# 不同常態轉換核平滑化無參數試題反應理論模式之蒙地卡羅模擬比較

劉湘川　　何志成　　林文質

亞洲大學生物資訊系　亞洲大學資訊工程系　亞洲大學資訊工程系

lhc@asia.edu.tw　maple0528@yahoo.com.tw　kbc@mail.ntcu.edu.tw

## 摘要

Ramsay(1991)首先提出「高低鑑別指數加權常態分位數轉」之核平滑化無參數試題反應理論模式，劉湘川(2000)以較靈敏之相關鑑別指數替代 Ramsay 之高低鑑別指數，提出改進之「相關鑑別指數加權分位數常態轉換」之核平滑化無參數試題反應理論模式。劉湘川(2007)進而以「機率值常態轉換」替代 Ramsay 之「分位數常態轉換」，提出「相關鑑別指數加權機率值常態轉換」之核平滑化無參數試題反應理論模式。本文以蒙地卡羅模擬比較，獲得結果為「相關鑑別指數加權機率值常態轉換」之模式有最佳表現，其次為「相關鑑別指數加權分位數常態轉換」之模式。

關鍵詞：高低鑑別指數、相關鑑別指數、分位數常態轉換，機率值常態轉換。

## 第一節　試題反應理論參數模式

現代測驗理論的重心是試題反應理論(item response theory;IRT)，它的特點是以機率的概念來解釋受試者能力和試題反應間之關係，也就是依據受試者的實際試題反應結果，經由理論的轉換運算，估計受試者的能力，此數學模式稱之為試題特徵函數，以圖形表示則稱為試題特徵曲線(item characteristic curve； ICC)。

試題反應理論在表達受試者能力和測驗反應間之關係上，因函數中所採用的參數個數不同，可區分為不同的模式，常用的模

式可分為參數模式和無參數模式，而參數型模式大致可分為單參數、雙參數及三參數等三種，各模式之試題特徵函數如下所示：

單參數模式

$$P_i\left(\theta_s\right)=\frac{e^{(\theta_s-b_i)}}{1+e^{(\theta_s-b_i)}} \qquad (1)$$

雙參數模式

$$P_i\left(\theta_s\right)=\frac{e^{a_i(\theta_s-b_i)}}{1+e^{a_i(\theta_s-b_i)}} \qquad (2)$$

三參數模式

$$P_i\left(\theta_s\right)=c_i+(1-c_i)\frac{e^{a_i(\theta_s-b_i)}}{1+e^{a_i(\theta_s-b_i)}} \qquad (3)$$

其中 $P_i\left(\theta_s\right)$：能力值為$\theta_s$之第$s$位受試者，其答對第$i$題的機率函數

exp：自然指數

$\theta_s$：第$s$位受試者之能力值

$a_i$：第$i$題的鑑別參數

$b_i$：第$i$題的難度參數

$c_i$：第$i$題的猜測參數

參數型試題反應理論主要以分析測驗中每一試題的難易度、鑑別度、猜測度等重要參數，再以這些參數為基礎，配合測驗目標進行組卷、施測，並將測驗結果的原始分數轉換為可代表學生真實能力的量尺分數，以估計學生之能力。

## 第二節　Ramsay「高低鑑別指數」及核平滑化法

Ramsay（1991） 引進logit函數轉換25％高低試題鑑別指數：$D_{25}$，即以高低分組通過率之差作為加權總分排序時的加權函數，原始總分排序前25％為高分組；原始總分排序後25％為低分組，分別以$P^{75}(i, j)$、$P^{25}(i, j)$ 表示第i 試題第j 選項之高分組通過率及第i 試題第j 選項之低分組通過率，$D_{25}=P^{75}(i,j)-P^{25}(i,j)$ 表示第i 試題第j 選項25％高低試題鑑別指數。則加權函數W(i, j) 如下式：

$$\begin{aligned}W(i,j)&=\mathrm{logit}\left[p^{75}(i,j)\right]-\mathrm{logit}\left[p^{25}(i,j)\right]\\&=\ln\left[\frac{p^{75}(i,j)}{1-p^{75}(i,j)}\right]-\ln\left[\frac{p^{25}(i,j)}{1-p^{25}(i,j)}\right]\end{aligned} \qquad (4)$$

洛吉數函數（logit function）：

設logit定義於開區間(0,1) 之函數，若

$$\log it(x)=\ln\frac{x}{1-x},\forall x\in(0,1) \quad (5)$$

則稱為洛吉數函數。

為方便以下介紹，將符號集中說明。

N ：表示受試者總人數。

n ：表示試題總題數。

$x_s$ ：表示第s 位受試者之測驗總分( s = 1, 2,…, N )，

則 $(x_1,x_2,x_3,...,x_N)$ 為全體受試者測驗總分之數列。$u_s(i)$：表示第s 位受試者選答第i 題填答情形之指示值，若 $u_s(i)$=1表示答對，為0則表示答錯，則 $(u_1(i),u_2(i),...,u_N(i))$ 為受試者指示值之數列(i = 1,2,…,m)。

故所得之加權總分 T 統計量定義如下式：

$$T_s=\sum_{i=1}^{m}W_i u_i^s,\quad s=1,2,...,N \qquad (6)$$

其中 $T_s$ 表受試者 $s$ 之鑑別指數加權總分值，$y_i^s$ 表受試者 $s$ 實際選答試題 $i$ 選項 $j$ 之指示值。

將 $T_s$ 值由小而大重新排序，可估得受試者 $s$ 之秩（rank）：$r_s$，再以下列機率積分轉換方式，可得標準常態分配的對應分位數（quantile）：$q_s$，$s=1,2,...,N$。

$$\int_{-\infty}^{q_s}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{t^2}{2}\right)dt=p(R\le r_s)=\frac{r_s}{N+1}, \quad (7)$$

因選答機率 $p$ 為機率估計值，故Ramsay(1991)採 Nadaray＆Watson 之 NW 核平滑化估計模式，如下

$$p_i(\theta)=\frac{\sum_{s=1}^{N}k\left(\frac{q_s-\theta}{h}\right)y_i^{(s)}}{\sum_{s=1}^{N}k\left(\frac{q_s-\theta}{h}\right)} \qquad (8)$$

其中 $s$ 表受試者：$s=1,2,...,N$，$i$ 表試題：$i=1,2,...,n$， $y_{ij}^{(s)}$ 表加權排序後，第 $r_s$ 序位受試者實際選答試題 $i$ 之指示值，$q_s$ 表第 $r_s$ 序位受試者加權總分經機率積分轉換之分位數。上式中 $h$ 表帶寬參數(bandwidth parameter)亦稱平滑參數(smoothing parameter)，Ramsay(1991)

採 $h=1.1N^{-0.2}$ ，其中 $N$ 為受試者人數，此外 Ramsay 選高斯函數（Gaussian function）為專有之核函數

$$k(u)=\exp\left[-\frac{u^2}{2}\right]\ ,\ -\infty<u<\infty \qquad (9)$$

綜合（6）（7）（8）三式，即得核平滑化無參數試題選項特徵曲線機率模式，如下

$$\hat{p}_i(\theta)=\frac{\sum_{s=1}^{N}\exp\left[-\frac{N^{\frac{2}{5}}(q_s-\theta)^2}{2.42}\right]y_i^{(s)}}{\sum_{s=1}^{N}\exp\left[-\frac{N^{\frac{2}{5}}(q_s-\theta)^2}{2.42}\right]} \qquad (10)$$

其中 $s$ 表受試者：$s=1,2,...,N$ ，$i$ 表試題：$i=1,2,...,m$ ，$y_i^{(s)}$ 表加權排序後第 $r_s$ 序位受試者實際選答試題 $i$ 之指示值，$q_s$ 表第 $r_s$ 序位受試者加權總分經機率積分轉換之分位數。

## 第三節　點二相關鑑別指數

劉湘川（民 90）提「點二系列相關鑑別指數」，在能力參數之估計上，特別引進隨機未作答虛擬選項，進行合併估計，因而能兼顧隨機未作答不完全資料之情況。「點二系列相關鑑別指數；$r_i$ 」之定義如下：

假設受試者有 $N$ 人（ $s=1,2,...,N$ ），試題有 m 題（ $i=1,2,...,m$ ）。

$r_i$ 表 $(x_1,x_2,...,x_N)$ 與 $(u_i^1,u_i^2,...,u_i^N)$ 之積差相關係數，稱為「點二系列相關試題鑑別指數」，即

$$r_i=\frac{\sum_{s=1}^{N}(x_s-\bar{x})(u_i^s-\bar{u}_i)}{NS_xS_{u_i}} \qquad (11)$$

其中 $x_s$ 表受試者 $s$ 之測驗總分，$u_i^s$ 表受試者 $s$ 是否實際選答試題 $i$ 選項 $j$ 之指示數值。

$$\bar{x}=\frac{1}{N}\sum_{s=1}^{N}x_s \quad , \quad S_x^2=\frac{1}{N}\sum_{s=1}^{N}(x_s-\bar{x})^2$$

$$\bar{u}_i=\frac{1}{N}\sum_{s=1}^{N}u_i^s \quad , \quad S_{u_i}^2=\frac{1}{N}\sum_{s=1}^{N}(u_i^s-\bar{u}_i)^2$$

因 $-1\le r_i\le 1$　取計分加權值

$$W_i=\frac{1+r_i}{2} \quad , \quad 0\le W_i\le 1$$

加權總分 T 統計量定義如下式：

$$T_s=\sum_{i=1}^{m}W_i u_i^s,\quad s=1,2,\ldots,N \qquad (12)$$

其中 $T_s$ 表受試者 $s$ 之相關加權總分值，$u_i^s$ 表受試者 $s$ 實際選答試題之指示值。

## 第四節　核平滑化無參數 IRT 模式之常態轉換改進估計

劉湘川（2007）提「核平滑化無參數 IRT 模式之常態轉換改進估計」,將原本之相關鑑別指數常態化，使得新的相關係數能介於 0 到 1 之間

$$r_i \Rightarrow \beta_i=\frac{1+r_i}{2},w_i=\frac{\beta_i}{\sum_{l=1}^{n}\beta_l} \qquad (13)$$

再以新的相關係數乘上答題指示值，求得總分，經過類似 Fisher 轉換之方法轉換後得到初步能力估計值

$$\begin{aligned}&Z_s=\sum_{i=1}^{n}w_i u_i^s,0\le Z_s\le 1\\&\Rightarrow \theta_s^{(0)}=\frac{1}{2}\ln\frac{Z_s}{1-Z_s}\end{aligned} \qquad (14)$$

接著以所得之估計值代入核平滑化公式求得答對之機率

$$\left[\theta_s^{(0)},u_i^s\right]\underset{\text{k.s}}{\Rightarrow} P_i^{(0)}\left(\theta_s^{(0)}\right) \qquad (15)$$

將每位受試者每題之答對機率平均並經由 Fisher 轉換後得最後之能力估計值，

$$\begin{aligned}&P^{(0)}\left(\theta_s^{(0)}\right)=\frac{1}{n}\sum_{i=1}^{n}P_i^{(0)}\left(\theta_s^{(0)}\right)\\&\Rightarrow \theta_s^{(1)}=\frac{1}{2}\ln\frac{P^{(0)}\left(\theta_s^{(0)}\right)}{1-P^{(0)}\left(\theta_s^{(0)}\right)}\end{aligned} \qquad (16)$$

## 第五節　實證研究

一、利用電腦隨機選取模擬參數值

產生平均數為 2 標準差為 0.5 之標準常態分配隨機亂數 a，做為試題之鑑別度。

產生標準常態分配介於 2 到-2 之隨機亂數 b，做為試題之難度。

產生均勻分配介於 0 到 0.25 之隨機亂數 c，做為猜測參數。

產生標準常態分配 N（0,1）之隨機亂數 $\theta_s$ ，設定為受試者的能力值。

二、利用步驟一設定之參數，使用三參數模式模擬答題者之答題情形：

第 s 位受試者答對第 i 題之機率函數為

$$P_i(\theta_s)=c_i+(1-c_i)\frac{e^{a_i(\theta_s-b_i)}}{1+e^{a_i(\theta_s-b_i)}} \quad (17)$$

三、利用均勻分配 U（0,1）隨機產生亂數 r，此時若 r 落於區段[0, $P_i(\theta_s)$]則代表受試者答對此題，答題指示值為 1，若 r 落於區段[$P_i(\theta_s)$，1]則代表受試者答錯此題，答題指示值為 0。

四、利用步驟三所得之答題指示值矩陣，當做原始分數，計算其總分並排序後，依公式分別求取「高低鑑別指數核平滑化無參數估計值」、「點二相關鑑別指數核平滑化無參數估計值」及「核平滑化無參數常態轉換改進估計值」，最後將三者之能力估計值與原先模擬之真實能力值求取 MSD 值，以比較何者有較佳之估計值。

五、本研究針分別就「擴張高低鑑別指數核平滑化無參數估計值」、「點二相關鑑別指數核平滑化無參數估計值」及「核平滑化無參數常態轉換改進估計值」三種不同模式下，對各種人數(400、800、1200 人)與題數均為 25 題之三種組合樣本進行能力值之估計分析。每組情形都模擬 10 次。下表為每一組皆模擬實驗 10 次之平均值的結果分析，以下即針對模擬資料所得結果進行分析說明：

表6.1高低鑑別指數、相關鑑別指數及改進之常態轉換估計能力值MSE平均值比較

| 模式 平均 取樣 | 高低鑑別指數 | 相關鑑別指數 | 常態轉換改進估計 |
|---|---|---|---|
| 400人25題 | 0.1566 | 0.137676 | 0.122439 |
| 800 人 25 題 | 0.1632 | 0.145734 | 0.09154 |
| 1200 人 25 題 | 0.158894 | 0.149437 | 0.076651 |

上表5-1之模擬資料所得結果可以看出，在估算受試者能力值時，採用「點二相關鑑別指數」、「擴張高低鑑別指數」及「常態轉換改進估計」三種模式做比較時，得到「常態轉換改進估計」模式估得之能力值與受試者真實能力值二者間之均方差小於「點二相關鑑別指數」模式，而「點二相關鑑別指數」模式估得之能力值與受試者真實能力值二者間之均方差又較「擴張高低鑑別指數」模式為小，亦即三者之加權模式所得之估計精準度為「常態轉換改進估計」最佳，「點二相關鑑別指數」其次，而「擴張高低鑑別指數」則最差。表6-1所得結果以折線圖表示如下(圖6-1)：

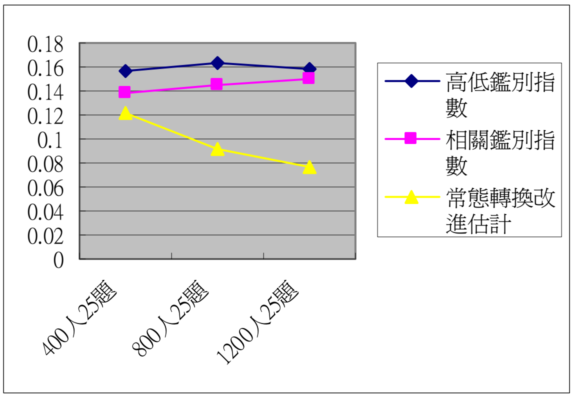


圖 6.1　擴張高低鑑別指數與點二相關鑑別指數估計能力值 MSE 比較

## 第六節　結論

「擴張高低鑑別指數」、「點二相關鑑別指數」「核平滑化無參數 IRT 模式之常態轉換改進估計」、,均為試題反應理論中為達成更佳的區辨力而被提出的加權法則，但「擴張高低鑑別指數」僅取前後25%的資料來做為求得加權值之依據，喪失了中間 50%受試者的作答訊息，故在做為能力估計值的預測上，無法反應所有受試者真實的答題情形，而「點二相關鑑別指數」，針對此一缺失加以改善，將所有受試者的答題組型均列入考慮，所以能夠較精確的估算出受試者的真實能力值，另「核平滑化無參數 IRT 模式之常態轉換改進估計」在相關係數的基礎上加以改善，故有更好的估算效果，在以上的模擬測試中已獲得實證。

## 參考文獻

[1] 劉湘川（2000）。點二系列相關試題鑑別指數之值譜分析及其在 IRT 之應用測驗。統計年刊第 8 輯，1-20 頁。台中市，國立台中師範學院。

[2] 劉湘川（2001）。核平滑化試題選項特徵曲線與選項關連結構整合擴充模式。統計年刊第 9 輯，1-18 頁。台中市：國立台中師範學院。

[3] 劉湘川（2002）。高階相關比累進加權核平滑化試題選項分析綜合模式。統計年刊第 10 輯，197-218 頁。台中市，國立台中師範學院。

[4] 劉湘川 （2003）。核平滑化試題選項分析模式之條件最大概似數值估計。統計年刊第 11 輯，17-40 頁。台中市，國立台中師範學院。

[5] 劉湘川 （2007）。試題反應理論講義。臺中縣霧峰鄉，亞洲大學資訊工程研究所，資訊教育組。(未出版)

[6] Ramsay, J. O. (1991). Kernel smoothing approaches to nonparametric item characteristic curve

estimation. Psychometrika,56,611-630.