斂于 c〕，但此名詞機率極限具有額外之利益（除出符號簡潔外），此即着重於有別于普通之極限。當 $Y_1, Y_2,\cdots\cdots$ 是一非隨機數串而具極限 c 時，我們可確定為 $|Y_n-c|$ 小於任何已給出之 ϵ，當n充分大時。當 $Y_1, Y_2,\cdots\cdots$ 是一隨機串而具機率極限 c 時，我只能說 $|Y_n-c|<\epsilon$ 成立之機率任意地接近於 1，當 n 充分大。但當 Y_n 之變域是無限，則永遠有一正的機率使得 $|Y_n-c|$ 大於 c，不管 n 如何大。此普通極限明顯的是機率極限之一特例，故關於後者之定理先天上可用於前者。

擴　展

機率收斂之觀念可直接擴展到 Y_n 及 c 為向量或矩陣而非數值之場合。然後 Y_n 之每一元素機率收斂于 c 的對應元素。更加，若 $g(\ )$ 是一連續函數而若 Y_n 之機率極限等於 c 如 (1.6) 陳述，然後

$$\operatorname*{plim}_{n\to\infty} g(Y_n) = g(\operatorname*{plim}_{n\to\infty} Y_n) = g(c) \tag{1.7}$$

以文字言：一個隨機串之機率極限是等于另一隨機串之機率極限的同一函數，若此原先隨機串之每一項為另一隨機串相應項之函數 $g(\ )$，假定此極限存在而此函數乃連續的。證明乃簡截的，而可摘述如下。若 $g(\ )$ 是連續的，它的值任意靠近 $g(c)$，當它的變量 Y_n 是充分地接近于 c。但 (1.6) 保證當 n 充分大時，此條件因機率任意靠近于1被滿足。故 $g(Y_n)$ 非常靠近于 $g(c)$當 $n\to\infty$ 時，增加其確定度。又此結果易推廣于連續向量及矩陣函數之上。

一致性估計量

機率收斂乃一特別重要觀念，當隨機串所涉者為一估計量而依

據于繼續增大樣本（$n=1, 2, \cdots\cdots$）及此串收斂之常數乃對應於未知參數時。特定地，命 θ 為一參數，此乃使某一分配所具之特徵，而假想一個隨機樣本（$X_1, \cdots\cdots, X_n$）乃由母體抽出。一個統計量 $\hat{\theta}_n$ 用來作 θ 之估計量；此添字 n 指示着此估計量根據的樣本大小。若 $\hat{\theta}_n$ 機率收斂于 θ：

$$(1.8) \qquad \lim_{n\to\infty} P[|\hat{\theta}_n-\theta|>\epsilon]=0 \quad \text{對任何 } \epsilon>0$$

或全等地說，若 θ 和 $\hat{\theta}_n$ 的機率極限合于一起，

$$(1.9) \qquad \operatorname*{plim}_{n\to\infty} \hat{\theta}_n=\theta$$

然後 $\hat{\theta}_n$ 是稱為 θ 之一致性估計量。故回到從具有有限變異數之母體抽得之樣本的樣本平均，我們證明 $\bar{X}_n$ 機率收斂于母體平均 μ；此等於說 $\bar{X}_n$ 是 μ 的一致性估計量。回憶到我們已證明 $\bar{X}_n$ 之期望值等于 μ 而它的變異數收斂于零。故下列之結果依據 Chebyshev 不等式乃可一般性地成立。一個估計量欲有一致性之充分條件為 $\mathscr{E}\hat{\theta}_n=\theta$（對全體之 n 及 θ 皆相同的）及 $\lim(\mathrm{var}\,\hat{\theta}_n)=0$，當 $n\to\infty$ 時。注意此些不是必須條件。此估計量亦許偏倚的，但仍可一致性倘若此偏倚收斂于零；閱 1.7問題。在某些情況，甚致於並不須要對任何有限之 n 言，此變異數欲存在的。若 $X_1, \cdots\cdots X_n$ 獨立而具有相同分配，其平均為 μ，則諸 X_i 之平均 $\bar{X}_n$ 為 μ 之一個一致性估計量。在8.3節，我們將回到對有限 n 言的估計量的變異數存在問題上去。此乃 Khintchine 定理供給一例也。

在標準線性模型條件下，LS 估計量之一致性

故一致性意味着當此樣本成為很大很大時，那只有很少機會使得此估計量和參數相差大於任何已給定之數。可證出在適切條件下，LS 估計于標準線性模型乃具此性質：

定理 8.1 （LS 係數及變異數估計量之一致性）假定對已給定 X言，3.2節之假定3.3是眞的。然後 LS 向量 $\mathbf{b}=(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ 爲 $\boldsymbol{\beta}$ 之一致性估計量，若矩陣 $(1/n)\mathbf{X}'\mathbf{X}$ 收斂于一個正定矩陣 $\mathbf{Q}$，$n\to\infty$ 時，而統計量 $s^2=\mathbf{e}'\mathbf{e}/(n-K)$，$\mathbf{e}=\mathbf{y}-\mathbf{X}\mathbf{b}$ 是 σ^2 的一致性估計量，若 $\mathbf{y}$（已知 $\mathbf{X}$）之分配是多元常態的。

$\mathbf{b}$ 之一致性可藉 3.3節之定理 3.2來證明，按照此定理，斯估計量乃有K度向之分配，附有平均向量 $\boldsymbol{\beta}$ 及共變異矩陣 $\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$②。此矩陣可寫成 $(1/n)\mathbf{X}'\mathbf{X}$ 之逆陣乘上一數值 σ^2/n 作乘數。後者之矩陣收斂于 $\mathbf{Q}$，當 $n\to\infty$ 時。因逆陣之元素，是原矩陣元素之函數（若此矩陣具有逆陣——但乃依假定而眞的），$(1/n)\mathbf{X}'\mathbf{X}$ 之逆陣收斂于 $\mathbf{Q}^{-1}$。故

$$\lim_{n\to\infty}\mathscr{V}(\mathbf{b})=\lim_{n\to\infty}\left[\frac{\sigma^2}{n}\left(\frac{1}{n}\mathbf{X}'\mathbf{X}\right)^{-1}\right]=0\mathbf{Q}=0 \tag{1.10}$$

因 σ^2/n 收斂于零。故 $\mathbf{b}$ 之全體元素是不偏的而其變異數收斂于零，故 Chebyshev 的不等式派上用場以建立有關 β 元素之一致性。

在常態條件下，欲證明 s^2 的一致性，回憶到隨機變數 $(n-K)s^2/\sigma^2$ 按照3.5節之定理 3.8，乃以 $\chi^2(n-K)$ 分配；故其平均爲 $n-K$ 而其變異數是 $2(n-K)$，故 s^2 是 σ^2 之不偏估計量，而具有變異數 $2\sigma_4/(n-K)$（閱3.5節定理3.7）。因此變異數收斂于零，當 $n\to\infty$，s^2 的確爲 σ^2 之一致性估計，此完成了證明。

② 也許欲以串的符號來限制而寫成 X_n 非寫成 X 來得好，但我們却不願意和以前數章之符號相脫節。

解釋變數動產的收斂條件

$\lim n^{-1}\mathbf{X}'\mathbf{X}=\mathbf{Q}$ 的條件在下面將大量被用到③。我們假想着

③ 注意此條件並非 s^2 的一致性之必要條件。另方面，若它被滿足，我們可用獨立及同一分配的擾亂項，這樣較弱條件來代常態條件，以獲得對 s^2 同樣一致性之結果，閱8.3 節下之定理 8.2。

對每一新的觀察，一個新列加在矩陣 $\mathbf{X}$ 上，而早先之列在此樣形式下保持着不受影響，即 $\mathbf{X}'\mathbf{X}$ 的 K^2 元素乃是 $O(n)$。外加，我們排除了 $n^{-1}\mathbf{X}'\mathbf{X}$ 之極限爲獨異之可能性。

假定 8.1 ($n^{-1}\mathbf{X}'\mathbf{X}$ 的收斂) 矩陣 $n^{-1}\mathbf{X}'\mathbf{X}$ 收斂于一 $K \times K$ 正定之矩陣 $\mathbf{Q}$，當 $n \to \infty$ 時。

讀者亦許想到此假定無過誤，若已知利用之資料永遠是有限數而極限過程 $n \to \infty$ 不過是一數學觀念。但此假定隱含着某些限制，此可由下列認清的。

(1) 考慮到最簡單場合 $\mathbf{X}' = [1 \quad 1 \cdots\cdots 1]$ 無斜率係數，只有一常數項。然後 $(1/n)\mathbf{X}'\mathbf{X} = 1$，此明顯的收斂于一個正值（一），當 $n \to \infty$ 時。明證的，在此場合無有困難發生。

(2) 取一直線趨勢而無常數項：$\mathbf{X}' = [1 \quad 2 \cdots\cdots n]$。然後 $\mathbf{X}'\mathbf{X} = O(n^3)$，故 $(1/n)\mathbf{X}'\mathbf{X}$ 並不收斂于一有限極限。雖然 LS 係數是一致性，當 定理8.1 之其他條件被滿足；閱問題1.9。

(3) 又取一個解釋變數及無常數項，但命那個變數之值的形式爲 $x_\alpha = (\frac{1}{2})^\alpha$: $x_1 = \frac{1}{2}$, $x_2 = \frac{1}{4}$, $x_3 = \frac{1}{8}$, $\cdots\cdots$ n 觀察值平方之和爲 $\Sigma x_\alpha^2 = [1-(\frac{1}{4})^n]/3$，此乃小於 $\frac{1}{3}$ 對任何有限n言，而 $n \to \infty$ 時，它收斂于那 $\frac{1}{3}$。因此，LS 係數具一變異數，它永不會小於 $3\sigma^2$，不管 n 多大。直覺來說，理由乃簡明的。當我們有一個解釋變數而無常數項，那個變數之係數被估計之準確性決定于（已知σ^2）此變數之絶對值。在 $x_\alpha = (\frac{1}{2})_\alpha$ 特定化下，此些值快速地收斂于零，因此從某一點向前言，所有後面觀察值之貢獻對於準確性皆可忽略。

(4) 假想我們有二個解釋變數而假定 $x_{\alpha 2} = kx_{\alpha 1}$ 對某一 k 及所有超過某數 N 的 α 成立。當 $(1/n)\Sigma x_{\alpha 1}^2$ 收斂于一有限正極限時，矩陣 $(1/n)\mathbf{X}'\mathbf{X}$ 收斂于一特異陣 $\mathbf{Q}$；閱問題1.10。此乃［在極

限內極端線性重合」之一例。

「在重複樣本內之常數」

上例解明了假定8.1須要所有的解釋變數的均方及乘積平均收斂于有限極限，此些極限形成一正定矩陣④。當除出已知觀察矩陣 $\mathbf{X}$ 外對解釋變數的行爲無所知時，體會出後者之值的簡便法如下。矩陣 $\mathbf{X}$ 含有首先 n 個觀察。假想此些變數的所有後面之觀察值皆可在 n 個組內獲得，每一個這樣的組完全同於原來的 $\mathbf{X}$ 矩陣。對 pn 個觀察值，我們有 $pn \times K$ 矩陣

$$(1.11)\qquad \begin{array}{l} \begin{bmatrix} \mathbf{X} \\ \mathbf{X} \\ \vdots \\ \vdots \\ \mathbf{X} \end{bmatrix} \end{array} \begin{array}{l} \text{第一組 } n \text{ 個觀察值} \\ \text{第二組 } n \text{ 個觀察值} \\ \\ \\ \text{第 } p \text{ 組 } n \text{ 個觀察值} \end{array}$$

倘若我們以其本身之轉置來前乘，然而以觀察數目（pn）之倒數來乘之，可得到

$$(1.12)\qquad \frac{1}{pn}(\mathbf{X}'\mathbf{X}+\cdots\cdots+\mathbf{X}'\mathbf{X})=\frac{1}{n}\mathbf{X}'\mathbf{X}$$

因此，n 及 $\mathbf{X}$ 皆固定；把觀察數目無限增加在此場合等於 $p\to\infty$ 的極限過程。可推出者爲假定8.1之 Q 觀等於固定的矩陣 $(1/n)\mathbf{X}'\mathbf{X}$，故此假定明顯的被滿足，當此（固定的）矩陣 $\mathbf{X}$ 具有完全行級時。此增加觀察數目之特殊方法可視作「重複樣本內之常數」，可就這樣說明的。注意此過程不能使用在每一型之變數上。顯明的，它可用於常數項變數上，因此對每一觀察皆取一，但不能用於一直線趨勢上。

④　注意此平均乘積包括了平均，當此廻歸帶有一常數項。

問　題

1.1 證明數串 $a_n=(-1)^n+2/n$ 是有界限的。

1.2 下列之數串皆爲 $O(n^k)$ 對某 k 言；在每一場合決定出 k。
(1) $2+4n-n^2$；(2) $1+5\sqrt{n}$；(3) $2-5/\sqrt{n}$。

1.3 考慮兩條數串：$a_n=O(n^h)$ 及 $b_n=o(n^k)$。證明 $a_nb_n=o(n^{h+k})$ 及 $a_n+b_n=O(n^h)$ 若 $h\geq k$, $O(n^k)$ 若 $h<k$。

1.4 證明 (1.3) 式定義之機率收斂亦可由下面相等之方法加以定義。對任何一對正數 (δ, ϵ)，不管如何小，必存有一數 $N(\delta, \epsilon)$ 而這樣的，$|Y_n-c|>\epsilon$ 之機率小於 δ，當 $n>N(\delta, \epsilon)$。

1.5 在 $\operatorname{plim} Y_n=c$ 及 $g(\ )$ 在一區間連續而含有常數 c 爲一內點之條件下，嚴密證明 (1.7)。

1.6 命 $\mathbf{A}_1, \mathbf{A}_2, \cdots\cdots, \mathbf{A}_n \cdots\cdots$ 爲一固定階數的隨機方陣的串具有機率極限 $\mathbf{B}$。證明 $\operatorname{plim} \mathbf{A}_n^{-1}=\mathbf{B}^{-1}$，若 $\mathbf{B}$ 是特異的。

1.7 證明 $\hat{\theta}_n$ 是對 θ 之一致性的，若 (1)$\hat{\theta}_n$ 爲不偏而此偏倚收斂于零，當 $n\to\infty$ 及 (2) 此估計量之變異數也收斂于零。〔提示。寫 $\hat{\theta}_n-\theta=(\hat{\theta}_n-\mathscr{E}\hat{\theta}_n)+(\mathscr{E}\hat{\theta}_n-\theta)$ 而證明 $\operatorname{plim}(a_n+b_n)=\operatorname{plim} a_n+\lim b_n$，在此 $a_n=\hat{\theta}_n-\mathscr{E}\hat{\theta}_n$ 是一隨機串而有一有限機率極限且 $b_n=\mathscr{E}\hat{\theta}_n-\theta$ 爲一非隨機串具有有限極限。〕

1.8 證明假定8.1可被較弱條件 $\lim(\mathbf{X}'\mathbf{X})^{-1}=\mathbf{0}$ 來代替，對 $\mathbf{b}$ 之有限性言，(定理 8.1) 而證明後者條件確較弱。

1.9 考慮在假定8.1下之例 (2) (無常數項之直線趨勢) 而證明在標準線性模型之假定下的 LS 係數的變異數是 $6\sigma^2/n(n+1)(2n+1)$。(提示：閱 1.7節末之問題 7.1) 。又證明此係數是相對應參數的一致估計量。其次，取一有常數項的線性趨勢：

$$\mathbf{X}'=\begin{bmatrix} 1 & 1 & \cdots\cdots & 1 \\ 1 & 2 & \cdots\cdots & n \end{bmatrix} \tag{1.13}$$

而證明此特定化亦破壞了假定 8.1。導出兩個 LS 係數的共變

異矩陣而證明它收斂于 2×2 零矩陣，故此建立了此些係數之一致性。（已知它們的不偏性在標準線性模型之假定下。）

1.10 考慮兩個解釋變數（無常數項）及假想 k 及 N＝數存在使得 $x_{\alpha 2}=kx_{\alpha 1}$ 對 $\alpha>N$ 成立。找出在 $(1/n)\Sigma x_{\alpha 1}^2$ 收斂于一正極限下之 假定8.1 之矩陣 $\mathbf{Q}$，且證明此 $\mathbf{Q}$ 是獨異的。

8.2　極限分配及中央極限定理AC

極限分配

一致性只意味着樣本誤很小之機率甚高，當樣本充分大時。此觀念並不提供關於樣本誤的陳述，例如標準誤如何。爲了那種目的，分析者必須加以擴充，所以我們回到我們的隨機變數串 Y_1, Y_2, …… 具有 $F_1(\ ), F_2(\ )$, …… 分配函數之上。此串被稱爲分配收斂於于分配於隨機變數 Y 附有分配函數 $F(\ )$，若 $F_n(x)$ 收斂於 $F(x)$ 當 $n\to\infty$ 對 $F(\ )$ 之全體連續點言。此分配 $F(\ )$ 被稱爲隨機變數串的極限分配。注意此定義把機率收斂於常數 c〔閱(1.3)〕作爲一特例；此極限分配然後消化使所有質量集中在 c。此觀念解釋於8.2圖對 $F_{10}(\ ), F_{100}(\ )$，及極限分配 $F(\ )$。明白的，$F_{100}(\ )$ 比之 $F_{10}(\ )$ 更接近於 $F(\ )$。

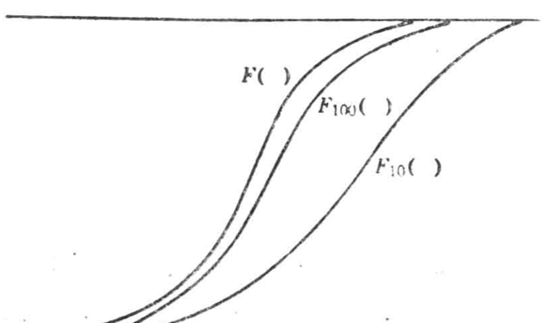

圖 8.2　分配收斂的解明

在本節可發見在前二章所論之漸近標準誤乃指相聯於估計量的極限分配的標準差。又可發見若串 $F_1(\ )$, $F_2(\ )$ 如前段所述收斂于 $F(\ )$，則極限分配 $F(\ )$ 之標準差 (σ) 不一定相同於分配 $F_1(\ )$, $F_2(\ )$ …… 的標準差串之極限。事實上，存存有一組重要之場合，此即 σ 存在，雖 σ_1, σ_2, …… 不存在。此些概念在 8.3 節將系統地追探。

特性函數及在決定極限分配中其任務C

欲導出已知的分配串之極限分配，我們回到 2.5節，在該處我們發展出一個隨機變數**X**之動差生成函數 $M(t) = \mathscr{E}(e^{tX})$。我們須要之定理並不對動產生成函數可眞正成立，但對特性函數 $\phi(t)$ 却可成立，此定義爲 e^{itX} 之期望值，在此 $i = \sqrt{-1}$ 爲虛數單位⑤。

⑤ 閱第二章之註⑥。

$$\phi(t) = \mathscr{E}(e^{itX}) \tag{2.1}$$

對某一定分配而求其特性函數是簡捷的，當動差生成函數已知時；那簡單的是以 it 代替 t 而把 i^2 作 -1 代入。故根據 2.5節之 (5.4) 及 (5.7) 式，我們可找出具有平均數 μ 及變異數 σ^2 之單元常態分配之特性函數：

$$\phi(t) = \exp\left(\mu it - \frac{1}{2}\sigma^2 t^2\right) \tag{2.2}$$

而對平均向量 μ 及共變異矩陣 Σ 之多元常態分配之特性函數爲：

$$\phi(\mathbf{t}) = \exp\left(i\mu'\mathbf{t} - \frac{1}{2}\mathbf{t}'\Sigma\mathbf{t}\right) \tag{2.3}$$

此特性函數亦可用來產出動差。在一向度場合，利用 2.5節第一段所言之 Taylor 展開，以 it 之冪代替t之冪：

$$\phi(t) = 1 + \mu it + \mu_2' \frac{(it)^2}{2!} + \cdots\cdots + \mu_k' \frac{(it)^k}{k!} + R_k \tag{2.4}$$

在此，μ, μ_2', μ_3', …… 乃沿着零的連續性動差；最後項 (R_k) 爲剩餘項。易證出對獨立隨機變數之線性函數函由動差生產函數所導出之規則同等地對特性函數亦可成立。〔閱2.5節之 (5.10) 及 (5.11) 式〕。

依據特性函數而決定隨機變數串 Y_1, Y_2, ……, 的極限分配之重要定理如下⑥。命 $\phi_1(t)$, $\phi_2(t)$, …… 爲 Y_1, Y_2, ……, 的特性函數串，而假想 $\phi_n(t)$ 收斂于 $\phi(t)$, 當 $n\to\infty$ 時，對全體 t；然後，Y_1, Y_2, …… 串分配收斂于極限分配，此極限分配之特性函數是 $\phi(t)$，假若 $\phi(t)$ 在 $t=0$ 上連續。故此步驟如下。證明特性函數 $\phi_n(t)$ 當 $n\to\infty$ 收歛于一函數 $\phi(t)$，檢查在 $t=0$ $\phi(t)$ 之連續性，而試去認識一種分配，它的特性函數爲 $\phi(t)$。那個分配是串 Y_1, Y_2, …… 的極限分配。

⑥ 有關證明，例如可閱 Cramér (1946, 96—98頁)。

中央極限定理 (Lindeberg-Lévy) C

上述之步驟現可用來解明一例，此例爲 n 個獨立相同之分配着的隨機變數 X_1, …… X_n，它們具有平均 μ 及變異數 σ^2，其等之平均爲 $\bar{X}_n$ 我們以均差來測量每一變數；$X_i-\mu$ 之特性函數以 (2.4) 言，可寫成 $1-\frac{1}{2}\sigma^2t^2+R_2$。其次，我們以 $\sigma\sqrt{n}$ 除 $X_i-\mu$ 而我們回憶及第k階動差是被 $(\sigma\sqrt{n})^k$ 除。$X_i-\mu$ 對 $\sigma\sqrt{n}$ 比的特性函數成爲 $1-\frac{1}{2}t^2/n+o(n^{-1})$；$o(n^{-1})$ 乃依據剩餘項 R_2 被除以 $(\sigma\sqrt{n})^k$, $k>2$ 之考慮。最後，我們把 n 個比率加起來：

$$\sum_{i=1}^{n}\frac{X_i-\mu}{\sigma\sqrt{n}}=\frac{(\bar{X}_n-\mu)\sqrt{n}}{\sigma} \tag{2.5}$$

因 n 個比率，即被加起來者，乃隨機獨立，它們和 (2.5) 之特性函數可以 $\phi_n(t)$ 來代表，此可取個別特性函數之乘積得之：

$$\phi_n(t)=\left[1-\frac{1}{2}t^2/n+o(n^{-1})\right]^n \tag{2.6}$$

然後兩邊取自然對數而把 $\log(1+x)$ 對小的 $|x|$ 作熟知的 Taylor 展開（在此等於一很大之數值 n）：

$$(2.7)\qquad \log\phi_n(t)=n\log\Big[1-\frac{1}{2}t^2/n+o(n^{-1})\Big]$$

$$=-n\Big[\frac{1}{2}t^2/n+o(n^{-1})\Big]\to-\frac{1}{2}t^2 \quad \text{當}\ n\to\infty$$

我們結論道：$n\to\infty$，隨機變數（2.5）之特性函數 $\phi_n(t)$ 收斂于 $e^{-t^2/2}$ 對全體 t，此爲標準常態分配之特性函數。注意 $e^{-t^2/2}$ 在 $t=0$ 上乃連續的；此乃上述之部分條件。

前段證明了 Lindeberg-Lévy 的中央極限定理變局，此中央極限定理可以文字敘述如下。我們已知從一含有平均 μ 及變異數 σ^2 常態母體所取之隨機樣本的樣本平均 $\bar{X}_n$ 本身具有以 μ 作平均 σ^2/n 作變異數之常態分配；或相等地說，隨機變數（2.5）遵守着一標準化常態分配，若樣本從一常態母體中取得。現我們發見當我們從任何具有平均 μ 及變異數 σ^2（兩者皆有限）之母體中抽出之樣本時，（2.5）之極限分配爲標準常態型。此常常如此地敘述着樣本平均 $\bar{X}_n$ 漸近地作常態分配，具有平均數 μ 及變異數 σ^2/n。注意 $\bar{X}_n$ 之極限分配退化而所有它的質量集中於 μ。當我們說到 $\bar{X}_n$ 漸近地遵守着平均 μ 及變異數 σ^2/n 常態分配時，$\bar{X}_n$ 之分配當趨向于退化分配之時。故卽使 $\bar{X}_n$ 並非常態分配，對任何 n 言，我們仍可作出近似性質之陳述，卽其分配近似地爲以 μ 作平均及 σ^2/n 變異數之常態分配，若 n 是相當大。

延　展C

把 Lindeberg-Lévy 定理以向量作一般化可形成如下。命 $\mathbf{x}_1, \cdots\cdots, \mathbf{x}_n$ 爲一獨立隨機抽樣，從 k 向度母體而來，該母體具有平均向量 μ 及有限共變異矩陣；然後，向量 $n^{-\frac{1}{2}}\Sigma_i(\mathbf{x}_i-\mu)$ 具有一

多元極限分配，其平均向量爲零向量及共變異矩陣 $\mathbf{A}$。相等地說，向量 $\bar{\mathbf{x}}=(1/n)\Sigma\,\mathbf{x}_i$ 是漸近地常態分配，具有平均向量 μ 及共變異矩陣 $(1/n)\mathbf{A}$。此項證明乃一向度場合之直接推廣耳。

第二種一般化乃指 $X_1,\cdots\cdots,X_n$ 仍爲獨立，而皆有平均μ，但却具不同之分配。我們假想此等分配乃連續而具有密度函數 $fi(\ )$，且有有限變異數 $\sigma_i^2,\ i=1,\cdots\cdots,n$。寫出 D_n 代表 $X_1+\cdots\cdots+X_n$ 之標準差（卽 D_n 是 $\sigma_1^2+\cdots+\sigma_n^2$ 之正平方根）而考慮比率

$$\text{(2.8)}\qquad \frac{1}{D_n}\sum_{i=1}^{n}(X_i-\mu)$$

此化成 (2.5)，當變異數 σ_i^2 皆相等時。此比率 (2.8) 可證明它分配收斂于一個標準化之常態變量，若

$$\text{(2.9)}\qquad \lim_{n\to\infty}\frac{1}{D_n^2}\sum_{i=1}^{n}\left[\int_{|x|>\epsilon D_n}(x-\mu)^2 fi(x)dx\right]=0$$

對任何已知 $\epsilon>0$⑦。此條件之涵義並不明顯可知，但可宣示出它須要 $D_n\to\infty$ 及 $\sigma_n/D_n\to 0$ 當 $n\to\infty$ 時。此意味着 $X_1+\cdots\cdots+X_n$ 之總變異數 D_n^2 趨向無窮，但在此總變異數中之每一成份却只貢獻得小部份。注意 (2.9) 之涵義爲 $D_n\to\infty$, $\sigma_n/D_n\to 0$ 但其逆不成立。

第二種一般化對吾們之基本方程式的擾亂分配乃重要的。若我們想像着此些擾亂代表許多被忽略因素的聯合影響而我們假定此些因素以線性式及獨立式作用着，然後諸擾亂近似地爲常態分配，假若此些因素之數目充分大而並無某些因素能支配其他因素。

對機率收斂及分配收斂之五個命題[C]

下面列出數個有關機率及分配收斂之規則。此些規則之正確性可訴之於直覺；其中之一之證明在 2.5問題中檢討着⑧。

⑦ 閱例如 Cramér (1937, 57—61頁)。

⑧ 詳細證明閱 C. R. Rao (1965a, 101—*106*頁)。