2012 TOPCO 崇越論文大賞

論文題目：

以貝氏 Ordered Probit Model 分析間斷型行銷問卷資料

Applying Bayesian Ordered Probit Model to Analyzing Discrete Marketing Survey Data

報名編號：D0012

以貝氏 Ordered Probit Model 分析間斷型行銷問卷資料

摘要

研究者在進行問卷資料分析時，常會面對以下三大挑戰：一、資料屬性與分析方法的統計假設不符，可能會造成分析結果偏差的情況。二、問卷資料多為橫斷性資料，任何統計方法都無法針對個別受訪者的異質性做探討。三、研究者多希望了解不同尺度問項彼此之間的關係。因此，本研究提出貝氏 Ordered Probit Model 的解決方案，使用蒙地卡羅馬可夫鏈模擬技術估計模式參數，並實際應用於分析美國啤酒消費問卷中。結果顯示，受訪者對於尺度選擇的偏好具異質性，亦即每位受訪者皆有專屬的一組尺度區間；結果亦顯示在李克特分數越低時，其所對應之尺度的變異程度越小，這個發現驗證了固定區間的不合理性。

關鍵字：問卷調查分析、貝氏模式、Ordered Probit Model

壹、 緒論

一、 研究背景

問卷調查一直是最為廣泛使用且最具彈性的行銷研究方法，根據研究目的不同，問卷設計可以不同尺度呈現，以蒐集最豐富之資訊。而在眾多問卷尺度中，Likert(1932) 提出李克特量表(likert scale)最常用來衡量受訪者對個別問項的態度及傾向。雖然，問卷資料多為間斷型資料或順序型資料，大多數的問卷資料仍以常態分配為基礎的多變量分析工具，進行資料分析與解讀。

針對上述資料屬性與分析方法的統計假設不符合之問題，本研究提出貝氏 Ordered Probit Model 的解決方案，並使用蒙地卡羅馬可夫鏈(Markov Chain Monte Carlo, MCMC)模擬技術估計模式參數，並將本研究提出的模型與估計方法，實際應用於分析美國啤酒消費問卷中。初步分析結果顯示，受訪者對於尺度選擇的偏好具異質性，亦即每位受訪者皆有自己專屬的一組尺度區間；研究結果亦顯示在李克特分數越低時，其所對應之尺度的變異程度越小，這個發現驗證了固定區間的不合理性。本研究的實證結果也發現，具有某些人格特質的啤酒消費者如：樂於與人溝通交流內心話、很有自己想法且不容易受他人影響的消費者較喜愛嘗試不同品牌的啤酒，對於新產品的接受程度較高，且在挑選啤酒時會傾向注意特殊品牌。

二、 研究動機

本研究認為，在分析間斷型或順序型問卷資料時，研究者會面對三大挑戰，第一、以往大多數的問卷資料仍使用常態分配為基礎的多變量分析工具，進行資料分析與解讀，而當資料屬性與分析方法的統計假設不符合時，可能會造成分析結果有偏差的情況。第二、問卷資料為橫斷性資料，亦即一位受訪者只有一組資料，任何統計方法都無法針對個別受訪者的異質性做分析探討。第三、研究者多希望了解某一組問項對另一組問項的解釋能力，但往往這些不同組的問項可能採用不同的尺度，增加分析資料的困難性。針對上述發現之問題，本研究欲運用貝氏 Ordered Probit Model，改善以往統計分析方法，如：迴歸分析(Regression Analysis)、相關性分析(Correlation Analysis)等處理間斷型之行銷問卷資料，與分析方法假設不符的情況。而透過貝氏統計建構模型，亦可針對不同尺度的問項探討彼此之間的解釋能力，期以更準確推論各問項之間的相關性，以取得有用的資訊進而做出正確的行銷決策。

三、 研究目的

在面對資料分析的問題時，研究者常常忽略，資料屬性和統計假設間的適切性，而貿然使用熟悉的統計工具進行資料分析。例如，傳統上，處理李克特尺度類型資料之前，會先將資料標準化，以解決資料變異的問題。但由於將離散型資料當作連續型資料而計算其平均數及標準差，反而會增加資料的干擾變數，無法準確地根據資料做預測(Rossi et al., 2001)。因此，本研究提出以 Ordered Probit Model，將間斷型資料轉換成連續型資料，並以貝氏統計方法，克服受訪者尺度異質性及小樣本的問題，進行模型參數的統計推論，希望能在行銷研究的實務應用上有所貢獻。本研究的學術貢獻包含以下四點：一、解決在分析間斷型行銷問卷資料時，統計分析方法之基本假設與資料特性不符的問題。二、解釋受訪者在李克特尺度選擇的異質性。三、將不同尺度類型的問卷資料透過模型整合，以了解變數之間的關係。四、驗證所提出之模型較傳統統計分析方法具備較精確的統計推論結果。

貳、 文獻探討

一、 貝氏統計

在古典統計方法中，所謂參數估計的抽樣分配或是參數檢定的統計量，基本上都是假設樣本資料是從符合某機率分配的母體中抽出的，當樣本趨近於無限大時，可透過 asymptotic theory，以最大概似法(maximum likelihood method)估計模型參數。然而大部分的行銷資料多為間斷型資料，且不符合大樣本的原則，因此建

構於 asymptotic theory 的統計推論方法，無法保證其參數估計值仍能保有 consistency 和 efficiency 的特性。而貝氏統計以 likelihood principle 與貝氏定理為其理論基礎，結合概似函數(likelihood function)和未知參數的先驗機率分配(prior distribution)，推導模型參數之後驗機率分配(posterior distribution)，以進行統計推論。因為貝氏統計不受限於大樣本，不以 asympotitic theory 進行統計推論，且具有非常好的有限樣本特質(finite sample property)，使得貝氏統計在量化行銷領域逐日受到重視。

近年來由於貝氏統計在參數估計演算法上的突破，使得貝氏統計在各個領域的應用急速增加，尤其在行銷研究方面。貝氏統計已經處理了如顧客異質性分析、產品定價、品牌選擇、和新產品設計等許多行銷問題。例如，Allenby and Rossi(1999)以層級貝氏模型分析顧客消費行為的異質性。Rossi and Allenby(2003)以層級貝氏模型，有效的進行市場區隔，協助企業擬定行銷策略。關於貝氏統計的理論基礎可參閱 Gelman, Carlin, Stern, and Roubin (2003) 和 Robert and Casella (2010)。關於貝氏統計在行銷領域的應用可參閱 Rossi, Allenby, and McCulloch(2005)。

二、 蒙地卡羅馬可夫鏈演算法

以貝氏統計估計參數最困難之處在於，需要處理複雜的積分問題。假設資料為 D，模型參數為 θ，概似函數為 p(D|θ)，參數的先驗機率分配為 π(θ)，資料 D 的邊際分配為 p(D)，則根據貝氏定理，θ 的後驗機率分配為：

$$p(\theta|D) = \frac{p(D,\theta)}{p(D)} = \frac{p(D|\theta)\pi(\theta)}{\int p(D|\theta)\pi(\theta)d\theta} \tag{2.1}$$

因為資料 D 的邊際分配為 p(D)不包含 θ，因此 θ 的後驗機率分配可以改寫為：

$$p(\theta|D) \propto p(D|\theta)\pi(\theta) \tag{2.2}$$

貝氏統計和古典統計一樣，常需要以敘述統計(summary statistics)以呈現參數的特性。例如，令$h(\theta) = (\theta - \bar{\theta})^2$，則 θ 的後驗變異數為：

$$E_{\theta|D}[h(\theta)] = \int h(\theta)p(\theta|D)d\theta = \frac{\int h(\theta)p(D|\theta)\pi(\theta)d\theta}{\int p(D|\theta)\pi(\theta)d\theta} \tag{2.3}$$

或是要以貝氏統計進行預測時，假設預測之觀測值為 D_f 則 D_f 的預測分配為：

$$p(D_f|D) = \int p(D_f, \theta|D)d\theta = \int p(D_f|\theta, D)p(\theta|D)d\theta \qquad (2.4)$$

無論是方程式(2.3)或(2.4)的計算，都是件極具挑戰性的工作，尤其當 θ 為高維度或是 $p(\theta|D)$非已知統計分配時，$E_{\theta|D}[h(\theta)]$和 $P(D_f|D)$都無法直接計算而得。

在處理後驗機率的多重積分的問題上，貝氏統計的估計是透過反覆抽樣的方式進行，其中又以蒙地卡羅-馬可夫鏈(Monte Carlo Markov Chain, MCMC)抽樣方法，最廣為使用。所謂的 MCMC 是一種從目標分配(Target Distribution)模擬出樣本積分的技術。當 MCMC 的模擬次數趨近於無限時，可以使其所模擬出的樣本分配代表欲積分函數的分配。也就是說，透過 MCMC 模擬方法所模擬出的樣本而建構的後驗機率分配，幾乎等同於真實的後驗機率分配。MCMC 的理論推導可參考 Nummelin(1984)、Tierney(1994)、Gelman, Carlin, Stern, and Roubin (2003) 和 Robert and Casella (2010)。

在眾多 MCMC 模擬方法中，吉氏抽樣法(Gibbs sampling)和 Metropolis -Hasting 演算法(Metropolis-Hastings Algorithm, M-H)最為常用。所謂 M-H 演算法最先是 Metropolis et al.(1953)根據馬可夫鏈所發展出來的理論，其後經由 Hasting(1970)建立整個演算法的架構，直到 Muller(1993)和 Tierney(1994)進一步闡述其演算法在後驗分配推估的價值，才逐漸提高 M-H 在統計學領域的應用。根據轉換函數設立的不同，M-H 演算法又可分為獨立性 Metropolis-Hastings 演算法(Independence Metropolis-Hastings Algorithm)和隨機漫步 Metropolis-Hastings 演算法(Random-walk Metropolis-Hastings Algorithm, RW M-H)。Gibbs 抽樣法是 M-H 演算法的一種特例，適用於參數的完全條件分配(Full Conditional Distribution)為已知分配的情況下進行疊代抽樣。以下分別針對本研究主要採用其中兩個 Gibbs 及 RW M-H 抽樣方法做詳細說明。

(一) Gibbs 抽樣法

Gibbs 抽樣法是能直接從個別參數的完全條件分配產生隨機變數，在 Markov Chain 收斂之後，所抽出的隨機變數即為所有參數的聯合機率分配的一項技術。其方法的理論基礎可追朔至 Metropolis et. al.(1953)相關文獻，進而由 Hastings(1970)發展較完整的架構，直到近代 Gelfand and Smith(1992)闡述 Gibbs 抽樣法為過去傳統的統計方法帶來新的影響，使得近代統計學家能解決處理一些複雜的統計計算。

使用 Gibbs 抽樣法的首要條件是，欲估計的參數的條件後驗分配，又稱為完全條件分配，必須為一般熟知的機率分配(亦即，完全條件分配有 close form)。假設

資料為 D，且模型有 P 個參數$\theta' = (\theta_1, \theta_2, \ldots, \theta_P)$。Gibbs 抽樣法就是分別建構 P 個參數的條件後驗分配，如下所示，然後以疊代的方式，進行抽樣。

$$\pi_1(\theta_1|\theta_2, \theta_3, \ldots, \theta_P, D)$$

$$\pi_2(\theta_2|\theta_1, \theta_3, \ldots, \theta_P, D)$$

$$\vdots$$

$$\pi_i(\theta_i|\theta_1, \theta_2, \ldots, \theta_{i-1}, \theta_{i+1}, \ldots, \theta_P, D)$$

$$\vdots$$

$$\pi_P(\theta_p|\theta_1, \theta_2, \ldots, \theta_{p-1}, D)$$

則 Gibbs 抽樣法主要程序如下：

1. 給定參數起始值$\theta^0 = (\theta_1^0, \theta_2^0, \ldots, \theta_P^0)$，並決定抽樣次數 R 次。
2. 在第 i 次抽樣，從條件後驗分配$\pi_1\left(\theta_1 \middle| \theta_2^{(i-1)}, \theta_3^{(i-1)}, \ldots, \theta_P^{(i-1)}, D\right)$中抽出$\theta_1^{(i)}$的值，用$\theta_1^{(i)}$取代$\theta_1^{(i-1)}$後，於$\pi_2\left(\theta_2 \middle| \theta_1^{(i)}, \theta_3^{(i-1)}, \ldots, \theta_P^{(i-1)}, D\right)$抽出$\theta_2^{(i)}$的值，再用$\theta_2^{(i)}$取代$\theta_2^{(i-1)}$後，於$\pi_3\left(\theta_3 \middle| \theta_1^{(i)}, \theta_2^{(i)}, \theta_4^{(i-1)}, \ldots, \theta_P^{(i-1)}, D\right)$抽出$\theta_3^{(i)}$的值，以此類推，反覆疊代抽樣，直到抽出$\theta_p^{(i)}$的值。
3. 利用抽樣出的新參數值重複第 2 步驟直到完成第 R 次抽樣。

在完成 R 次抽樣之後，觀測每一個變數的時序圖以判斷 Markov Chain 是否收斂。假設所有參數的 Markov Chain 皆在第 K 次抽樣後收斂，則第 K 次抽樣後所得之參數樣本 $\{\theta_1^{(i)}, \theta_2^{(i)}, \theta_3^{(i)}, \cdots, \theta_p^{(i)}; \forall i = k, k+1, .., p\}$可視為由參數的聯合機率分配$\pi\left(\theta_1, \theta_2, \theta_3, \ldots, \theta_p \middle| D\right)$所抽出之樣本。

(二) Random-walk Metropolis-Hastings 演算法

雖然 Gibbs 抽樣法為相當有效的樣本模擬技術，特別是用在一般熟知的機率分配上，但在實務上仍存在著無法解決的問題，原因是若條件後驗分配並非為一個 close form 的分佈，Gibbs 抽樣則無法針對條件後驗分配進行樣本模擬。換句話說，當欲估計參數其後驗分配不符合一般熟知的機率分配時，我們就需要使用 Metroplis-Hasting 演算法來產生模擬樣本。Random-walk Metropolis-Hastings 演算法的步驟如下：

1. 給定參數起始值為$\theta^{(0)}$。

2. 在第 i 抽樣，由轉換函數 $g(\theta)$抽出新的參數$\theta^{(n)}$，且 $g(\theta)=\theta^{(i-1)}+s*\varepsilon$，$\varepsilon$符合常態分配，平均數為 0，變異數為對稱矩陣或單位矩陣。

3. 在連續均勻分配 U(0,1)中抽出 u 值，並與$\alpha=\min\left\{1,\frac{\pi(\theta^{(n)})}{\pi(\theta^{(i-1)})}\right\}$比較，若 $u>\alpha$則$\theta^{(i)}$的值以$\theta^{(n)}$取代，否則$\theta^{(i)}$以$\theta^{(i-1)}$取代。

4. 重複上述步驟直到完成 R 次抽樣, 直到所抽出之θ值收斂為止。

在步驟 2 中的 s 為一常數，用以控制每次疊代的跳動幅度，及新參數接受率。若 s 設定過大，則參數的接受率會較小，也就是透過轉換函數抽樣的θ_n^*被拒絕的機率較大，可能會導致參數疊代收斂至真實值的速度變慢，或是新參數值一直不被接受，造成馬可夫鍊無法移動，參數無法收斂。然而，若 s 設定過小，參數的接受率會較大，但導致參數收斂速度太快，容易陷入區域最佳值。因此根據 Muller(1993)和 Tierney(1994)的建議，跳動幅度要控制在一定的範圍內，一般所建議的參數接受率為 25%~40%之間。

三、 資料擴增演算法和潛在變數模型

行銷人員常常希望能透過了解顧客需求與偏好，以制定更有效的行銷策略。然而，顧客行為具異質性，且受到一些內在無法觀察到的因素影響。即使透過行銷資料庫，研究者最多僅能觀察到顧客的決策結果，無法了解決策的過程與動機及外在環境對顧客決策的影響。行銷研究者或許可以透過問卷設計，了解受訪者對產品或服務的認知、態度與想法。但是，問卷調查受限於其 cross-sectional 的特性，且結果仰賴受訪者誠實回答的意願，某些影響決策過程的微妙心思意念，是無法透過問卷調查取得，造成受訪者認定的"自己"未必會反映在問卷調查中。潛在變數模型(Latent Variable Model)便是希望將無法直接觀察之變數，透過經濟學或是心理學的理論基礎，建構於統計模型中，以協助行銷研究者，了解消費行為可能的驅動原因。

以常用來分析二元選擇問題的 Binary Probit Model 為例，假設 t 為時間，y_t為購買決策。當$y_t=1$時，該顧客決定購買；而當$y_t=0$時，該顧客決定不購買。假設該顧客的決策與若干獨立變數x_t(如價格、產品屬性)及其對應的個人偏好β有關，而ε_t為殘差項，代表所有無法為獨立變數所解釋的部分。若在時間點 t，該顧客取得產品的效用$x_t\beta+\varepsilon_t$大於0，顧客會購買(即$y_t=1$)；若該顧客取得產品的效用$x_t\beta+\varepsilon_t$小於0，顧客則不會購買(即$y_t=0$)。因此，Binary Probt Model 可以下列數學式表示:

$$y_t=\begin{cases}1 & \text{if}\quad x_t\beta+\varepsilon_t\geq 0\\ 0 & \text{if}\quad x_t\beta+\varepsilon_t<0\end{cases}\qquad \varepsilon_t\sim N(0,1) \tag{2.5}$$

若進一步假設殘差項ε_t為標準常態分配，β的後驗分配為$\pi(\beta)$。則根據貝氏定理，β的後驗分配為：

$$p(\beta|\{y_t\},\{x_t\}) = \left(\iint \cdots \int \varepsilon_1\varepsilon_2 \cdots \varepsilon_T d\varepsilon_1 d\varepsilon_2 \cdots d\varepsilon_T\right) \cdot \pi(\beta) \tag{2.6}$$

由β的後驗分配可看出，若要取得β的後驗分配估計值，必須要先對參數的後驗分配直接做多重積分的計算。

為了避免直接處理多重積分的計算，潛在變數模型通常使用貝氏統計的資料擴增(Data Augmentation)演算法，進行模型估計。資料擴增(Data Augmentation)演算法最先由 Tanner & Wong(1987)提出，資料擴增演算法藉由潛在變數擴增觀察值資料的機制，使得後驗分配機率計算得以簡化。因此，若以 data augmentation 演算法估計 Binary Probit Model, 則第(2.5)式的模型，可改寫成:

$$y_t = \begin{cases} 1 & \text{if} \quad z_t \geq 0 \\ 0 & \text{if} \quad z_t < 0 \end{cases} \tag{2.7}$$

$$z_t = x_t\beta + \varepsilon_t\text{，}\varepsilon_t \sim N(0,1) \tag{2.8}$$

其中z_t為 Data Augmentation 演算法所引進的潛在變數。雖然在(2.7) (2.8) 式的 binary probit model 中，參數β的後驗分配為$p(\beta|y_t, z_t)$，若要得到$p(\beta|y_t)$的後驗分配，必須對潛在變數z_t積分。然而透過 data augmentation 演算法及 MCMC，所有的積分及估計過程得以簡化，原本參數估計困難的問題可獲得解決。換句話說，在推導出對條件後驗分配$p(z_t|y_t, \beta)$和$p(\beta|y_t, z_t)$後，分別對潛在變數z_t和未知參數β進行抽樣。並在完成所有既定的抽樣次數後，觀測β的時序圖以判斷 Markov Chain 是否收斂。假設所有參數的 Markov Chain 皆在第 K 次抽樣後收斂，則第 K 次抽樣後所得之參數樣本$\{\beta^{(i)}; \forall i = k, k+1, \cdots, p\}$可視為由參數的聯合積率分配(2.7)式所抽出之樣本。

潛在變數模型可視為一種將「可直接觀測的離散性變數(y_t)」轉換成「無法觀測的連續變數(z_t)」的媒介，因此研究者可將各樣理論(如，個體經濟學中效用極大化的理論)作為建構潛在變數模型之基礎，以描述並解釋觀測資料的產生機制。也因為間斷型變數並無機率密度的概念，所以必須透過潛在變數模型的轉換，才能建構資料的概似函數，相關概念在 Rossi et al.(2001) 和 Marshall and Bradlow(2002) 中皆有提及。最後，因為 Data Augmentation 與 MCMC 演算法讓潛在變數模型更容易估計，使得潛在變數模型在經濟學與行銷學上的應用更加蓬勃發展。

參、 研究方法

一、 模型建構

假設問卷受測總人數 H 人，回答李克特尺度問題數共 T 題，使用李克特 C 點評分等級。令y_{ht}為第 h 位受訪者在第 t 題李克特問項的間斷型觀察值。且數值 k 為第 h 位受訪者在第 t 題的評等分數，因此 k 值的範圍為 1 至 C 的任一整數值；$z_{h,t}$為$y_{h,t}$所對應之連續型潛在變數，X 為用來解釋問項評等分數之獨立變數，可為連續型或離散型觀察值。本研究所提出之 Ordered Probit Model 表示如下：

$$y_{h,t} = k \quad \text{if} \quad \gamma_{y_{h,t}} < z_{h,t} < \gamma_{y_{h,t+1}} \tag{3.1}$$

$$z_{ht} = X_h\beta_t + \varepsilon_{ht} \qquad \varepsilon_{ht} \sim N(0, I) \tag{3.2}$$

$$\forall\, h = 1,2,\cdots,H\,;\ \forall\, t = 1,2,\cdots,T$$

亦即，當第 h 位受訪者在第 t 題所勾選的分數為 1 時，可知$y_{h,t} = 1$，其所對應之潛在變數$z_{h,t}$會介於γ_1與γ_2之間。以此類推$y_{h,t} = 2$至$y_{h,t} = C$ 所對應之潛在變數的範圍。

(3.2)式中的觀察值 X 為問卷中其他可用以解釋李克特問項$y_{h,t}$之問項，假設這些解釋問項共有 m 題，尺度不拘。因此，

第一位受訪者的資料結構：

$$y_1 = \left(\mathbf{y_{1,1}}, y_{1,2}, \cdots, y_{1,T}\right)'$$

$$z_1 = \left(\mathbf{z_{1,1}}, z_{1,2}, \cdots, z_{1,T}\right)'$$

$$\Gamma_1 = \left(\gamma_{1,1}, \gamma_{1,2}, \cdots, \gamma_{1,C}, \gamma_{1,(C+1)}\right)'$$

$$x_1 = \left[x_{1,1}, x_{1,2}, \cdots, x_{1,m}\right]'$$

第二位受訪者的資料結構：

$$y_2 = \left(\mathbf{y_{2,1}}, y_{2,2}, \cdots, y_{2,T}\right)'$$

$$z_2 = \left(\mathbf{z_{2,1}}, z_{2,2}, \cdots, z_{2,T}\right)'$$

$$\Gamma_2 = \left(\gamma_{2,1}, \gamma_{2,2}, \cdots, \gamma_{2,C}, \gamma_{2,(C+1)}\right)'$$

$$x_2 = \left[x_{2,1}, x_{2,2}, \cdots, x_{2,m}\right]'$$

$$\vdots$$

以此類推至所有受訪者的資料結構。令$Y_t = \left(y_{1,t}, y_{2,t}, \cdots, y_{H,t}\right)'$為一個 Hx1 的向量，代表所有受訪者回答第 t 題李克特問項的觀察值之集合，$Z_t = \left(z_{1,t}, z_{2,t}, \cdots, z_{H,t}\right)'$為一個 Hx1 的向量，是與$Y_t$對應之連續型潛在變數之集合，

則所有受訪者在第 t 題李克特問項所對應之潛在變數Z_t的迴歸式結構如下：

$$Z_t = X \cdot \beta_t + \varepsilon_t$$

其中

$$X = \begin{bmatrix} 1 & x_{1,1} & x_{1,2} & \cdots & x_{1,m} \\ 1 & x_{2,1} & x_{2,2} & \cdots & x_{2,m} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{H,1} & x_{H,2} & \cdots & x_{H,m} \end{bmatrix}_{H\times(m+1)}$$

$$\beta_t = \begin{bmatrix} \beta_{0,t} \\ \beta_{1,t} \\ \vdots \\ \beta_{m,t} \end{bmatrix}_{(m+1)*1}$$

$$\varepsilon_t = \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \\ \vdots \\ \varepsilon_{H,t} \end{bmatrix}_{H*1}$$

本研究所提出的模式與以往研究不同之處在於：(1)我們將 Likert 資料透過潛在變數模型轉換成與其對應之連續變數，使資料符合分析方法之分配假設，提升推論結果的正確性與解釋力；(2)在本研究中我們試圖結合兩組不同尺度類型的資料，觀察某一組資料對另一組資料的解釋能力；(3)雖然受限於問卷資料為橫斷型資料的緣故，本研究無法針對每位受訪者的異質性探討分析，但是藉由本模式的建構，我們依然可以討論受訪者在尺度使用上之異質性。

二、 參數估計

在瞭解上述資料結構與本研究所提出之模型後，我們將進一步針對所建構模式中的參數後驗分配推導進行說明。首先資料的概似函數描述如下：

$$\left[\{Z_t\},\{\beta_t\},\Gamma_h,\bar{\beta},V_\beta \middle| \{Y_t\},\{X_h\}\right] \propto \prod_{t=1}^{T}\prod_{h=1}^{H}[y_{ht}|z_{ht}]\,[z_{ht}|X_h,\beta_t,\Gamma_h]$$

接著將參數 $\beta_t, d_h, \Gamma_h, \bar{\beta}, V_\beta$ 的先驗機率陳列如下：

$$\beta_t \sim MN_m(\bar{\beta}, V_\beta) \qquad \forall t = 1,2,\cdots,T \tag{3.3}$$

$$d_h^{\,*} \sim MN_{C-2}(\bar{d}^*, A_d^{\,-1}) \tag{3.4}$$

$$d_h = \exp(d_h^{\,*}) \tag{3.5}$$

$$\Gamma_h = (\gamma_1,\gamma_2,\cdots,\gamma_{C+1})' \qquad \forall h = 1,2,\cdots,H \tag{3.6}$$

其中， $\gamma_1 = -\infty,$

$\gamma_2 = 0$

$$\gamma_3 = \gamma_2 + d_1$$
$$\vdots$$
$$\gamma_C = \gamma_{C-1} + d_{C-2}$$
$$\gamma_{C+1} = \infty$$
$$\bar{\beta} \sim MN_m(\mu_0, V_0) \tag{3.7}$$
$$V_\beta \sim IW_m(u_0, S_0) \tag{3.8}$$

值得注意的是，因為本研究所提出之模型中，只有X和Y為觀察值，因此並非所有模型參數都可以同時被估計，此為計量模型常見的"Identification Problem"，表示對潛在變數同時加減乘除某一常數後，模型之概數函數值不變。常見的 Identification Problem 有 Location Shift 和 Scale Shift。關於 Identification Problem 的詳細討論可參 McCulloch and Rossi(2000)。因此，本研究將$\gamma_{h,2}$固定於 0 以解決 Location Shift 造成無法同時 identification 的問題。另外，為了解決 Scale Shift 的問題，本研究固定共變異數矩陣V_β為單位矩陣。

在本研究提出之模式中，需要估計的參數有Γ_h、Z_t、β_t、$\bar{\beta}$、V_β共 5 個，其中Γ_h的後驗分配之推導結果並非一般熟知的機率分配，所以本研究利用 Random-walk Metropolis-Hastings 演算法對Γ_h的後驗分配進行抽樣；另外，Z_t, β_t, $\bar{\beta}$, V_β後驗分配推導的結果是我們所熟知的分配，因此則採用 Gibbs 抽樣法對其後驗分配進行抽樣。茲將整理 MCMC 估計程序如圖 1 所示：

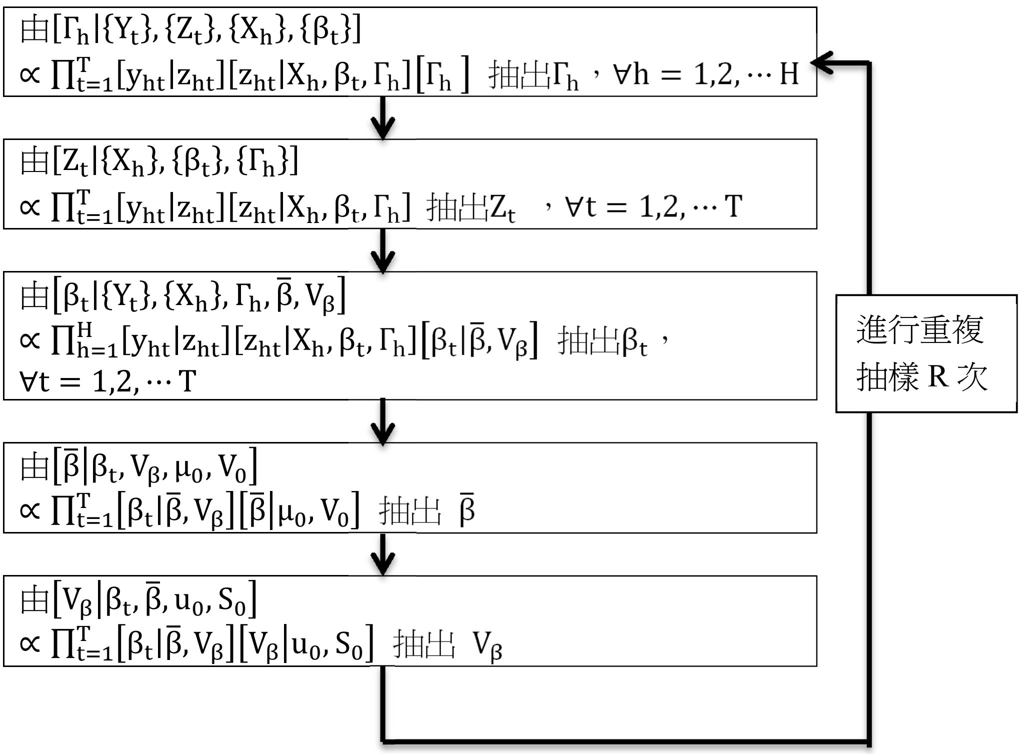

圖 1 MCMC 估計程序圖

肆、　實證分析

一、　研究樣本與資料來源

為驗證本研究中所提出之模式，我們使用美國某啤酒公司所提供的啤酒消費意向調查資料進行實證研究分析。這項研究是針對美國家庭 21 歲以上最常使用或購買啤酒的受測者進行調查，該資料原始共回收 13147 份問卷。因本論文主要為分析順序尺度的離散型資料，因此將所有問卷進行過濾並篩選出有使用啤酒且會購買啤酒的受訪者，探討受訪者的生活態度是否影響啤酒品牌之偏好，篩選結果得到 8930 份調查問卷。生活態度問項為二元變數尺度，而啤酒品牌偏好問項則使用李克特 5 點尺度進行評量，其中 1 分代表"非常同意"；5 分代表"非常不同意"。我們由 8930 份問卷中隨機抽出 500 份問卷資料進行實證分析。

二、　貝氏模式分析

在本研究，我們使用圖 1 之 MCMC 估計程序估計模型參數，在進行模型估計之前，我們需要對各先驗分配內的參數進行初始值的設定，本研究設定各參數之先驗機率如表 1：

表 1 各參數先驗機率設定值表

參數	$d_h^* \sim MN_3(\bar{d}^*, {A_d}^{-1})$	
	$\bar{d}^*$	${A_d}^{-1}$
初始值	$\begin{bmatrix}0\\0\\0\end{bmatrix}_{3\times1}$	$\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix}_{3\times3}$
參數	$\beta_t \sim MN_5(\bar{\beta}, V_\beta)$	
	$\bar{\beta}$	V_β
初始值	$\begin{bmatrix}0\\0\\0\\0\\0\end{bmatrix}_{5\times1}$	$\begin{bmatrix}100&0&0&0&0\\0&100&0&0&0\\0&0&100&0&0\\0&0&0&100&0\\0&0&0&0&100\end{bmatrix}_{5\times5}$
參數	$\bar{\beta} \sim MN_5(\mu_0, {V_0}^{-1})$	
	μ_0	${V_0}^{-1}$
初始值	$\begin{bmatrix}0\\0\\0\\0\\0\end{bmatrix}_{5\times1}$	$\begin{bmatrix}0.01&0&0&0&0\\0&0.01&0&0&0\\0&0&0.01&0&0\\0&0&0&0.01&0\\0&0&0&0&0.01\end{bmatrix}_{5\times5}$
參數	$V_\beta \sim IW_5(u_0, S_0)$	
	u_0	S_0
初始值	10	$\begin{bmatrix}10&0&0&0&0\\0&10&0&0&0\\0&0&10&0&0\\0&0&0&10&0\\0&0&0&0&10\end{bmatrix}_{5\times5}$

在設定模擬次數為 8000 次，每 10 筆估計值儲存為一組的情況下，所有參數皆已經達收斂狀態(如圖 2 所示)，因此我們採用後 4000 次參數模擬的平均值結果作為參數的估計值。在Γ_h參數方面，每位受訪者皆有自己 3 個估計值($\gamma_3, \gamma_4, \gamma_5$)的收斂圖(如圖 3 所示，以第 20 位受訪者為例)；在β_t參數方面，每題李克特問項分別有 5 個估計值($\beta_{.0}, \beta_{.1}, \beta_{.2}, \beta_{.3}, \beta_{.4}$)的收斂圖(如圖 4 所示，以問卷中第 6 個問項為例)；另外$\bar{\beta}, V_\beta$兩個參數為所有李克特問項的共有參數(如圖 5 及圖 6 所示)。

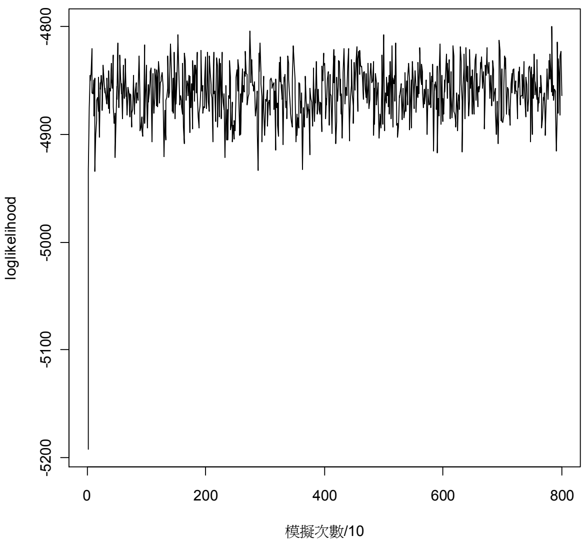

圖 2　概似函數收斂圖

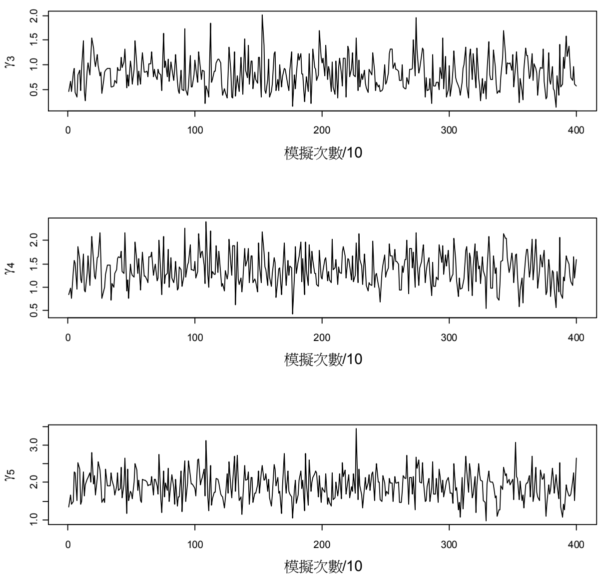

圖 3　γ估計值之收斂圖

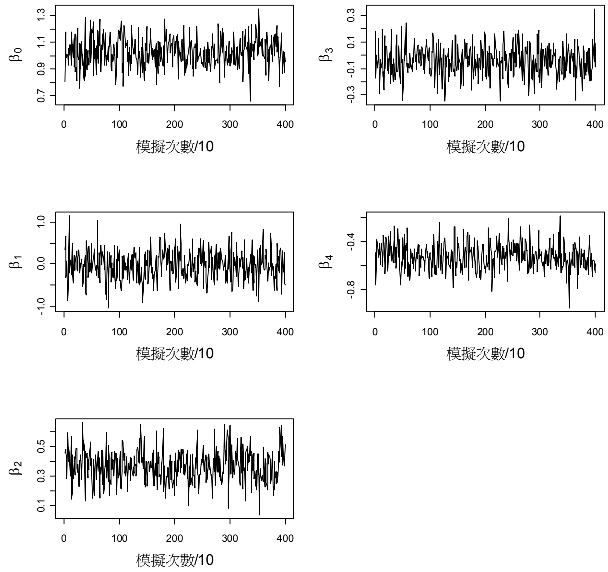

圖 4　β估計值之收斂圖

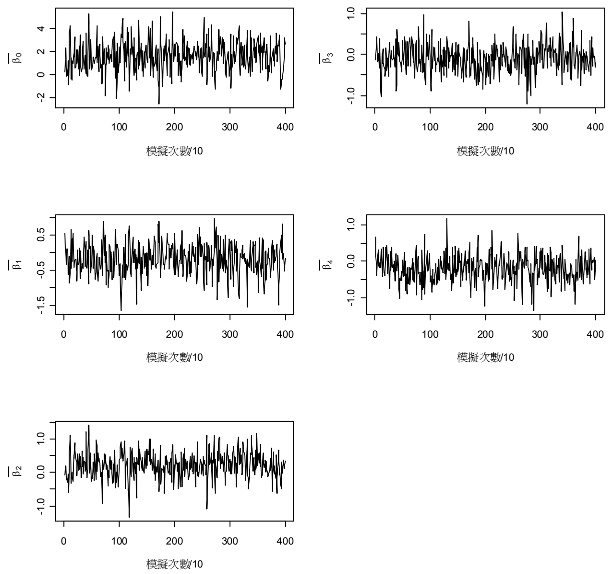

圖 5　$\bar{\beta}$估計值之收斂圖

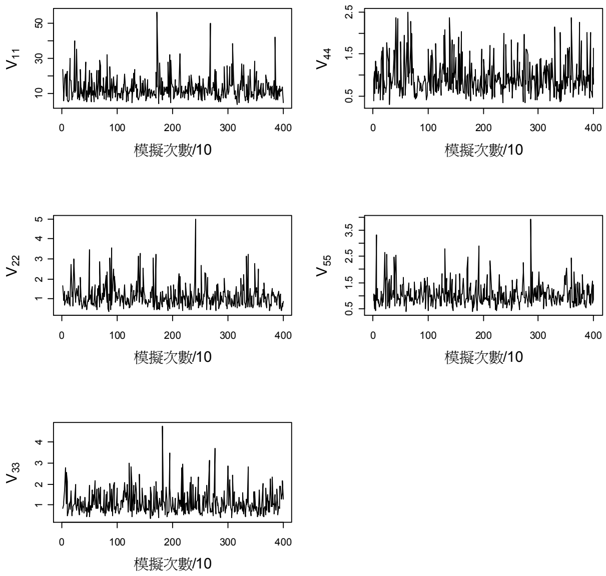

圖 6　V_β估計值之收斂圖

除了收斂圖外，我們逐步解讀模式中的各項參數估計結果以及其行銷意義。首先由圖 7 的盒鬚圖，我們可觀察每位受訪者$\gamma_3,\gamma_4,\gamma_5$的分布情形，因$\gamma_3,\gamma_4,\gamma_5$後驗估計值的變異數分別為 0.236、1.241、2.781，顯示γ_5的變異程度最大，γ_3的變異程度最小。這個結果顯示，在李克特分數越低時，所對應之尺度變異程度越小，受訪者之間的潛在變數Z_t值差異不大;而當受訪者所勾選李克特分數越大時，所對應之尺度變異程度越大，受訪者之間的Z_t值差異較大。這表示，勾選李克特分數越小的受訪者，其生活型態對品牌選擇的解釋度都差不多；反之，勾選李克特分數越大的受訪者，即生活型態對品牌選擇的解釋度較具有明顯差異性。值得注意的是，因為每個人都有自己專屬的尺度區間，且$\gamma_3,\gamma_4,\gamma_5$的後驗估計值，變異相當顯著，故本研究驗證了受訪者在尺度選擇上的異質性，及固定區間的不合理性。

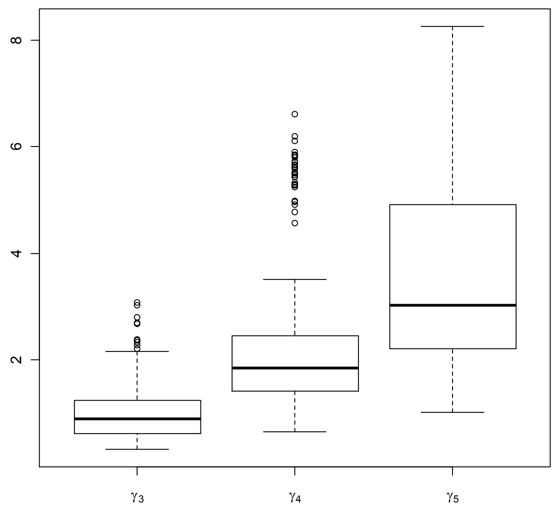

圖 7　γ盒鬚圖

在解釋完每位受訪者對於尺度選擇偏好異質性後，現在解釋每題李克特問項的迴歸式的行銷意涵。如表 2 所示，將具有顯著性的變數以粗體表示。結果可觀察到所有迴歸式的常數項皆具有顯著性，表示當所有二元尺度問項的勾選皆為 0 時，受訪者對於李克特問項所已具備某種認同傾向。例如，就 Q11 而言，其對應之β_0 數值為 3.44，表示每個受訪者對 Q11 的認同傾向的隱藏變數值$z_{h,7}$，至少有 3.44。

另外，表 2 的結果亦顯示，β_2和β_4的估計結果對於李克特問項 Q6「我喜歡嘗試不同的啤酒品牌」和李克特問項 Q8「我期望在未來購買小眾或是進口啤酒」較有顯著性影響。因為β_2和β_4其所對應之生活型態問項分別為「當我和朋友一起喝酒時，我不習慣和朋友分享心內話($x_{h,2} = 1$)」和「我傾向和別人選擇不同的啤酒品牌($x_{h,4} = 1$)」。因此可推論，樂於與人溝通交流內心話，卻很有自己想法、不容易受他人影響的消費者，較喜愛嘗試不同品牌的啤酒，對於小眾品牌和進口品牌的接受程度較高，在挑選啤酒時也會傾向注意特別的產品。

除此之外，β_1和β_4的估計結果對於李克特問項 Q12「由啤酒或酒精飲料品牌的選擇，可以看出一個人的人格特質」也有顯著性影響。因為β_1其所對應之生活型態問項為「狂歡時，我很放得開($x_{h,1} = 1$)」，可推論成功的品牌形象塑造，有助於吸引狂歡時無節制又喜歡與眾不同者的認同。最後，由β_2對李克特問項 Q9「在叫啤酒前，我會希望先知道今天的特別推薦品牌」的顯著性，我們可看出樂於與人溝通交流內心話的受訪者，比較容易受店家特別推薦品牌的影響。

表 2　貝氏模式β參數之平均後驗估計值

參數β	估計值							
	Q6	**Q7**	**Q8**	**Q9**	**Q10**	**Q11**	**Q12**	**Q13**
$\boldsymbol{\beta_0}$	**1.030**	**0.976**	**1.624**	**1.340**	**1.433**	**3.44**	**2.351**	**1.334**
$\boldsymbol{\beta_1}$	-0.029	0.058	-0.025	0.171	-0.619	-0.184	**-0.817**	-0.031
$\boldsymbol{\beta_2}$	**0.370**	0.194	**0.542**	**0.432**	0.116	0.030	0.210	0.095
$\boldsymbol{\beta_3}$	-0.043	-0.099	-0.152	-0.134	-0.113	-0.111	-0.210	0.009
$\boldsymbol{\beta_4}$	**-0.524**	-0.115	**-0.726**	-0.116	0.021	0.073	**-0.343**	-0.093

三、　模型比較

一般來說，目前多變量分析工具多是以常態分配為基礎，因此研究者大多希望資料能趨近常態分配利於分析。雖然統計學上假設當樣本數夠大，研究者可將任何非常態分配資料視為常態分配，以進行統計分析，但是在實務上，行銷資料多屬間斷性，且難從個別消費者採集到夠大的樣本數，更何況透過 asymptotic theory 所建構的統計推論，有限樣本的特性未明，無法保證參數估計值能保有 consistency 和 efficiency 的特性，因此本研究提出以貝氏方法建構模型以解決資料型態與分析假設不符之情形。為了驗證使用貝氏方法估計參數較精確，我們另使用統計軟體 SPSS 18.0 分析以常用之多元迴歸模型同一筆資料。

首先將資料做常態性檢定，由於樣本數達 500 筆，因此使用 Kolmogorov-Smirnov 檢定(K-S 檢定)來檢定樣本是否呈現常態分配。K-S 檢定是一種無母數統計方法檢定，是以樣本分配函數與理論分配函數比較差距為基礎，其基本之虛無假設為樣本呈常態分配，當 p-value 值越大則愈不易拒絕 H_0，樣本越有可能呈常態分配。表 3 顯示所有資料的 p- value 值皆小於 5%之信賴水準，因此 K-S 檢定結果否定了樣本呈常態分配的假設。

接著我們比較使用貝氏模式與使用線性迴歸模式所估計之參數，貝氏模式估計結果參考表 2，而線性迴歸模式結果整理如表 4 所示，所有具統計顯著性之參數估計值皆以粗體加以區別。由於我們證明資料的常態性檢定並不顯著，因此我們可以得到使用一般傳統統計方法所估計的參數，顯示出正值被低估而負值被高估的情況。且在 Q12 的李克特問項中，β_1和β_4兩個參數在貝氏模式的估計結果為顯著；而在線性迴歸模型的估計結果卻為不顯著。因此，透過模型比較，本研究證實了當資料特性與模型假設不符合時，除了得到錯誤或有偏差的估計結果，可能讓行銷研究者無法真正了解消費者的偏好並做出正確的決策行為，因此在選用資料的分析工具上必須小心謹慎。

表 3　樣本之常態性檢定表

樣本	Kolmogorov-Smirnov檢定		
	統計量	自由度	顯著性
Q6	.260	500	**.000**
Q7	.231	500	**.000**
Q8	.154	500	**.000**
Q9	.194	500	**.000**
Q10	.222	500	**.000**
Q11	.370	500	**.000**
Q12	.198	500	**.000**
Q13	.266	500	**.000**

表 4　線性迴歸模式 β 參數估計值之結果

參數β	估計值							
	Q6	**Q7**	**Q8**	**Q9**	**Q10**	**Q11**	**Q12**	**Q13**
$\boldsymbol{\beta_0}$	**2.730**	**2.725**	**3.212**	**3.029**	**2.997**	**4.413**	**3.780**	**2.892**
$\boldsymbol{\beta_1}$	0.284	0.418	0.115	0.707	-0.550	0.206	-0.562	0.274
$\boldsymbol{\beta_2}$	**0.275**	0.089	**0.431**	**0.311**	0.091	-0.011	0.146	0.074
$\boldsymbol{\beta_3}$	-0.043	-0.101	-0.132	-0.167	-0.062	-0.041	-0.172	-0.005
$\boldsymbol{\beta_4}$	**-0.430**	-0.061	**-0.556**	-0.050	0.096	0.055	-0.209	-0.037

伍、　研究結論與建議

本研究提出一行銷計量模型 Bayesian Ordered Probit Model，以解決在分析由李克特量表所蒐集之行銷問卷資料時，所遇到的三大挑戰：第一、當資料與分析方法的統計假設不符合時，可能會造成分析結果有偏差的情況。第二、問卷資料為橫斷性資料，亦即一位受訪者只有一組資料，任何統計方法都無法針對個別受訪者的異質性做分析探討。第三、研究者希望能了解某一組問項對另一組問項的解釋能力。因此，根據實證分析的結果，本研究不但成功利用貝氏 Ordered Probit Model 解決問卷資料與統計方法假設不符的問題，也證實了受訪者在選擇尺度的異質性。本研究也透過模型建構，結合問卷中不同尺度類型的資料，進一步分析彼

此之間的相關性，並向業者提出行銷建議。本研究在後續發展上可朝下列幾點進行：

一、在潛在變數模型的選擇上，未來可以考慮非線性之模型，期以達到最佳的估計結果。但值得注意的是，在行銷研究上講求模型能提供合理之解釋，因此除了最佳化模式的估計結果外， 也需要考量是否合乎現況。

二、由於本研究無估計參數d_h^*之平均數$\bar{d}^*$及變異數A_d^{-1}，無法解釋每位受訪者專屬的尺度區間$[\gamma_1, \gamma_2, \cdots, \gamma_{C+1}]_h$之間的差異程度，期望未來能著重分析受訪者選擇尺度異質性行為的部分。

陸、 參考文獻

1. Allenby, G. M. and P. E. Rossi 1999. Marketing Models of Customer Heterogeneity, Journal of Economtrics, 89, 57-78.
2. Gelfand, A. E., and Smith, A. F. M. 1990. Sampling-Based Ap-proaches to Calculating Marginal Densities, Journal of the American Statistical Association, 85, 398-409.
3. Gelman, Andrew, Carlin, John B., Stern, Hal S., and Rubin, Donald B. 2003. Bayesian Data Analysis, Second Edition. Chapman & Hall/CRC.
4. Hastings, W. K. 1970. Monte Carlo Sampling Methods Using Markov Chains and Their Applications, Biometrika, 57, 97-109.
5. Likert, R. 1932. A Technique for the Measurement of Attitudes, Archives of Psychology, 140, 1-55.
6. Metropolis, N., Rosenbluth, A. W., Rosenbluth,M. N. , Teller,A. H., and Teller, E. 1953. Equations of State Calculations by Fast Compution Machines, Journal of chemical Physics, 21, 1087-1091.
7. Muller, P. 1993. A Generic Approach to Posterior Integration and Gibbs Sampling, Technical Report 91-09, Department of Statustucs, Purdue University.
8. McCulloch, Robert, Nicholas Polson, Peter Rossi. 2000. Bayesian analysis of the multinomial probit model with fully identified parameters, J. Econometrics , 99, 173-193.
9. Marshall, Pablo, Eric T. Bradlow, 2002. A unified approach to conjoint analysis models, Journal of the American Statistical Association, 97, 674-682.

10. Nummelin, E. 1984. General Irreducible Markov Chains and Non-Negative Operator, Cambridge Univ. Press.

11. Rossi, Peter E., Zvi Gilula, and Greg M. Allenby. 2001. Overcoming Scale Usage Heterogeneity: A Bayesian Hierarchical Approach, Journal of the American Statistical Association, Vol. 96, no. 453, 20-31.

12. Rossi, Peter E. and Greg M. Allenby 2003. Bayesian Statistics and Marketing, Marketing Science, 22, 304-328.

13. Rossi, P.E., Allenby, G.M., and McCulloch 2005. *Bayesian Statistics and Marketing*, New York : John Wiley & Sons.

14. Robert, Christian P. and Casella, George. 2010. Monte Carlo Statistical Methods. Springer.

15. Smith, A. F. M., and Gelfand, A. E. 1992. Bayesian Statistics Without Tears: A Sampling-Resampling Perspective, The American Statistician, 46, 84-88.

16. Tanner, M.A and Wong.W.H 1987. The calculation of posterior distributions by data augmentation (wira discussion), J.Amer. Statist. Assoc.82, 528-550.

17. Tierney, L. 1994. Markov Chains for Exploring Posterior Distributions, Annals of Statistics, 22, 1701-1762.