Bayesian CAT-SoFun

內容

[題庫 (Item Bank) 2](#_Toc437010892)

[計分向量與設計向量 2](#_Toc437010893)

[測量模式 (Measurement Model) 2](#_Toc437010894)

[功能估計 (Ability Estimation) 3](#_Toc437010895)

[適性選題 (Adaptive Item Selection) 6](#_Toc437010896)

本文件首先簡述SoFun測驗之題庫概況、介紹用來分析資料的測量模式，進而說明CAT-SoFun在「功能估計」與「適性選題」兩種程序之算則。

# 題庫 (Item Bank)

SoFun測驗的所有試題皆測量單一建構，為單向度測驗，總題數為24題，每題皆有5個計分類別（1~5分）。利用ConQuest針對預試資料進行PCM分析後，得到24個試題整體難度以及96個階難度，詳細的試題參數數值請參考**環境設定.xlsx**。

## 計分向量與設計向量

由上述題庫描述，可以分別建構出大小為120\*1的**計分矩陣**（24\*5=120種試題\*計分類別配對，1個能力向度)與大小為120\*96的**設計矩陣**（120種試題\*計分類別配對，24個試題整體難度+72個階難度)，在計分矩陣與設計矩陣中每個元素的係數根據量表計分規則加以設定，詳細的元素數值請參考**環境設定.xlsx**。

# 測量模式 (Measurement Model)

若將個案在*p*個向度的能力表示為**θ**={θ1, θ2, …, θ*p*}，此能力向量分別影響個案在某一試題或某幾道試題之作答。由於SoFun測驗為多元計分題型，採用Multidimensional Random Coefficients Multinomial Logit Model (MRCMLM, Adams, Wilson, & Wang, 1997)來描繪個案的*p*個能力值、試題的難度參數，以及該題之作答反應之關係，其數學函數如公式(1)所示。其中，

 (1)

*Pnik*為第*n*位個案在第*i*題得到*k*分之機率

*Uni*為第*n*位個案在第*i*題之作答反應

*Ki*為第*i*題的計分類別數

**θ***n*為第*n*位個案之能力向量

**ξ**為試題參數向量

**b***ik*為第*i*題中得*k*分的計分向量

**a***ik*為第*i*題中得*k*分的設計向量

一旦將能力向量(**θ***n*)、試題參數(**ξ)**、計分向量(**b***ik*)以及設計向量(**a***ik*)的數值代入公式(1)，便能算出第*n*位個案在第*i*題得到*k*分之機率，進而預測其可能的作答反應。實務上，雖然**ξ**、**b**以及**a**的數值在建置題庫的過程中便已可獲得，然而個案的功能程度通常是未知的。CAT-SoFun的主要功能便是根據公式(1)建立的函數關係，蒐集個案在多個試題上的作答資料，反向估算個案的功能程度。

為了能精準有效的評估個案功能，CAT-SoFun內建**適性選題算則**以及**功能估計算則**。適性選題算則能從題庫中抽選最適合個案功能程度的試題進行施測，在每施測一道試題後，功能估計算則便依據個案作答資料列出概似機率函數(likelihood function)，如公式(2)所示，使得此函數最大化的數值，便是個案功能程度的最佳估計值。

 (2)

其中，*v*為被選到的試題集合，*Pnik*(**θ**)可經由公式(1)計算而得。

在實務上，多採用貝氏估計取向以克服最大概似取向的缺點，因此CAT-SoFun分別採用Maximum a Posterior (MAP)法以及Maximum Fisher Information法，進行功能估計與適性選題，詳細說明以下。

# 功能估計 (Ability Estimation)

貝氏估計法，有別於最大概似估計法，同時採用個案能力的事前分布(prior distribution)、以及其作答向量構成的概似函數，建構出能力事後分布(posterior distribution)。根據貝氏定理，能力事後分佈可以表示如下

 (3)

*g*(**θ**)為能力事前分佈，若假設能力向量服從平均數向量為**μ**、變異共變數矩陣為**Σ**之多變量常態分佈，則*g*(**θ**)之機率密度函數可表示如下

 (4)

*f*(**u**)為**u**的邊際機率，針對函數*L*(**u**|**θ**) *g*(**θ**)從-∞~∞之能力區間進行積分而得。

貝氏能力估計法，根據對能力事後分布所進行運算之差別，可分為兩種 (Mislevy, 1984)：Maximum a Posterior (MAP)以及Expected a Posterior (EAP)。MAP能力估計值(modal estimates)，為針對能力事後分布進行最大值(mode)計算後而得；EAP能力估計值，為針對事後分布進行期望值(expected value)計算後而得。在單向度情境下，兩種方法所耗費的運算時間相去不遠，然而當向度個數增加時，EAP法的運算量呈指數倍增加，MAP法的運算量因而低於EAP法(Segall, 1996)，因此，CAT-SoFun採用MAP法進行功能估計。

多向度MAP能力估計值，為使得能力事後分布有最大值的一組數值，其估計程序可分為以下幾個步驟：

1. 設定*f* (**θ**)、根據個案作答反應向量列出*L*(**u**|**θ**)，並利用公式(3)建構出*f* (**θ|u**)。

2. 對 *f* (**θ|u**)進行自然對數計算，建構出ln *f* (**θ|u**)，如下所示

 (5)

3. 對ln *f* (**θ|u**)進行一階偏微分並令其為0，如公式(6)，這*p*個方程式的解即為。

 (6)

將公式(5)代入公式(6)可以得到

 (7)

若向度個數為2，則，

在公式(7)中，第*r*個方程式為

 (8)

若測量模式為Rasch model，公式(8)可簡化表徵如下

 (9)

在CAT-SoFun中的測量模式為MRCML model，公式(8)重新表徵如下

 (10)

其中，*bik*為第*i*題第*k*類別在第*r*向度上的計分係數

*Ei*(**θ**)為個案在第*i*題第*r*向度上的期望得分。

由於公式(6)沒有封閉解，所以利用數值方法進行疊代求取近似解。Newton-Raphon法為求近似解的常用方法之一，若在第*j*次疊代過程中計算得到的能力向量為**θ**(*j*) 數值變化量向量為**δ**(*j*)，將兩者代入公式(11)後可計算出第*j*+1次疊代的能力向量。

 (11)

**δ**(*j*)可根據公式(12)計算而得。

 (12)

其中，**J**(**θ**)為針對ln *f* (**θ|u**)進行二階微分所構成的*p*×*p*對稱矩陣。

 (13)

當測量模式為MRCML model，公式(13)可重新表徵如下

 (14)

令

則**J**(**θ**)的第*r*個對角線元素可表示如下

 (15)

非對角線元素可表示如公式(16)

 (16)

一旦計算出第*j*次疊代的數值變化量向量(**δ**)，便能得到第*j*+1次疊代的能力向量，最終重複多次疊代程序直至**δ**的每個元素值，皆小於事前訂定的數值(如，0.001)方終止疊代。

# 適性選題 (Adaptive Item Selection)

多向度適性測驗之選題策略為：選擇「能最大幅度降低多變量常態橢圓體體積」之試題，作為下一道施測試題。Segall (1996)推導出訊息量矩陣之行列式值，與該體積之下降幅度有關。從推論結果來看，若將已經施測的*S*道以及額外施測第*S*+1道試題後得到的訊息量矩陣分別表示為以及，使得兩者相加之後的訊息量矩陣具有最大行列式值者，即為最佳的下一個施測試題(Segall, 1996)。

當測量模式為MRCML model，已經施測*S*道試題對能力為的個案提供的訊息量，可利用公式(17)進行計算

 (17)

在單向度適性測驗，訊息量矩陣縮減成訊息純量，當已經施測*S*道試題後可以根據個案的暫時能力估計值，計算每道試題提供的訊息量後進行累加計算，進而據此測驗訊息量估算出能力估計標準誤。已知，意即當測驗訊息量越高則估計標準誤越小，越能縮減能力估計值之信賴區間，當能力值可能變動程度越小、表示測量越精準。若能在尚未施測試題集合中選取具有最大訊息量者，一旦將該題訊息量累加至現有測驗訊息量中，將使得新的累加訊息量維持最大化、也能最大幅度降低估計標準誤。因此，具有最大試題訊息量之未施測試題，為下一道施測試題之最佳依據 (Lord, 1980, p.72)。