

「文化共識理論」(cultural consensus theory, CCT) 是一以認知模式為本之分析方法，由學者 Batchelder 及其同事於 1980 年代發展出。此方法旨在分析群組人士在文化影響下對於某些知識或意見之共識，特別是當正確答案屬未知狀態。此方法最初發展僅限用於文化情境底下二分回答選項之資料，爾後 Batchelder 及其同事結合信號偵測理論及測驗理論的一些元素，將之延伸至可處理排序型類別資料(包括李克特量表)。本研究在 CCT 架構下提出另一路徑，發展一結合反應信心於閾值理論的分析方法處理排序型類別資料。我們藉電腦模擬評估此法之可行性。

潛在真值模型 (Latent Truth Model, LTM)

我們假設有 N 位受試者 (informants or subjects) 將回答 M 個題項的測驗，而該測驗沒有先驗的正確答案 (answer key)。在 CCT 架構下我們考慮以下情境，首先，每一題項有其潛在的「文化真值」(cultural truth) $\mathbf{T} = (T_1, \dots, T_M)^\top$ ，其中 $T_k \in [0, 1] \forall k$ 。另外，對每一題受試者被限制只能回答「1」/「0」(或「是」/「否」，「真」/「假」)。因此整套問卷施測完所有受試者後，研究者可得到一個回答矩陣 $\mathbf{X} = (X_{ik})_{N \times M} \in \{0, 1\}^{N \times M}$ 。Batchelder and Anders (2012) 將此種設計稱為「潛在真值模型」(Latent Truth Model, LTM)。¹

具體而言，LTM 主要由三類參數所組成：(1) 前述提到過的文化真值 \mathbf{T} ；(2) 受試者的文化能力 (cultural competence) $\mathbf{E} = (E_i)_{N \times 1}$ 且 $E_i > 0 \forall i$ ；以及 (3) 偏誤閾值 (bias threshold) $(B) = (b_i)_{M \times 1}$ 且 $b_i \in [0, 1] \forall i$ ，其公理定義如下：

公理 1 (文化真值). 每一道題項都有一個文化真值參數 $T_k \in [0, 1], k = 1, \dots, M$ 適用於所有的受試者。

公理 2 (條件獨立). 每位受試者對於每道題目都會有個潛在真值 (*latent truth*) 答案 Y_{ik} ，且所有人對於每道題目的潛在真值隨機矩陣 $\mathbf{Y} = (Y_{ik})_{N \times M}$ 在其他的參數固定下乃有條件獨立，亦即，隨機矩陣 \mathbf{Y} 在可能的實現值 $(y_{ik})_{N \times M}$ 下，其聯合機率分配 $h(\cdot)$ 為

$$h((y_{ik}) | \mathbf{T}, \mathbf{E}) = \prod_{i=1}^N \prod_{k=1}^M g(y_{ik} | T_k, E_i) \quad (1)$$

其中 $g(y_{ik} | T_k, E_i)$ 為潛在真值 Y_{ik} 在給定的相關參數下之邊際反應分配。

¹ 附帶一提，若將文化真值設定為類別變項則稱為「廣義孔多賽模型」(General Condorcet Model, GCM)，為 Batchelder 團隊先前研究中頗受歡迎的模型 (Batchelder & Anders, 2012; Batchelder & Romney, 1986, 1988; Romney et al., 1987; Weller, 2007)。

公理 3 (回答的邊際分配). 式 (1) 中的邊際機率密度函數為

$$Y_{ik} \sim Beta(T_k E_i, (1 - T_k) E_i) \quad (2)$$

公理 4 (反應偏誤). 每位受試者都有自己的偏誤閾值 (*bias threshold*)

$B = (b_i)_{N \times 1} \in [0, 1]^N$ ，他們在回答該道題項的兩種選項中的選擇機率為

$$\Pr(X_{ik} = x_{ik} | Y_{ik}, b_i) = \begin{cases} \Pr(Y_{ik} > b_i) & \text{if } x_{ik} = 1 \\ \Pr(Y_{ik} \leq b_i) & \text{if } x_{ik} = 0 \end{cases} \quad (3)$$

Batchelder and Anders (2012) 對於 LTM 的設定中，公理 1 說明題項的文化真值 T_k 必須得落在 $[0, 1]$ 之間，對於同一批的受試者 (在相同文化下的人們) 而言，此一文化真值代表其共識。公理 2 假定受試者知覺到題項內容後，會產生一個潛在的真值隨機變數 Y_{ik} (研究者觀察不到的)，決定其對該題項的回答 X_{ik} (研究者可觀察到的)。此外，LTM 也假定在其他參數給定下，每一題得到的潛在真值是相互獨立的。換句話說，受試者對於每道題項得到的潛在真值不受到其他題項的潛在真值或是其他受試者的回答影響。

公理 3 假設潛在真值的機率分配為 Beta 分配。若一個隨機變數 $Z \sim Beta(\alpha, \beta)$ (其中 $\alpha, \beta > 0$ 為分配之參數)，則其機率密度函數為

$$f(z | \alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} z^{\alpha-1} (1 - z)^{\beta-1}$$

其中， $0 \leq z \leq 1$ 以及 $\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt \forall x > 0$ 為 Gamma 函數。Beta 分配的期望值與變異數分別為

$$E(Z) = \frac{\alpha}{\alpha + \beta}, \quad \text{Var}(Z) = \frac{E(Z)(1 - E(Z))}{\alpha + \beta + 1}$$

而公理 3 進一步將潛在變數 Y_{ik} 之分配重新參數化，使之成為具有模型意義的參數變項 T_k 和 E_i 組成，其中 $\alpha_{ik} = T_k E_i$ 以及 $\beta_{ik} = (1 - T_k) E_i$ 。透過這樣的設計，我們可看出 Y_{ik} 的期望值剛好就是 T_k ，而變異數則為 $T_k(1 - T_k)/(1 + E_i)$ 。亦即，受試者的文化能力 E_i 扮演者精確參數 (precision parameter) 的角色，當受試者的文化能力 E_i 越高，他對於該題項的潛在真值 Y_{ik} 也就會集中在該題項文化真值 T_k 附近。此外，由於題項回答的設計僅限於二分變項，公理 4 假設受試者會回答「1」如果其潛在真值 Y_{ik} 超過該受試者心中的閾值 b_i ；反之，若其潛在真值 Y_{ik} 小於閾值 b_i 的話，則會回答「0」。

結合式 (2) 和式 (3)，我們可寫出某位受試者對於某題項回答為「1」之概似函數 (likelihood function)：

$$\begin{aligned} L(X_{ik}, T_k, E_i, B_i) &= \Pr(X_{ik} = 1 | T_k, E_i, b_i) \\ &= 1 - F(b_i | T_k, E_i) \quad \forall i, k. \end{aligned} \quad (4)$$

其中， $F(.|T_k, E_i)$ 表示 Beta 分配的累積密度函數 (cumulative distribution function)。從上式可推出在 LTM 設定下，所有受試者對於每道題項回答的概似函數為

$$\begin{aligned} L(\mathbf{T}, \mathbf{E}, \mathbf{B}) &= \Pr(\mathbf{X}|\mathbf{T}, \mathbf{E}, \mathbf{B}) \\ &= \prod_{i=1}^N \prod_{k=1}^M [1 - F(b_i|T_k, E_i)]^{x_{ik}} F(b_i|T_k, E_i)^{1-x_{ik}}. \end{aligned} \quad (5)$$

LTM 的精髓在於，從 (觀察到的) 受試者對題項的回答矩陣 \mathbf{X} ，我們可以估計出題項的文化共識真值 \mathbf{T} 、受試者的文化能力 \mathbf{E} 、以及選擇偏誤 \mathbf{B} 。

潛在真值信心模型 (Latent Truth Confidence Model, LTCM)

在 LTM 公理 4 中，Batchelder 等人刻劃每位受測者在感知題項到決策的歷程中，有一個潛在偏誤閾值 (bias threshold) b_i 決定了兩個回答選項的反應機率為何。然而在過去文獻中，有許多證據顯示受試者對於選項的選擇機率，會受到與題項內容無關的個人偏誤影響，例如實驗的指導語、獎金酬賞、個人動機等。而這些因素會被認為會影響受試者的回答閾值 (response threshold) 或回答準則 (response criteria)，但不會影響到與題項內容有關的感知閾值 (sensory threshold)。這兩種不同類型的偏誤應可在 LTM 區分出。Luce (1963) 發展的 (兩狀態) 低閾值理論 (two-state low-threshold theory) 即假定，當受試者接收到實驗刺激 (「訊號」或「訊號加雜訊」) 時，會先進入「偵測到」或是「沒偵測到」刺激這兩種狀態，且進入這兩種狀態分別的可能性乃受到刺激本身與感知閾值決定。接續，這兩種狀態分別對回答選項 (「有訊號」與「沒有訊號」) 有各自的選擇機率，其取決於回答閾值的大小而有不同。此模型說明了即使某位受試者感知上是在「有偵測到」刺激的狀態，但他還是有可能在回答選擇時，選擇「沒有訊號」的選項；反之亦然。閾值理論在後續的發展中，也考量多個信心程度不同的回答選項，其在心理物理學上的展延可參考 Krantz (1969) 以及 Hsu and Doble (2015)。

LTM 之延伸模型

我們將 Luce (1963) 發展的 (兩狀態) 低閾值理論概念帶入 LTM，建構含信心程度的回答選項，使模型能處理排序型類別資料 (包括李克特量表)。我們修改 LTM 的公理 4 和加入新的公理，將原先的偏誤閾值區分出感知閾值所造成的不同心理狀態，以及加入回答閾值的設定來描述信心程度回答選項的選擇機制，並將此延伸的模型命名為「潛在真值信心模型」(Latent Truth Confidence Model, LTCM)。以下描述其使用的情境與公理。

我們假設每位受試者在填答每道題項時，可從 J 個次序選項類別中擇一個 $j \in \{1, \dots, J\}$ 當作回答，其中這 J 個選項反應出對答案是否為在 0 到 1 之間的信心程度，其中最大值的選項 J 表示「對題項 k 答案是『1/真』有最大的信心」(是『0/假』有最低的信心)，選項 $J - 1$ 則表示「對題項 k 答案是『1/真』有次大的信心」(是『0/假』有次低的信心)，…，選項 1 則表示「對題項 k 答案是『1/真』有最低的信心」(是『0/假』有最大的信心)。若 J 為奇數的話，正中間選項 $\lceil \frac{J}{2} \rceil$ 可視為「不確定對題項 k 答案是『1/真』或是『0/假』」。對此，研究者可得受試者的回答隨機矩陣 $\mathbf{X} = (X_{ik})_{N \times M}$ 的某次實現值 (realization)

$$X_{ik} = j \quad \text{if informant } i \text{ assigns category } j \text{ to item } k. \quad (6)$$

我們依照公理 1 假設題項的文化真值是介於 0 到 1 之間，但對於受試者回答則不同於 LTM 中只能回答 (極端的) 二分選擇，LTCM 允許在 0 ~ 1 之間有不同次序大小信心程度的回答選項。據此，我們將閾值理論引入公理 4，將之拓展，其延伸之公理設定如下：

公理 4' (狀態偏誤). 每位受試者都有自己的感知閾值 (*sensory threshold*) $\mathbf{B} = (b_i)_{N \times 1} \in [0, 1]^N$ 。當接受題項的內容後，受試者會有機率進入到覺得該題項答案是「1」(或是「0」) 的狀態 S_{ik} ，該機率為

$$\Pr(S_{ik} = s_{ik} | Y_{ik}, b_i) = \begin{cases} \Pr(Y_{ik} > b_i) & \text{if } s_{ik} = 1 \\ \Pr(Y_{ik} \leq b_i) & \text{if } s_{ik} = 0 \end{cases} \quad (7)$$

公理 5' (回答準則). 在題項有 $J \geq 2$ 個回答信心選項類別時，受試者在兩種狀態 ($S_{ik} = 1$ 或 0) 下，分別對於 J 種信心程度選項有各自的回答準則 (*response criterion*)， $\mathbf{D} = (d_{ij})_{N \times J}$, $d_{ij} \in [0, 1] \& \sum_{j=1}^J d_{ij} = 1$ 和 $\mathbf{G} = (g_{ij})_{N \times J}$, $g_{ij} \in [0, 1] \& \sum_{j=1}^J g_{ij} = 1$ ，且該機率與題項的內容是無關的。因此，此受試者在某題項回答第 j 個信心程度選項時的機率為

$$\Pr(X_{ik} = j | S_{ik}, d_{ij}, g_{ij}) = \Pr(S_{ik} = 1) d_{ij} + \Pr(S_{ik} = 0) g_{ij} \quad (8)$$

其中 $j = 1, \dots, J$ 。

值得注意的是，公理 5' 的設計使得 LTCM 比 LTM 增加了一些參數，其個數多於資料量所能提供的自由度，因此使得模型具有不可辨識性 (non-identifiability)，對此我們設

定一些限定條件使得模型是可辨識的²。Krantz (1969) 對於信心回應的低閾值模型，提出一個合理的「**限制 1**」： $\frac{d_j}{g_j}$ 對於 j 是單調遞增 (monotonic increasing) 的函數。換句話說， $\frac{d_j}{g_j}$ 會隨著 j 而增加 (對答案「1」的回答信心程度增加)。另外，我們給與更進一步的「**限制 2**」： $d_{i1} = g_{iJ} = 0 \forall i$ 。這是一個合理的限制，原因在於如果受試者已經是在狀態 0 的情況下，他不應該會選擇最肯定 (信心程度最高) 答案是 1 的選項，也就是第 J 個選項，因此 $d_{iJ} = 0$ ；同理，受試者在狀態 1 的情況下，也不該選擇最肯定答案是 0 的選項，也就是第 1 個選項， $g_{i1} = 0$ 。最後，當 $J = 2$ 時，LTCM 實際上等同於 LTM，雖然我們區分了狀態與選擇兩階段的歷程，但在上述的「**限制 2**」下，只要受試者進入狀態 0，他就只會選擇對 0 信心程度最高的「選項 1」；反之，若受試者是進入狀態 1，則他也只會選擇對 1 信心程度最高的「選項 2」。

LTCM 之層級式貝氏模型與分析

近幾年來階層模型 (hierarchical model) 與貝氏推論 (Bayesian inference) 作為模型的建構與統計推論，在各個不同領域中越來越多研究者採用。心理學的認知模型也開始有學者介紹使用階層貝氏模型所帶來的好處 (Lee, 2011; Lee & Wagenmakers, 2014; Rouder & Lu, 2005)。對於認知模型有階層式的假設是假定於每位受試者或題項中所關心的變量，都是服從於某一特定分配族 (distribution family)，能解釋個人層次變量之間的隨機效果 (random effect)。而貝氏推論會將這些分配族帶有的參數進一步假設有某個先驗分配 (prior distribution)，納入研究者對於參數可能發生值的想法，在藉由收集到的資料資訊和貝氏定理推導，最終能得到模型參數值的各種發生值的可能性強度，稱之為後驗分配 (posterior distribution)。有了參數後驗分配後便可進行統計推論，包括點估計、區間估計、模型預測與比較 … 等 (Gelman et al., 2013; Hoff, 2009; Kruschke, 2014)。研究者通常在模型建立時便已有參數與選定分配之假設，而貝氏統計須指定參數的先驗分配為何以及說明設定上的合理性。

² 過去文獻中，不同閾值理論對於 $D&G$ 的限制不盡相同。早在 Luce (1963) 的兩階段閾值模型中，對於二分「是/否」選項時，限制 $g_1 = 0$ 或是 $d_1 = 1$ 。另外也有學者提出別的限制，對於每個選項 j 假設 $g_j = 0$ 或 $d_j = 0$ (Broadbent, 1966; Watson & Bourbon, 1965)，或是允許有部分重疊的，譬如在五個反應選項中限制 $d_1 = d_2 = g_4 = g_5$ (Nachmias & Steinman, 1963) 等。詳細限制式的探討可參考 Krantz (1969)。

我們依據公理 1-3、4' 和 5' 所建立的 LTCM 的階層模型假設如下：

$$T_k \sim Beta(\mu_T \tau_T, (1 - \mu_T) \tau_T)$$

$$E_i \sim Gamma(\mu_E^2 / \sigma_E^2, \mu_E / \sigma_E^2)$$

$$b_i \sim Beta(\mu_b \tau_b, (1 - \mu_b) \tau_b)$$

$$\mathbf{d}_i \sim Dirichlet(\boldsymbol{\alpha}_d)$$

$$\mathbf{g}_i \sim Dirichlet(\boldsymbol{\alpha}_g)$$

其中 $i = 1, \dots, N$ 、 $k = 1, \dots, M$ 、 $j = 1, \dots, J$ 。我們以粗體字來表示行向量 (column vector)，包括： \mathbf{d}_i 、 \mathbf{g}_i 、 $\boldsymbol{\alpha}_d$ 、& $\boldsymbol{\alpha}_g$ 。由於 T_k 與 b_i 可能值的範圍在 $[0, 1]$ 之間，Beta 分配是一個很適當的選擇，其中先驗分配中參數化的方式也如同式 (2) 般，利用平均參數 (μ_T & μ_b) 與精確參數 (τ_T & τ_b) 來刻劃。另一方面， E_i 可能的範圍為 $[0, \infty)$ ，且 E_i 做為式 (2) 中 Beta 分配的精確參數，因此 Gamma 分配常會是此先驗的考量 (Kruschke, 2014)，其中 Gamma 分配本身也有兩個參數 (形狀參數與尺度參數) 來描述，但在這裡我們將之轉換為平均參數 μ_E 和變異參數 σ_E^2 以方便模型的設立與事後的理解與解釋。至於不同狀態下的回答準則 \mathbf{d}_i 和 \mathbf{g}_i 得符合 $\sum_i^J d_{ij} = \sum_i^J g_{ij} = 1$ ，而 Dirichlet 分配可滿足之。然而，為了滿足「限制 1」與「限制 2」的設定，實際上階層式貝氏模型的設計會與以上描述略有不同，詳細語法設定可參考附件 A 中對於 LTCM 階層式貝式模型的程式碼。

接續，為使用貝氏統計推論方法，我們為 \mathbf{T} 、 \mathbf{E} 、 \mathbf{B} 、 \mathbf{G} 、& \mathbf{D} 之先驗分配中的參數設置超先驗分配 (hyperprior) 如下：

$$\mu_T \sim Beta(\alpha, \alpha), \quad \alpha = 2$$

$$\tau_T \sim Gamma(\mu_{\tau_T}^2 / \sigma_{\tau_T}^2, \mu_{\tau_T} / \sigma_{\tau_T}^2), \quad \mu_{\tau_T} = 10, \quad \sigma_{\tau_T} = 10$$

$$\mu_E \sim Uniform(c1, c2), \quad c1 = 3, \quad c2 = 14$$

$$\sigma_E \sim Gamma(\mu_{\sigma_E}^2 / \sigma_{\sigma_E}^2, \mu_{\sigma_E} / \sigma_{\sigma_E}^2), \quad \mu_{\sigma_E} = \mu_E / 4, \quad \sigma_{\sigma_E} = \mu_E / 8$$

$$\mu_b = 1/2$$

$$\tau_b \sim Gamma(\mu_{\tau_b}^2 / \sigma_{\tau_b}^2, \mu_{\tau_b} / \sigma_{\tau_b}^2), \quad \mu_{\tau_b} = 10, \quad \sigma_{\tau_b} = 10$$

$$\boldsymbol{\alpha}_d = (0, 1, \dots, 1)^\top$$

$$\boldsymbol{\alpha}_g = (1, \dots, 1, 0)^\top$$

此超先驗分配的設置仿照 Batchelder and Anders (2012) 對於 LTM 的設定，除了將原先透過公理 4 直接做出回答的模型設計轉為兩個階段公理 4' 和 5'，並加入相對應參數的

超先驗分配。我們假設任意的文化真值 T_k 的平均參數 μ_T 之超先驗分配為 Beta 分配，因其可能範圍在 $[0, 1]$ ，但在這裡我們加入些微的先驗資訊 (prior information)，設定 μ_T 可能位於在 $1/2$ 上下附近。而相對應的文化真值 T_k 的精確參數 τ_T 則設定為無訊息 (non-informative) 的 Gamma 分配，讓資料來說話。與之相似的狀態閾值 b_i ，其平均參數 μ_b 直接給定在 $1/2$ ，因為我們認為受試者看完題項後進入「狀態 0」或是「狀態 1」，主要是會受到題項本身的文化真值與此受試者的文化能力影響，而比較不會是受到內在的偏誤而使其容易比較偏向某一種狀態，但實際上受試者的狀態閾值 b_i 還是保有隨機性的，因此對於其中的精確參數 τ_b 不設微定值，而是採用無訊息的 Gamma 分配。

需要注意的是，我們在受試者能力 E_i 的超先驗分布設計上給予先驗較強的資訊假設，主要的原因在於，受試者在進入到「狀態 0」或是「狀態 1」的機率除了收到題項本身的文化真值大小外，也同時受到該受試者本身的文化能力 E_i 與狀態閾值 b_i 同時競爭的影響。由於 E_i 與 b_i 是我們要同時估計的參數，在我們的階層模型中的設計下此二參數有可能有不可辨識性的問題，也就是說假設在進入「狀態 1」同樣的機率下

$$\begin{aligned} \Pr(S_{ik} = 1|Y_{ik}, b_i) &= \Pr(Y_{ik} > b_i) \\ &= \int_0^b \frac{\Gamma(E_i)}{\Gamma(T_k E_i) + \Gamma((1 - T_k) E_i)} y^{T_k E_i - 1} (1 - y)^{(1 - T_k) E_i - 1} dy \end{aligned}$$

受試者相關的參數 E_i 和 b_i 實際上可能有不止一組解，且 E_i 在潛在真值 Y_{ik} 中扮演的是精確參數，因此對於 $\Pr(S_{ik} = 1|Y_{ik}, b_i)$ 變更所需要的幅度會比從用 b_i 來調整還要來的大。此外，在模擬分析的過程發現，即便在 b_i 的超先驗分配上有設定一些先驗資訊，平均值要落於 $1/2$ ，然而在這種設定下還是很容易產生所有受試者的文化能力 E_i 都得到相似的後驗分配，表示每個人的能力在此模型下的估計都差異不大。對此，Batchelder and Anders (2012) 對文化能力的超先驗分布給予較強先驗資訊，文化能力的平均值 μ_E 設定在 Uniform 分配，範圍從 3 到 20 都有相同的機率密度，使得貝式估計法有足夠大範圍值去探索，而在標準差 σ_E 則設定為 Gamma 分配並隨著 μ_E 做變化，讓 E_i 的變化能夠隨著平均值大小的升高而有足夠大的變化來讓每位受試者文化能力估計時有所差異。在這樣的超先驗分配的設定下，Batchelder and Anders 在用模擬資料與實徵資料的分析發現，不管在對 σ_E 各種可能值的敏感度分析 (sensitivity analysis) 和後驗預測檢驗 (posterior predictive check) 上，或是對其他參數的後驗平均 (posterior mean) 作點估計，其結果都差異不大，因此他們認為這種超先驗分配的設定大體上是可以接受的。但是在本研究使用的 LCTM 模擬研究發現，這樣的設定其實沒有明顯地改善文化能力 E_i 間差異不大的問題，詳細的內容請見下一章節的說明。我們的 LCTM 模擬設定與原先 LTM 不同之處還有回答準則的超參數 (hyperparameter) 的設計，我們除了在

限制 2 中將受試者的 d_{i1} 與 g_{iJ} 限制為 0， d_i 與 g_i 在其他選項的機率我們預先也沒有任何資訊，因此都設定相同權重值。

模擬資料分析

本章節主要說明在既有的模擬資料下，使用 LTCM 階層式貝氏模型及其統計推論的表現為何。貝氏推論是利用以觀察到的資料與模型建立的概似函數 (5) 與研究者對參數給定先驗分配，來得到預估計參數的後驗分配進行後續的統計推論。然而在許多實務情境下我們即便有這些資訊，也難以用解析的方式得到後驗分配的確切長相。因此，一種常被使用的數值方法為「馬可夫鏈蒙地卡羅法」(Markov Chain Monte Carlo method, MCMC method)，此演算法可以得到參數後驗分配的抽樣樣本 (Gelman et al., 2013; Hoff, 2009; Kruschke, 2014)。近十年來 MCMC 在諸多領域中被廣泛運用，這也使得貝氏統計得以快速發展。MCMC 在 CCT 中的使用介紹可參考 Karabatsos and Batchelder (2003)。當今有許多程式可以執行 MCMC，而本研究使用 R 語言 (R Core Team, 2020) 透過 rjags (Plummer, 2019) 與 R2jags (Su & Yajima, 2020) 兩個 R 套件來連結 JAGS (Plummer, 2003) 進行吉布斯採樣 (Gibbs sampling)。吉布斯採樣為 MCMC 演算法中常見的採樣算則，乃透過全條件分配 (full conditional distributions) 產生一串馬可夫鏈，是梅特羅波利斯－黑斯廷斯 (Metropolis-Hastings) 算法的特例。另外，R 與 JAGS 皆為免費的開源軟體，前者足以讓研究者在做統計 (包含貝氏統計) 甚至在統計以外的領域上使用。LTCM 執行 MCMC 時所需要的 JAGS 語法可參見附件 A。

下文將包括三部分：(1) **模擬資料生成**：透過設定模型中的參數值，產生一筆模擬資料來接續貝氏統計的分析；(2) **參數估計回復性 (parameter recovery)**：模擬資料中藉 MCMC 所得到的各參數的後驗分配，檢驗是否能還原參數的設定值。(3) **後驗預測檢查 (posterior predict check)**：利用 MCMC 得到的參數後驗分配的樣本，檢驗是否能重新模擬出原本真實資料的樣貌 (例如該筆資料本身的觀察值或是某個統計量)。透過上述模擬資料的分析，我們可以評估研究者所建立的模型是否有能力估計出其所感興趣的參數，以及模型是否有捕捉到受試者在回答題項的認知行為，以確保在實際資料分析時該模型是適用的且可以被估計的 (Wilson & Collins, 2019)。

模擬資料生成

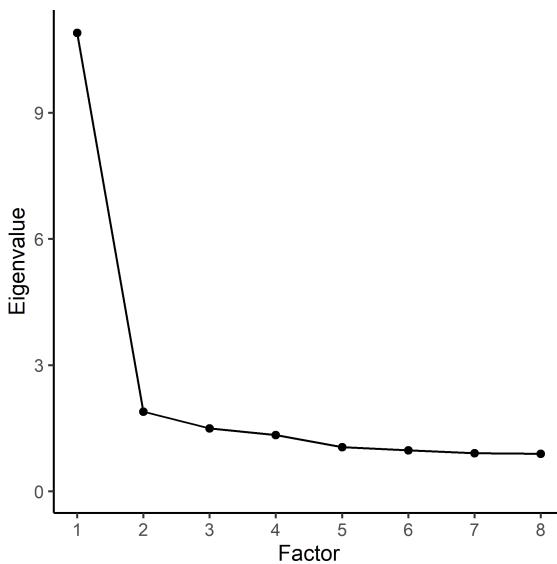
我們模擬受試者人數 $N = 25$ 在題項數目 $M = 50$ 以及信心程度選項有 $J = 5$ 個下的回答資料。其方法是透過預先給定 LTCM 之階層式貝式模型中的超參數的值，依照其先驗分配的形式產生每位受試者具有的特質變項 (包括： E, B, D, G) 和題項的文化共識真值 (T)，同時滿足限制 1 與 2 的條件，令每位受試者 i 兩狀態下的回答準則比值 d_{ij}/g_{ij} 對於 j 乃非遞減的函數且 $d_{i1} = g_{iJ} = 0$ 。有了每位受試者與題項的設定值後，我們依照概似函數模擬所有人的回答矩陣 $(x_{ij})_{N \times M}$ 。

然而在進行 LTCM 的分析之前，得需考量此筆資料是否真的符合公理 1：該文化 (所有 25 位模擬受試者) 下有一組文化共識答案。在此我們計算出受試者間的相關矩陣並進行因素分析，來檢驗是否有明顯的單一因子。計算上我們採用最小殘差因素分析法 (minimal residual factor analysis, MINRES, Comrey, 1962)，在 R 軟體中使用 psych 套件中的 fa() 即可實作 (Revelle, 2020)。圖 1 展示此筆模擬資料各因素所佔的特徵值 (eigenvalue)，當特徵值的數值越高表示該因子能解釋整體的變異量也越多。在此筆模擬資料中我們發現有個單一因子的特徵值特別高，從第二個以後的因子呈現線性遞減。表示此筆資料很有可能受此單一因子就能解釋掉大部分的資料變異，也顯示著每道題項只有一組文化共識真值來影響受試者的回答，而剩下的因子則可解釋掉相關矩陣的殘餘變異。附帶一提，就目前為止並沒有唯一的方法或統計檢定來判定陡坡圖該選取幾個因子。經驗法則可採取手肘法 (elbow method) 或是比較第一和第二因子的特徵值比值是否超過 3，然而有研究發現經驗法則的判定方式不一定保證該筆資料只有一組文化共識真值所組成。不同文化共識下的群體比例及能力程度皆會影響陡坡圖的長相，因而造成誤判 (Batchelder & Anders, 2012; Weller, 2007)。其實，貝氏統計中可以對某個統計量進行後驗預測模型檢查，該方法可以檢驗模型的假設是否正確，詳細流程會在後續章節說明。

本次模擬分析中，執行的 MCMC 設定了 6 條採樣鏈 (chains)，每一條採樣鏈執行 30000 次疊代並捨棄掉前 10000 次疊代採樣值做為預燒期 (burn-in period) 以避免初始值設定的影響。在先前的檢驗中發現，受試者文化能力的自相關 (autocorrelation) 在延遲好幾期後依舊特別高，導致每個樣本真正帶有的資訊量並不高。換算成有效樣本 (effective sample size, ESS) 是非常的低，因此我們把疊代次數設定為 30000 以累積更多訊息，但同時為了省電腦記憶體的資源，因而採取每 10 次疊代才紀錄 1 次值 (thinning)。總結來說，每個參數實際真正機錄下的共有有 12000 ($= 6 \times (30000 - 10000)/10$) 個取樣點。然而執行 MCMC 需要確保這些採樣的馬可夫鏈達到收斂的狀態，才能確信這些得到的樣

圖 1

模擬資料之陡坡圖。



本具有代表性。我們採視覺檢驗法，包括監控歷史痕跡圖 (trace plot) 與密度圖 (density plot)，確認六條疊代鏈都有交疊在一起，以及數值檢驗上採用 Gelman–Rubin potential scale reduction factor (PSRF, \hat{R}) 診斷指標，得到每個參數的該數值皆遠小於經驗值標準 1.1，且很接近 1.0 (完全收斂狀態)。雖然理論上當疊代次數接近無限次時才能保證收斂，但就目前的檢驗結果看來在這樣的 MCMC 疊代設置下，我們有足夠的信心所得到的樣本可使用 (MCMC 診斷與說明可以參考 Gelman et al., 2013; Kruschke, 2014)。接下來兩章節的貝氏統計分析是利用上述 MCMC 設定得到的後驗分配樣本所得的結果。

參數估計回復性

本章節旨在確認在 LTCM 此模型下，運用貝氏統計的推論方法估計出的參數是否能還原真實參數值。貝氏學派在統計推論上與古典頻率學派最大的差異在於不會只將參數視為未知定值，而是認為其具有不確定性，因此在推論上不會只單單使用點估計 (如：後驗平均值 (posterior mean)、後驗中位數 (posterior median) 或後驗眾數 (posterior mode))，還會利用區間估計 (如 95% 最高 (事後) 密度區間 (highest (posterior) density interval, HDI)) 來表示經由觀察的資料下我們對於參數可能出現的 95% 可能性是在多少範圍內。當然，在最好的情況下我們希望直接使用後驗分配來顯示我們對參數的先驗機率以及觀察資料下的的綜合評估。然而在溝通上可能會相較困難些，因此在接下來的參數估計回復性的評估上。本研究會報告後驗期望值與 50%, 75% & 95% HDI 來表示

對於參數可能發生值以及相對應的可能性大小。

首先，在題項的文化真值部分，一種很簡易直覺的估計方式就是以該題項所有人回答的平均 \hat{T}_k 來當作 T_k 的估計，其計算方式為

$$\hat{T}_k = \frac{1}{N} \sum_{i=1}^N X_{ik}^*, \text{ 其中 } X_{ik}^* = \frac{X_{ik} - 1}{J - 1} \in [0, 1]$$

($k = 1, \dots, M$)。圖 2 的左上圖是將每道題項所估計的 \hat{T}_k 與真實設定值畫出散佈圖，而右上圖則是改為 LTCM 對 T_K 的後驗平均值作為估計值，並分別計算出估計值與真實值的皮爾森相關 (Pearson correlation coefficient, r) 和均方根誤差 (root mean square error, RMSE)。我們發現使用模型 LTCM 來對題項的文化真值做點估計比起用一般平均的方式，整體表現要來得好 (r 稍大且 RMSE 較小)，趨勢上也落在斜率為 1 的斜值線上 (真實值與估計值皆相等的線)；反倒是在沒有模型之下，直接使用受試者的回答平均當作估計值，在真實值較小的區域有產生高估的情形。另一方面，若我們看 HDI 的表現，圖 2 下方為毛毛蟲圖 (caterpillar plot)，其中紅色叉叉表示真實值，而藍色空心圓圈表示後驗平均值，藍色線段則分別表示不同機率下的 HDI。從圖中可觀察到，大多數題項的 95% HDI 都有包含到真實參數值 (除了其中 3 題)。整體而言，LTCM 模型對於 T 的參數估計回復性表現是不錯的。

圖 2

題項文化真值 T_k 之後驗平均散佈圖，以及 50%, 75% & 95% HDI 毛毛蟲圖。

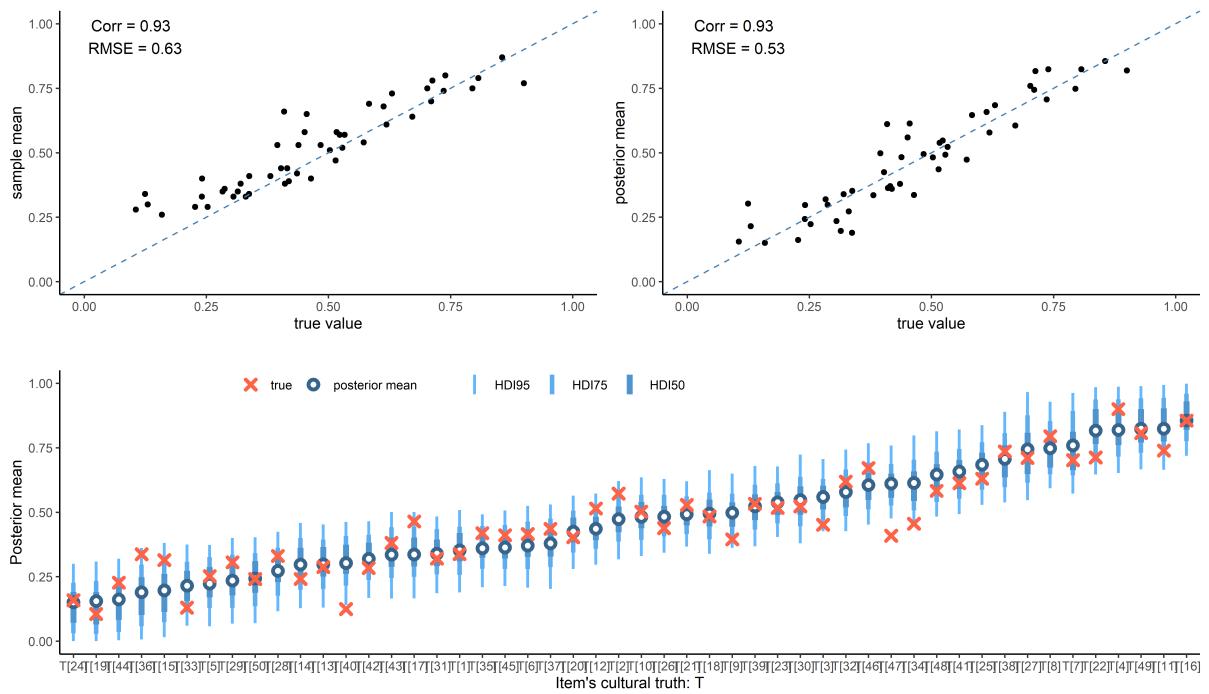


圖 3 的左上與中間分別為受試者的文化能力 E_i 的散佈圖與毛毛蟲圖，而右上與最下方

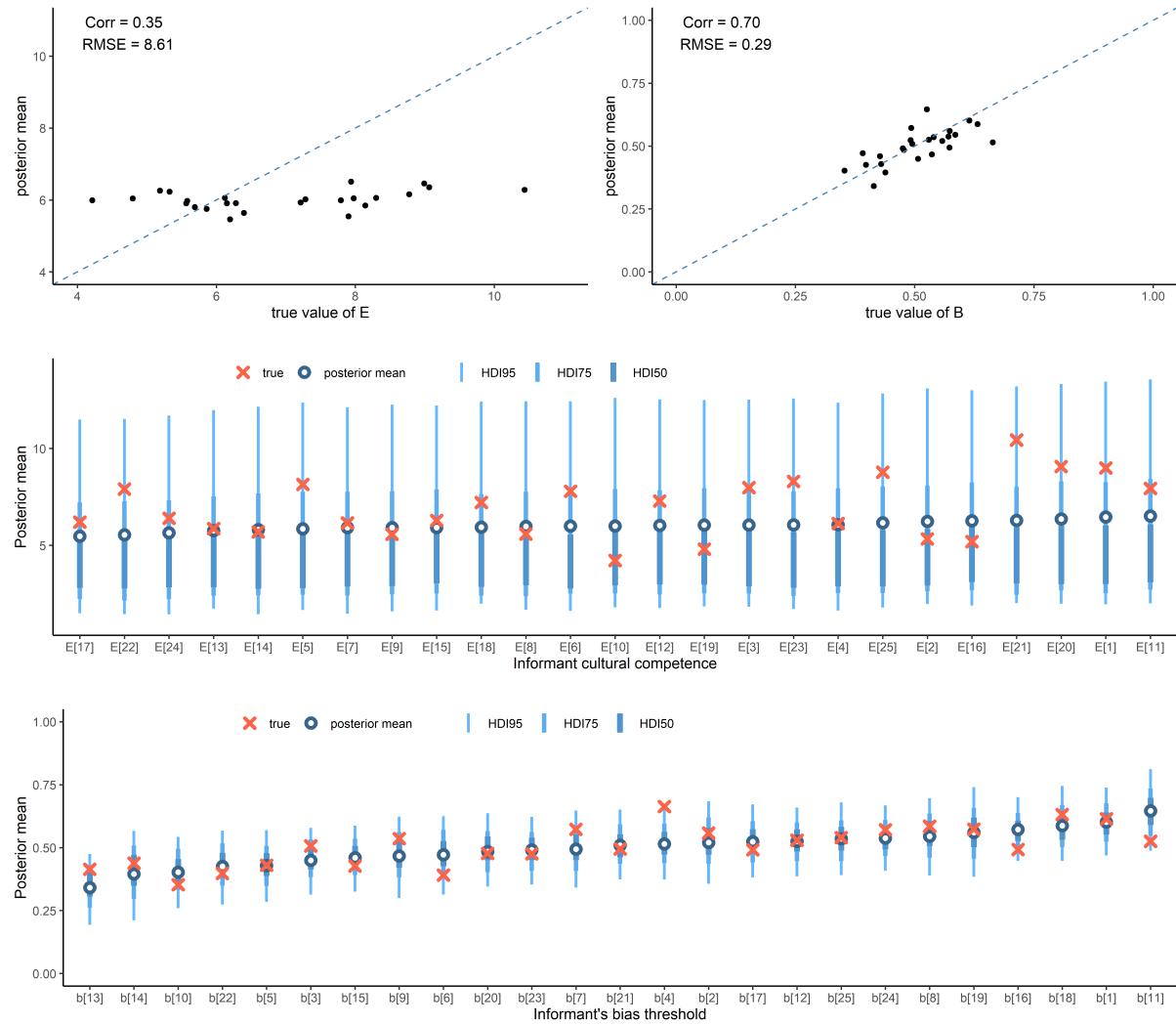
分別為感知閾值 b_i 的散佈圖與毛毛蟲圖。其中偏誤閾在參數估計回復性上表現不錯，後驗平均與真實值差異不大 (RMSE 小)，其 95% HDI 也皆都包含到真實值。然而在文化能力的估計上卻不怎麼理想。我們從散佈圖可以很明顯發現，模擬的 25 位受測者的能力不管是後驗平均或是後驗分配皆沒有很明顯地感覺到差別；雖然 95% HDI 皆有包含在真值內，但實際上我們點估計的表現在真實值高的地方會低估，而在真實值低的地方會高估。其實這種「收縮」(shrinkage) 的現象在階層式模型容易出現，此乃因為每個人的 E_i 皆服從於更高一層的分配 ($E_i \sim \text{Gamma}(\mu_E^2/\sigma_E^2, \mu_E/\sigma_E^2)$)，在 E_i 後驗分配的估算中同時會考慮到 i 本身資料與上一層群體分配的共同作用，因此所得估計會較只用自身資料更傾向於群體集中的位置。

很顯然地，在 LTCM 階層式貝式模型架構下，每個人 E_i 後驗平均值過於集中到群體的平均 μ_E (其後驗平均估計為 6.03)，以至於看不出受試者間的差異，即便我們在題項的文化真值與受試者的感知閾值上估計的不錯。我們推測造成差異不大的可能原因在於：(1) 先前有提到過 b_i 與 E_i 之間可能存在不只一組解，彼此在估計上會相互影響；(2) 同樣身為受試者相關的參數 G_i 與 D_i 可能與 b_i 一同解釋掉 E_i 之間的變異，導致整體估計不太理想；(3) E_i 的先驗分配被給予太強的先驗資訊。雖然 Batchelder and Anders (2012) 在他們 LTM 先驗分配上運用同樣的設計，在他們的圖 5 (毛毛蟲圖) 認為這樣的結果是可以被接受的 (他們並沒有畫出散佈圖)，但實際上受試者的文化能力差異之間的差別也是不大的。也許這一方面值得後續的研究探討，亦即 b_i 與 E_i 在模型中之間的關係，以及該如何設計較好的先驗分配和超先驗分配。附帶一提，若將題項的信心回答選項改為兩個 ($J = 2$)，LTCM 其實是與 LTM 是相同的，且在各個參數估計回復性上與 Batchelder and Anders (2012) 的結果差異不大。

接續，我們檢驗狀態 0 與狀態 1 下的回答準則估計上的表現。圖 4 的第一列顯示 G_{ij} 與 D_{ij} 後驗平均和真實值的散佈圖，整體狀況在斜直線上，雖然看起來沒有存在系統性誤差，但指標卻沒有想像中的理想，RMSE 看起來還是偏大些。由於在回答選擇的機率上 $\sum_{j=1}^J G_{ij} = \sum_{j=1}^J D_{ij} = 1$ ，因此我們在圖中的第二列分別顯示每位受試者在兩個狀態下回答準則的疊狀長條圖，如此較容易觀察估計值與真實值的關係。舉例來說，第一位受試者在狀態 0 下的 G_{1j} 分別在四種 (因為 $G_{i5} = 0$) 回答準則的後驗平均之間的比例，與真實值之間的大小關係相似；然而同樣是在狀態 0 下，最後一位受試者的 G_{25j} ，其 $G_{25,1-3}$ 皆明顯高估許多，而 $G_{25,4}$ 就一定會是低估。若我們觀察區間估計的部分，也就是圖中最下方兩列的 95% HDI 毛毛蟲圖，也可發現有許多參數很明顯地沒有被可信區間包含到，顯示在回答準則的估計上回復性不太好。如前段所述，我們推測原因

圖 3

受試者文化能力 E_i 與感知閾值 b_i 之後驗平均散佈圖，以及 50%, 75% & 95% HDI 毛毛蟲圖。

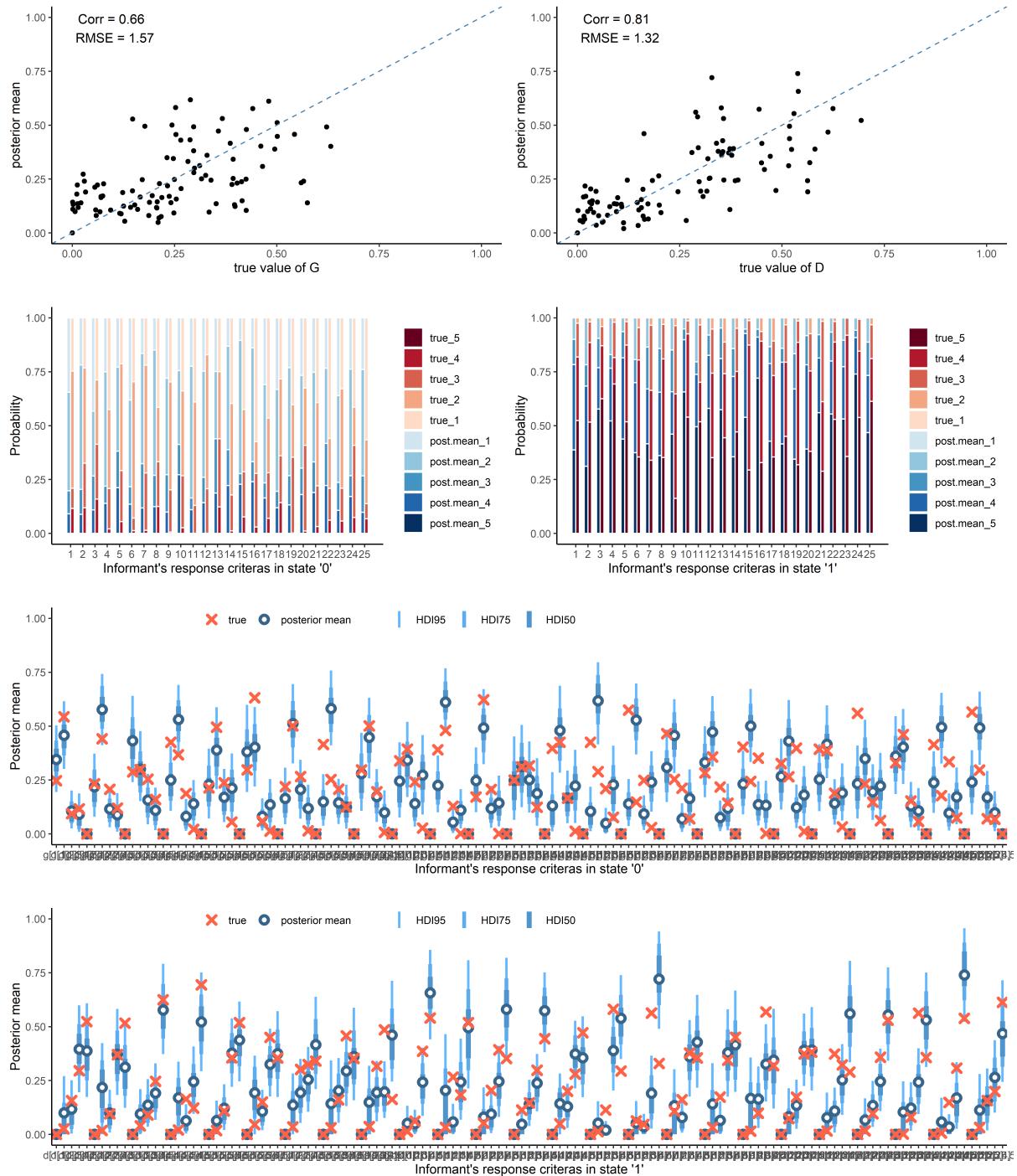


在於受試者參數 E_i 、 b_i 、 G_i 和 D_i 之間可能會互相解釋掉彼此的變異，導致在這幾個參數上 (除了 b_i 外) 估計得不甚理想。也許日後可藉增加題項數目來改進估計表現。

最後，在超參數的部分 (見圖 5)，參數估計的回復性表現看起來不錯，真實值皆含括在 95% HDI 內。但需注意 τ_b 的後驗分配範圍特別大，其主要原因乃 τ_b 在 b 的 Beta 分配中扮演精確參數，因此需要較大幅度的改變才能影響到 b 的準確性。

圖 4

回答準則 G_i 與 D_i 之後驗平均散佈圖、後驗平均疊狀長條圖以及 50%, 75% & 95% HDI 毛毛蟲圖。

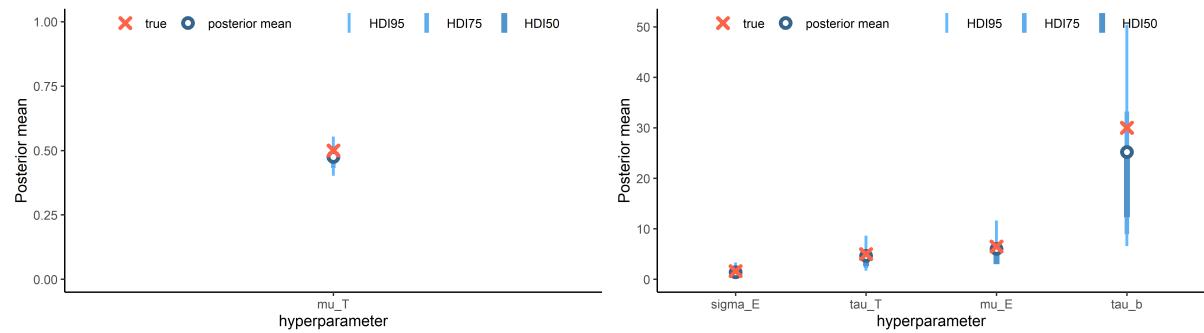


後驗預測檢查

如果模型估計良好，我們從估計出的參數模型產生的複製資料 (replicated data) 應該要與原始資料相似才對。換句話說，原始資料應該會被資料後驗預測分配中給包含住。

圖 5

超參數 $(\mu_T, \tau_T, \mu_E, \sigma_E, \tau_b)$ 之後驗平均散佈圖，以及 50%, 75% & 95% HDI 毛毛蟲圖。



假若後驗分配沒有涵蓋原始資料，很有可能暗示我們模型估計的不適當。我們執行後驗預測檢查，主要有四大步驟：(1) 選擇一個研究者感興趣的統計量 $T(x)$ (為資料 x 的一個函數轉換，也可就純粹資料本身)；(2) 透過先前 MCMC 得到參數的一系列後驗分配樣本，逐一產生新的複製資料；(3) 將這些複製資料逐一計算其統計量 $T(x^{rep})$ ，並畫出該統計量的後驗預測抽樣分配 (sampling distribution)；(4) 觀察原始資料所計算的統計值 $T(x)$ 是否在抽樣分配的可靠區間 (credible interval) 內。前述所提之複製資料所產生的後驗預測分配可表達為

$$p(x^{rep}|x) = \int p(x^{rep}|\theta)p(\theta|x)d\theta \quad (9)$$

其中 θ 為模型的參數 (在 LTCM 中即為 T, E, B, D, G)， $p(\theta|x)$ 為參數的後驗分配 (由 MCMC 得到)， $p(x^{rep}|\theta)$ 為模型的概似函數 (式 (5))。另外，統計量的後驗預測分配計算可直接從複製資料的後驗預測分配進行轉換得知 (Gelman et al., 2013)。

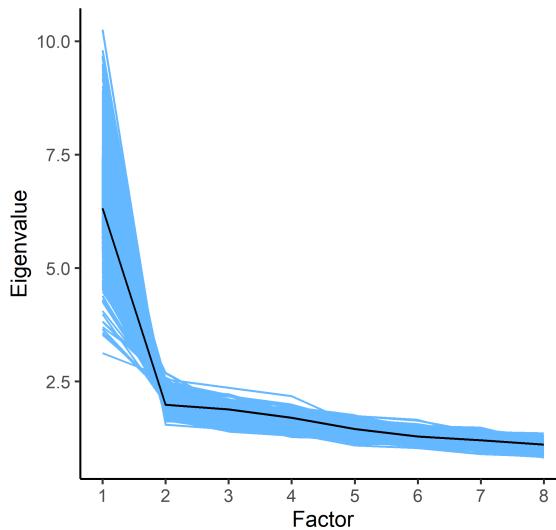
圖 6 為對陡坡圖進行後驗預測檢查的結果。黑線為模擬資料的原始特徵值 (即圖 1)，藍色區域部分為 1000 個從後驗預測分配得來的隨機複製資料。我們據此分別去計算出特徵值，並將之畫成陡坡圖疊加。如圖所示，模擬資料的原始特徵值確實被包含在後驗預測分配之內，說明 LTCM 通過特徵值檢驗，意謂著此筆資料對於所有受試者來說，該 50 道題項應有一組文化共識真值，符合 LTCM 的公理 1。

討論

本研究在文化共識理論 (CCT) 架構下建立一個新的模型，從原先潛在真值模型 (LTM) 只處理題項回答是二分變項 (真/假、是/否或是 0/1)，延伸至可處理排序型類別資料 (包括李克特量表) 的潛在真值信心模型 (LTCM)，題項回答可以從 J 個不同信心程度

圖 6

陡坡圖的後驗預測分析圖。黑線為模擬資料的特徵值，藍色區間為根據 1000 個後驗預測分配得來的隨機複製資料所計算出的特徵值疊加。



選項中選擇。特別是，LTCM 納入閾值理論中兩狀態低閾值理論模型，將原先受試者做選擇時受到的選擇偏誤，區分出與題項內容相關的感知閾值以及與題項無關的選擇準則。前者決定受測者會進入到支持題項真值為 0 或 1 的狀態，而後者描繪受測者在分別狀態下選擇 J 個不同信心程度選項的機率。當然，與其他 CCT 的模型目標相同，LTCM 的目標在於估計出這些題項背後所存在的文化共識真值為何，同時也能估計出受試者在該文化下的能力程度和兩種偏誤閾值。我們依據模型公理 1-3、4' 與 5' 以及限制 1 和 2，建立出相對應的階層式貝氏模型，並以電腦模擬方式評估該模型與統計方法上之可行性。

在貝氏統計推論架構下，我們藉模擬資料探討參數估計的回復性和進行後驗預測檢查。在特徵值陡坡圖中，後驗預測檢查顯示我們的模擬資料應該有一組文化共識真值，符合 LTCM 的公理 1 假設。在參數估計的回復性表現上，題項文化真值 T 與受試者的感知閾值 B 在後驗平均與真實值差異不大，且 95% HDI 也涵蓋住參數真實設定值。值得注意的是，利用 LTCM 模型來估計文化真值，比起直接用使用該題項對所有受試者的回答平均來當作估計值，更能精準估計出這群受試者中的共識答案（見圖 2 中的比較）。然而較差強人意之處在於，在受試者的文化能力 E 估計上，雖然 95% HDI 有包含住真實參數值，然而此區間過於寬廣，每個人的後驗平均差異並不大，難以看出每個人的能力差異。有一種可能原因是 E 本身在模型中是作為受測者在該題項之潛在真值 Y 分配中的精確參數（見式 (2)），此需較大值的變化才能在模型中看到較明顯的改變（我們

觀察到其 HDI 的範圍相較其他類型參數要明顯大上許多)。另一方面，受試者的在不同狀態下的選擇準則 G & D ，在參數的回復性上表現不佳，有些受試者的 95% HDI 並沒有包含住真實參數值。目前我們推測一種可能的原因是超參數的先驗份配設計得並不理想，譬如在 E 超參數的先驗分配上我們給予很強烈的訊息(如同 Batchelder and Anders (2012) 在 LTM 中的設計)，但此先驗參數配置有可能並不適合在 LTCM 中使用。另外一種可能的原因是，受試者內的參數 E_i, b_i, B_i & D_i 之間會相互解釋掉彼此之間的變異，導致這幾個參數表現不佳。我們曾嘗試過將所有人的 B 固定為 0.5 生成模擬資料，但並沒有限制模型中的 B 為定值，之後進行 MCMC 的估計發現 \hat{b} 會圍繞在 0.5 上下，然而 \hat{E} 有系統性的低估，甚至對於 T , G & D 估計比原先的還差，這可能暗示著 E 與 B 之間的關係沒有想像中的簡單。但究竟其中原因為何還待後續研究。³

未來研究方向

首先，我們可能得瞭解受試者參數之間在模型中估計上的關係，亦需釐清先驗分配與模型可辨識性的問題，並在實際收集的資料是否也具有良好的估計性質也有待確認。Batchelder 與其同事在 CCT 的後續發展中，考慮了許多模型上的統計性質、納入題像難度參數、缺失值處理以及在同一群體中可能存在多個文化共識的模型 (Batchelder et al., 2018)。本篇研究尚未處理這些問題，這些是 LTCM 之後擴展的方向之一。此外，Batchelder 團隊的 CCT 後續發展中也有針對排序型類別資料發展模型，稱為「潛在真值評分模型」(Latent Truth Rater Model, LTRM) (Anders & Batchelder, 2013)，然而該模型對多個類別選項的處理方式是加入數個選擇偏誤(如同公理 4)，而非如 LTCM 區分出感知與選擇兩個階段，亦即，LTCM 與 LTRM 在受試者認知處理設計上有明顯不同的考量。未來研究上可比較兩模型表現的優劣。

³ 如果 $J = 2$ ，LTCM 實際上就是 LTM。我們據此進行 MCMC 方法對參數估計，發現模型表現與 Batchelder and Anders (2012) 該篇中 LTM 的表現一樣好。

参考文献

- Anders, R., & Batchelder, W. H. (2013). Cultural consensus theory for the ordinal data case. *Psychometrika*, 80(1), 151–181. <https://doi.org/10.1007/s11336-013-9382-9>
- Batchelder, W. H., & Anders, R. (2012). Cultural consensus theory: Comparing different concepts of cultural truth. *Journal of Mathematical Psychology*, 56(5), 316–332. <https://doi.org/https://doi.org/10.1016/j.jmp.2012.06.002>
- Batchelder, W. H., Anders, R., & Oravecz, Z. (2018). Cultural consensus theory. In E.-J. Wagenmakers (Ed.), *Stevens' handbook of experimental psychology and cognitive neuroscience* (4th ed., pp. 201–264). Wiley. <https://doi.org/10.1002/9781119170174.epcn506>
- Batchelder, W. H., & Romney, A. K. (1986). The statistical analysis of a general Condorcet model for dichotomous choice situations. In B. Grofman & G. Owen (Eds.), *Information pooling and group decision making* (pp. 103–112). JAI Press.
- Batchelder, W. H., & Romney, A. K. (1988). Test theory without an answer key. *Psychometrika*, 53(1), 71–92. <https://doi.org/10.1007/bf02294195>
- Broadbent, D. (1966). Two-state threshold model and rating-scale experiments. *The Journal of the Acoustical Society of America*, 40(1), 244–245. <https://doi.org/https://doi.org/10.1121/1.1910047>
- Comrey, A. L. (1962). The minimum residual method of factor analysis. *Psychological Reports*, 11(1), 15–18. <https://doi.org/https://doi.org/10.2466/pr0.1962.11.1.15>
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis* (3rd ed.). CRC Press.
- Hoff, P. D. (2009). *A first course in Bayesian statistical methods*. Springer. <https://doi.org/https://doi.org/10.1007/978-0-387-92407-6>
- Hsu, Y.-F., & Doble, C. W. (2015). A threshold theory account of psychometric functions with response confidence under the balance condition. *British Journal of Mathematical and Statistical Psychology*, 68(1), 158–177. <https://doi.org/10.1111/bmsp.12040>
- Karabatsos, G., & Batchelder, W. H. (2003). Markov chain estimation for test theory without an answer key. *Psychometrika*, 68(3), 373–389. <https://doi.org/10.1007/bf02294733>
- Krantz, D. H. (1969). Threshold theories of signal detection. *Psychological Review*, 76(3), 308–324. <https://doi.org/10.1037/h0027238>

- Kruschke, J. (2014). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan* (2nd ed.). Academic Press.
- Lee, M. D. (2011). How cognitive modeling can benefit from hierarchical Bayesian models. *Journal of Mathematical Psychology*, 55(1), 1–7.
<https://doi.org/10.1016/j.jmp.2010.08.013>
- Lee, M. D., & Wagenmakers, E.-J. (2014). *Bayesian cognitive modeling: A practical course*. Cambridge University Press.
- Luce, R. D. (1963). A threshold theory for simple detection experiments. *Psychological Review*, 70(1), 61–79. <https://doi.org/10.1037/h0039723>
- Nachmias, J., & Steinman, R. M. (1963). Study of absolute visual detection by the rating-scale method. *Journal of the Optical Society of America*, 53(10), 1206–1213.
<https://doi.org/10.1364/JOSA.53.001206>
- Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. *Proceedings of the 3rd international workshop on distributed statistical computing*, 124, 1–10.
- Plummer, M. (2019). *Rjags: Bayesian graphical models using MCMC* [R package version 4-10]. <https://CRAN.R-project.org/package=rjags>
- R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. <https://www.R-project.org>
- Revelle, W. (2020). *Psych: Procedures for psychological, psychometric, and personality research* [R package version 2.0.9]. Northwestern University. Evanston, Illinois.
<https://CRAN.R-project.org/package=psych>
- Romney, A. K., Batchelder, W. H., & Weller, S. C. (1987). Recent applications of cultural consensus theory. *American Behavioral Scientist*, 31(2), 163–177.
<https://doi.org/https://doi.org/10.1177/000276487031002003>
- Rouder, J. N., & Lu, J. (2005). An introduction to Bayesian hierarchical models with an application in the theory of signal detection. *Psychonomic Bulletin & Review*, 12(4), 573–604. <https://doi.org/10.3758/bf03196750>
- Su, Y.-S., & Yajima, M. (2020). *R2jags: Using R to run ‘JAGS’* [R package version 0.6-1]. <http://CRAN.R-project.org/package=R2jags>

- Watson, C. S., & Bourbon, W. T. (1965). Rating scales and two-state threshold models. *The Journal of the Acoustical Society of America*, 38(4), 667–668.
<https://doi.org/10.1121/1.1909772>
- Weller, S. C. (2007). Cultural consensus theory: Applications and frequently asked questions. *Field Methods*, 19(4), 339–368. <https://doi.org/10.1177/1525822x07303502>
- Wilson, R. C., & Collins, A. G. (2019). Ten simple rules for the computational modeling of behavioral data. *eLife*, 8, e49547. <https://doi.org/10.7554/eLife.49547>

附件

LTRM 之 JAGS 程式碼

```

1 # Latent Truth Confidence Model [LTCM]
2 model{
3   # likelihood model
4   for (k in 1:M) {
5     T[k] ~ dbeta(mu_T*tau_T, (1-mu_T)*tau_T)
6   }
7   for (i in 1:N) {
8     for (k in 1:M) {
9       X[i, k] ~ dcat(p_X[i, k, 1:J])
10      p_X[i, k, 1:J] ← p_S[i, k]*d[i, 1:J] + (1-p_S[i, k])*g[i, 1:J]
11      p_S[i, k] ← 1 - pbeta(b[i], T[k]*E[i], (1-T[k])*E[i])
12    }
13    b[i] ~ dbeta(mu_b*tau_b, (1-mu_b)*tau_b)
14    E[i] ~ dgamma(mu_E^2/sigma_E^2, mu_E/sigma_E^2)
15    # d_i/g_i non decreasing setting
16    g[i, 1:J] ~ ddirch(alpha_g[1:J])
17    for (l in 1:(J-1)) {
18      a[i,l] ~ dgamma(shape_a, rate_a)
19    }
20    a_sort[i, 1:J] ← c(0, pow(sort(a[i, 1:(J-1)]), 2))
21    gJ[i] ~ dunif(0, g[i, (J-1)])
22    d0[i, 1:J] ← a_sort[i, 1:J]*c(g[i, 1:(J-1)], gJ[i])
23    d[i, 1:J] ← d0[i, 1:J]/sum(d0[i, 1:J])
24  }
25
26  # hyperprior
27  mu_T ~ dbeta(alpha_mu_T, alpha_mu_T) T(0.0001, 0.9999)
28  alpha_mu_T ← 2
29  tau_T ~ dgamma(mu_tau_T^2/sigma_tau_T^2, mu_tau_T/sigma_tau_T^2)
30  mu_tau_T ← 10; sigma_tau_T ← 10
31  mu_E ~ dunif(c1, c2)
32  c1 ← 3; c2 ← 20
33  sigma_E ~ dgamma(mu_sigma_E^2/sigma_sigma_E^2, mu_sigma_E/sigma_sigma_E
34  ^2)
34  mu_sigma_E ← mu_E/4; sigma_sigma_E ← mu_E/8
35  mu_E ← 6.5; sigma_E ← mu_E/4

```

```
36 tau_b ~ dgamma(mu_tau_b^2/sigma_tau_b^2, mu_tau_b/sigma_tau_b^2)
37 mu_tau_b ← 10; sigma_tau_b ← 10
38 #alpha_d[1:J] ← c(0, rep(1, J-1))
39 alpha_g[1:J] ← c(rep(1, J-1), 0)
40 shape_a ← 2; rate_a ← 0.1
41 }
```