試題反應理論的介紹(四)

——能力與試題參數的估計

英譯：(The Estimation of Ability and Item Parameters)

余民寧　著

英譯名：(Yu Min-Ning)

應用試題反應理論的方法來分析某份測驗資料的首要步驟，是估計我們所選用的試題反應模式的參數。有了滿意的模式參數估計方法，整個試題反應理論的應用，才不致有濫用與誤用等遺憾的情形發生。

前又說過，在試題反應模式裡，正確反應的機率端賴兩種因素，一爲考生的能力參數，另一爲試題參數。不論是能力或試題參數，二者都是未知的，我們唯一知道的是一群考生在一組測驗試題上的作答情形(亦即是考生們的反應組型)。因此，參數估計的問題，便成爲運用何種有效的方法，從現有的考生反應組型裡，去推估適當的考生能力參數值和試題參數值的問題。這個問題很類似於迴歸分析中估計迴歸係數的問題，唯一不同者有兩點：一爲迴歸模式通常是直線的，而試題反應模式則是非直線的；另一爲迴歸分析中的迴歸變項即(自變項)是觀察得到的，而試題反應模式中的迴歸變項（即θ變項）是觀察不到的，需要進行估計才能得知。因此，假設θ爲已知或觀察得到的變項，則試題參數的估計問題，便相當於迴歸分析中去估計迴歸係數的問題；同樣的，如果試題參數爲已知，則能力參數的估計問題亦會變得相當地簡單。本文的目的，即在討論試題參數爲已知下的能力估計，和能力參數爲已知下的試題參數的估計方法：

能力參數的估計

假設某位考生在一份具有 n 個試題的測驗上的反應組型(response pattern)爲$(U_1, U_2, \cdots, U_j, \cdots, U_n)$，其中$U_j$的值不是 1（代表正確反應），就是 0（代表不正確反應）。基於局部獨立性的假設，上述觀察到的反應組型的聯合機率(joint probability)可以說是每一個試題反應機率的連乘積，亦即

$$P(U_1, U_2, \cdots, U_j, \cdots, U_n \mid \theta) = P(U_1 \mid \theta)\ P(U_2 \mid \theta) \cdots P(U_j \mid \theta) \cdots P(U_n \mid \theta)$$

或許也可以簡化成

$$P(U_1, U_2, \cdots, U_n \mid \theta) = \prod_{j=1}^{n} P(U_j \mid \theta)$$

由於U_j的值不是 1 就是 0，所以我們可以把近似值函數(likelihood function)表示成

$$P(U_1, U_2, \cdots,\ U_n \mid \theta) = \prod_{j=1}^{n} P(U_j \mid \theta)^{U_j} [1 - P(U_j \mid \theta)]^{1-U_j}$$

或者簡化成

$$P(U_1, U_2, \cdots,\ U_n \mid \theta) = \prod_{j=1}^{n} P_j^{U_j} Q_j^{1-U_j} \quad \text{（公式一）}$$

其中$P_j = P(U_j \mid \theta), Q_j = 1 - P(U_j \mid \theta)$。

其實，公式一是某個反應組型的聯合機率的表示公式；當這個反應組型爲已知時，亦即$U_j = u_J$，這種機率的解釋方式便不再是恰當的，此時，對這種聯合機率

的表示公式便稱作近似值函數，並且記作 $L(u_1, u_2, \cdots, u_j, \cdots, u_n \mid \theta)$ ，其中 u_j 代表在試題 j 上的實得反應。因此，

$$L(u_1, u_2, \cdots, u_n \mid \theta) = \prod_{j=1}^{n} P_j^{u_j} Q_j^{1-u_j}$$ （公式二）

由於 P_j 和 Q_j 都是 θ 和試題參數的函數，近似值函數也是 θ 和試題參數的函數。

舉例來說，假設我們有五位考生和五個試題，這些考生的反應組型和試題參數都是已知，詳如表一所示。

表一　試題參數和五位考生在五個試題上的反應組型

	試題參數			考生的反應組型				
試題	a_i	b_i	c_i	1	2	3	4	5
1.	1.27	1.19	0.10	1	1	0	0	0
2.	1.34	0.59	0.15	1	0	0	1	0
3.	1.14	0.15	0.15	1	1	0	1	0
4.	1.00	-0.59	0.20	0	0	1	1	0
5.	0.67	-2.00	0.01	0	0	1	1	1

表一中的反應組型，1 代表答對該試題，0 代表答錯該試題。以第三位考生爲例，$u_1 = 0, u_2 = 0, u_3 = 0, u_4 = 1, u_5 = 1$ ，因此，這位考生的近似值函數可以公式二的表示方法表示如下：

$$L_3(u_1,u_2,u_3,u_4,u_5|\theta) = (P_1^0 Q_1^1)(P_2^0 Q_2^1)(P_3^0 Q_3^1)(P_4^1 Q_4^0)(P_5^1 Q_5^0)$$
$$= Q_1 Q_2 Q_3 P_4 P_5$$

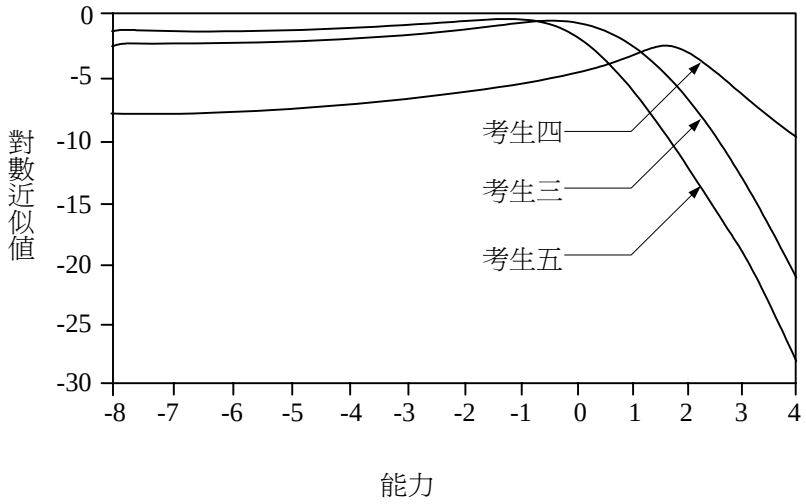
而第一位考生的近似值函數則可表示成：

$$L_1(u_1, u_2, u_3, u_4, u_5 \mid \theta) = P_1 P_2 P_3 Q_4 Q_5$$

由於 P 和 Q 都是試題反應函數，它們的數學公式端視試題參數而定（在表一中的例子是三個參數對數形模式），而在本例中的試題參數已經是已知的情形，所以針對某個固定的 θ 值，便可算出其精確的近似值函數值；我們也可以根據不同的 θ 值，畫出其相對應的近似值函數圖來。由於近似值函數是每個試題反應的機率之連乘積，而每個機率都是介於 0 與 1 之間，因此這個近似值函數的值會變得非常的小，不便於畫圖。有鑑於此，一個較好的量化方式，便是把近似值函數轉換成自然對數的形式，再來進行估計參數或畫圖。因此，公式二取自然對數後（稱作對數近似值 log-likelihood）可以寫成：

$$\ln L(u \mid \theta) = \sum_{j=1}^{n} [u_j \ln P_j + (1 - u_j) \ln(1 - P_j)]$$（公式三）

其中 u 代表試題反應的向量(vector)。根據考生能力及其相對應的對數近似值，可以圖一來表示，其中第三位考生的對數近似值在 $\theta = -0.5$ 時最高，第四位考生在 $\theta = 1$ 時的對數近似值最高，而第五位考生的對數近似值在 $\theta = -1.5$ 時最高。此時，能夠使某位考生的近似值函數（或相對應的對數近似值）達到最高點的 θ 值，便定義成該考生的 θ 的最大近似估計值（maximum likelihood estimate，簡寫成 MLE）。

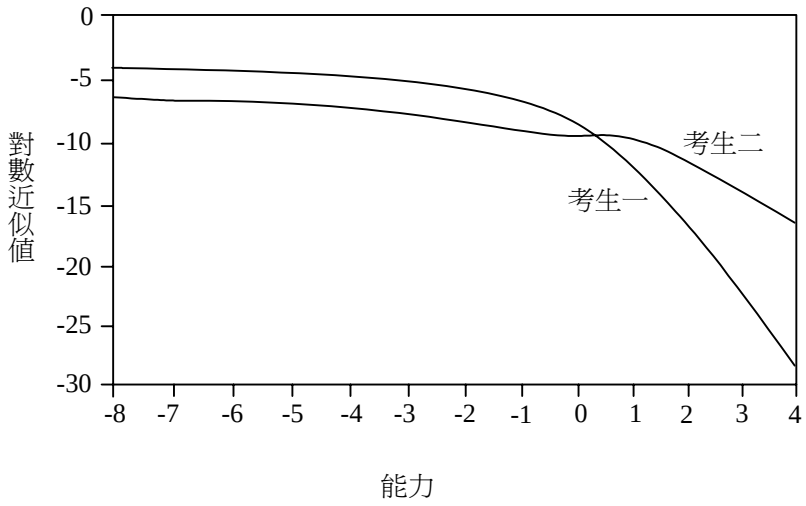


圖一　三位考生的對數近似值函數

圖一所示的圖解法，並不是找出最大近似值函數的好方法，尤其是當考生人數和試題數目增多時，這種方法更是行不通。較有效的辦法，便是利用近似值函數曲線的數學特性，亦即通過近似值函數最高點的函數斜率（以該曲線的第一階導數來代表）必定爲零。因此，我們可以利用微積分中求解函數的微分方式，把近似值（或對數近似值）函數方程式的第一階導數(first derivative)求出來，並且設定爲零，再解這個方程式中相關參數的值，便可求得這些試題參數和能力參數的最大近似估計值。由於第一階導數方程式中往往同時包含一個以上的參數，因此，最大近似估計值無法由該方程式中直接求出，我們必須再求出其第二階導數，再套入 Newton-Raphson 的遞迴估計程序(iteration procedure)，透過現成的電腦程式（如：BILOG 或 LOGIST），把參數的最大近似估計值求出，這整個估計過程可以詳見 Hambleton & Swaminathan (1985, PP. 76-88)，本文不再贅述。

可惜的是，近似值（或對數近似值）函數有時不會出現只有一個固定的最大值，這種情況尤其在考生全答對或全答錯試題時，便會產生。此時，能力參數的最大近似估計值會變成 $\theta = +\infty$ 和 $\theta = -\infty$ 。以圖二所示便可得知，第二位考生在 $\theta = 0.9$ 的點上的對數近似值最大，但其實在 $\theta = -\infty$ 的對數近似值才是真正的最大；同樣的，第一位考生的最大對數近似值出現在 $\theta = -\infty$ 的點上。因此，這兩位考生的最大近似估計值並不存在。

其實，出現上述這種現象的原因是由於這兩位考生的反應組型都是特異的(aberrant)：考生答對部份相當困難和有鑑別度的試題，卻答錯部份相當容易的試題。在這些情況下進行最大近似值估計法，最大的近似值往往無法收斂，以致無法獲得一個明確固定的最大近似估計值。像這種特異的反應組型所產生的問題，通常只出現在三個參數模式上，而不會出現在一個和兩個參數模式裡(Hambleton & Swaminathan, 1985)，有時也出現在 40 個試題以上的測驗裡。



圖二　兩個具有特異反應考生的對數近似值函數

最大近似估計值有個特殊的特性，那就是當它存在時，它具有大樣本的漸近性(asymptotic property)。由於我們所談論的只是一位考生，漸近的意思是指逐漸增加的測驗長度而言。當測驗長度增加時，θ的最大近似估計值，記作$\hat{\theta}$，會呈現以θ爲平均數的一種常態分配；這意謂著$\hat{\theta}$的漸近分配會以真正的θ值爲中心點，而呈現左右對稱的常態分配，因此，$\hat{\theta}$值在較長的測驗中是一種不偏的估計值(unbiased estimate)。$\hat{\theta}$的標準差，叫作標準誤(standard error)，記作$SE(\hat{\theta})$，是θ的一種函數，表示成

$$SE(\hat{\theta}) = 1/\sqrt{I(\theta)}$$ （公式四）

其中的$I(\theta)$叫作訊息函數(information function)，我們將留待後文再介紹它的特性及其對測驗編製的重要性。由於我們無法事先知道θ值，所以我們必須將$\hat{\theta}$值代入公式四裡的θ，才能計算出θ所對應的訊息函數值。

有了$\hat{\theta}$值等常態特性，我們也可以建立θ的信賴區間(confidence interval)。θ的$(1-\alpha)$ %的信賴區間，可以表示如下：

$$(\hat{\theta} - z_{\alpha/2}SE(\hat{\theta}), \hat{\theta} + z_{\alpha/2}SE(\hat{\theta}))$$

其中$SE(\hat{\theta})$便是在$\hat{\theta}$上的標準誤，而$Z_{\alpha/2}$是常態分配中上$(1-\alpha/2)$百分位數點；例如，95%的信賴區間的$\alpha = .05$，而$Z_{\alpha/2} = 1.96$。信賴區間可以提供研究者對θ估計值的精確性，一個參考的指標。

<u>試題參數的估計</u>

上一節所討論的是假設試題參數爲已知時，如何進行能力參數的估計。相反的，我們也可以假設能力參數爲已知時，然後來進行試題參數的估計。

假設每位考生的能力參數爲已知，我們可以針對一群考生進行一組試題的施測，然後求出N位考生在每個試題的反應的近似值函數，即

$$L(u_1,u_2,\cdots,u_N|\theta,a,b,c)=\prod_{i=1}^{N}P_i^{u_i}Q_i^{1-u_i}$$ （公式五）

其中 a, b 和 c 是試題參數（假設以三個參數模式爲例）。公式五是假設 N 位考生在每個試題上的反應是獨立的，在這個假設滿足後，公式五才算成立。

估計試題參數的方法與估計能力參數者雷同，仍然以常用的最大近似值估計法爲之：我們分別針對 a, b 和 c 參數，求出近似值函數的第一階導數，再把三個導數方程式設定爲零，再同時解出這三個非直線方程式的解；對二個參數模式而言，有兩個參數解，而一個參數模式則有一個參數解。接下來，可以 Newton-Raphson 的遞迴估計法，來求出這些方程式的解。當每位考生的能力參數爲已知時，每個試題可以分別進行估計，而不必考慮其他試題的存在。所以，估計程序必須重覆 n 次，每次估計一個試題。

其他估計方法與電腦程式

其實，在實際的估計情境中，我們往往無法事先得知能力和試題參數，因此，它們必須同時進行估計。我們可以採用上述的最大近似值估計法來進行參數的估計，這種同時進行估計能力與試題參數的最大近似值估計法，便叫作聯合的最大近似值估計法（joint maximum likelihood estimation，簡寫成 JMLE）。由於詳細的計算過程非常的繁瑣，本文不擬在此討論，有興趣的讀者可以參考 Hambleton & Swaminathan（1985，頁 129-138）。

除了聯合的最大近似值估計法外，尚有其他方法，如：邊緣的最大近似值估計法(marginal maximum likelihood estimation) (Bock & Aitkin, 1981)、條件化最大近似值估計法(conditional maximum likelihood estimation) (Andersen, 1972; Rasch, 1960)、聯合的和邊緣的貝氏估計法(Bayesian estimation) (Mislevy, 1986; Swamithan & Gifford, 1982, 1985, 1986)、啓發式估計法(heuristic estimation) (Urry, 1974)、和非直線因素分析法(nonlinear factor analysis) (McDonald, 1967, 1989)等，由於這些方法的數學公式艱深難懂，有興趣的讀者可以逕行參閱該原始文獻，本文不在此贅述。

估計能力和試題參數的過程雖然繁瑣，但站在試題反應理論的應用觀點來看，使用者不需要瞭解這些詳實的估算過程，只要知道它們如何被估計出來，並且知道如何使用它們便可以。很值得慶幸的是，目前已有數種電腦程式問世，使用者只要會使用這些程式，便可獲取能力與試題參數的估計值。有關這些電腦程式的簡介，可參見附錄一。

參考書目

Baker, F. B. (1985). The basics of item response theory. Portsmouth, NH: Heinemann.

Baker, F. B. (1987). Methodologty revieus: Item parameter estimation under the one-, two-, and three-parameter logistic models. Applied Psychological Measurement, 11, 111-142.

Hambleton, R. K., & Swaminathan, H. (1985). Item response theory: Principles and

applications. Beston, MA: Kluwer.
Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). Fundamentals of item response theory. Nowbury Park, CA: SAGE.

附錄一　目前常見的試題反應理論參數估計的電腦程式

程式名稱	來源	適用模式	估計方法	所需電腦配備	優點(＋) 缺點(－) 特性(★)
BICAL (BIGSCALE)	Wright 等 (1979)；Wright 等(1989)	一個參數	非條件化最大近似值	大多數的大電腦	(＋)廉價的 (＋)提供標準誤 (＋)提供畫圖／適合度考驗指標
RASCAL	評量系統公司(1988)	一個參數	非條件化最大近似值	個人電腦	(＋)包括適合度分析 (★)併入 Micro CAT 套裝程式裡
MICRO-SCALE	媒體交換科技公司(1986)	一個參數 多元類別	非條件化最大近似值	個人電腦	(＋)BICAL 的 PC 版 (★)資料可以放入電子試算表裡
ANCILLES	Urry(1974)	三個參數	啓發法	大多數的大電腦	(＋)廉價的 (－)常會刪除試題／考生 (－)估計方法不嚴謹 (★)不太廣泛使用
ASCAL	評量系統公司(1988)	一個參數 二個參數 三個參數	修改過的貝氏估計法	個人電腦	(＋)包括適合度分析 (＋)併入 Micro CAT 套裝程式裡 (★)使用貝氏估計法

LOGIST	Wingersky(1983)；Wingersky 等(1982)	一個參數 二個參數 三個參數	非條件化最大近似值	IBM / CDC 大電腦（第四版）	(＋)LOGIST 5 提供標準誤 (＋)具彈性，選擇多 (＋)允許未完成／空白未答的反應 (－)資料輸入繁瑣 (－)成本高 (－)難與非 IBM 相容設備聯線 (－)設定許多限制，以便獲得收斂的參數估計值
BILOG	Mislevy & Bock(1984)	一個參數 二個參數 三個參數	邊緣的最大近似值	IBM 大電腦 個人電腦版	(＋)選擇性貝氏估計值 (＋)可避免極值的估計值出現 (－)在大電腦上的執行成本很高 (－)錯誤的前置項假設會導致錯誤的估計值
NOHARM	Fraser & McDonald(1988)	一個參數 二個參數 三個參數	最小平方法	大多數的大電腦 個人電腦	(＋)適用於多向度的模式 (＋)包含殘差值分析 (－)C 參數是固定的 (★)在美國地區的使用不廣

MULTILOG	Thissen(1986)	多元類別		IBM 大電腦	(★) 把 BILOG 程式擴展成能夠處理多元類別資料的程式
MIRTE	Carlson(1987)	一個參數 二個參數 三個參數	非條件化最大近似值	IBM 大電腦 個人電腦	(＋)適用於多向度的模式 (＋)提供標準誤 (＋)包含殘差值分析 (－)C 參數是固定的
RIDA	Glas(1990	一個參數	條件化或邊緣的最大近似值	個人電腦	(＋)提供完整的考生與試題分析 (＋)處理測驗對換的不完整設計 (＋)包含適合度分析