

致謝詞

首先，誠摯的感謝我的指導教授---俞一唐博士，老師專業的指導與督促使我得到很大的收穫與幫助，論文得以順利完成。在此，謹致上最誠摯的敬意與謝意。同時，也感謝張玉媚教授與林彩玉教授在口試時殷切的指正，以專業角度，給予我諸多精闢的見解與修正，使得本論文更臻完善。

感謝系辦各位助教的大力幫助與悉心指導，對我的學習和工作給予了很大的幫助和支持，在我的工作和論文撰寫中，他們都給我提出了寶貴的意見和建議，在此表示深深的謝意。

感謝研究室夥伴謝尚儒、王舒涵、周群益、劉奕，在研究過程中鼎力相助與鼓勵，讓研究過程充滿樂趣。感謝全體研究室同學陪我走過這段忙碌充實又令人難忘的統計研究所生活。

最后，我要感謝我的父母，他們一直是我努力和學習的動力，感謝他們為我所付出一輩子的辛勞，感謝我的哥哥，感謝我的朋友和親人。

摘要

因子篩選實驗因實驗徑個數的不足在混淆狀況複雜的影響下會導致在因子篩選的錯誤。MBMM (modified Box-Meyer method) 為 BMM (Box-Meyer method) 建立在平均模型基礎下的推廣，在一些實際的實驗中，我們可以利用來解決上述的問題。在本研究中，我們將 MBMM 推廣到二元資料的因子篩選實驗。當資料型態為二元資料時，邊際分配將不再具有封閉形式，我們利用蒙地卡羅積分法來計算邊際分配。最後我們以磁磚不良率測驗的實驗為例，比較本研究方法與傳統方法篩選結果。

關鍵字：Box-Meyer method、Modified Box-Meyer method、貝氏分析、因子篩選實驗



目錄

第一章 前言.....	1
第二章 BMM 與 MBMM.....	4
第三章 MBMM 在二元資料上之應用.....	7
第四章 分析實例.....	11
第五章 結論.....	15
參考文獻	



第一章 前言

在一般的工業實驗中，一開始通常會考慮很多個可能會影響實驗結果的因子（factor）。然而，根據經驗真正具體有實質影響力的因子卻只占少數。因此，我們常常會利用因子篩選實驗（screening experiment）從多個因子中找出少數幾個真正有影響力的活性因子（active factor）。由於因子篩選實驗通常只是一個正式實驗的先前試驗，所以在時間跟金錢的考慮下，通常不會使用太多的實驗徑（run）。因此，部分因子設計（fractional factorial design）和正交矩陣設計（orthogonal array design）常被來設計因子篩選實驗。在因子篩選實驗中，實驗徑的個數通常無法提供足夠的自由度來估計各類型的效果（effect），進而造成資料分析上的困難。為了補救實驗徑個數少的影響，一些專家學者提出了以下兩個準則來分析因子篩選實驗（Wu and Hamada, 2009）：

- (一) 階級順序準則（hierarchical ordering principle）：低階的交互作用比高階的交互作用更可能會有顯著的影響。
- (二) 效果遺傳準則（effect heredity principle）：一個交互作用要有顯著的影響，通常此一交互作用裡面至少有一個主效果是有顯著影響的。

根據上面兩個原則，我們可以得到以下常用來分析因子篩選實驗的方法。首先，我們會先估計各個互相垂直的對比（contrast），接者利用常態分配機率圖（normal probability plot）或半常態機率分配圖（half-normal probability plot）找出影響力較大的對比。根據階級順序準則，我們會將其影響歸咎於造成此一對比中彼此混淆（aliased）效果中階數最低的（通常是主效果）。當一對比中出現兩個以上最低階彼此混淆的效果要被選擇時（通常是二階交互作用），在沒有其他專業判斷的情況下，效果遺傳準則將會被用來解決此一問題。此一方法在重要效果都是主效果或混淆狀況不會太複雜的情況下，可以有效地找出真正有影響力的活性因子。

然而當混淆狀況複雜時，例如，有重要的二階交互作用跟主效果彼此混淆時，此一傳統方法將無法篩選出造成此一二階交互作用的活性因子。

為了解決因混淆嚴重的關係導致活性因子的判別錯誤的問題。在貝氏模型平均法（Bayesian model averaging）的架構之下，Box 和 Meyer 提出了 Box-Meyer method（BMM）（Box and Meyer, 1993）來分析因子篩選實驗。由於貝氏模型平均法同時考慮多個模型，在不同的候選模型下，我們還是有機會分別考慮兩個彼此混淆的效果。因此，可以用來解決以上因混淆狀況複雜造成活性因子篩選錯誤的問題。不同於一般貝氏模型平均法在迴歸分析上的應用（Raftery, Madigan, and Hoeting, 1997），BMM 在建構候選模型方面是考慮所有可能活性因子的組合而非所有迴歸模型中效果的組合，因此大幅降低了所需要的計算時間。BMM 的概念非常一般，經由統計模擬和實際應用，它可以應用在多種實驗設計下的因子篩選實驗的分析（Cossari, 2008）。

再給定一個可能重要的活性因子組合之下，BMM 利用分析因子實驗常用的效果模型（effects model）來建模，其考慮的效果模型包含了所有在這個候選模型之下活性因子所形成的主效果、二階交互作用...一直到一個事先所選定的最高階交互作用。然而，再一些例子中，我們可以發現當所選定的最高階交互作用不夠高時，BMM 還是會錯失活性因子的篩選。為了解決此一問題，Yu（2013）提出以平均數模型（means model）替代效果模型的 MBMM（modified Box-Meyer method）來分析篩選實驗。利用 MBMM 來篩選活性因子，我們不需要選定最高階交互作用，所考慮的最高階交互作用完全由實驗本身決定。因此可以解決在使用 BMM 時因誤判最高階交互作用而造成活性因子誤判的問題。此外 MBMM 只考慮在每個實驗徑下的單元平均數（cell means），所需的參數個數最多為實驗徑的個數，相對 BMM 來說個數少的很多。因此，會比 BMM 在計算上節省很多時間。

在本研究中，我們將考慮將 MBMM 推廣到二元資料 (binary data) 的因子篩選實驗。當資料為常態分配時，由於常用的常態分配先驗分配 (prior) 與概似函數 (likelihood) 具有共軛 (conjugate) 的特性，我們可以很容易地計算將無論是 BMM 或 MBMM 中所需的邊際分配 (marginal distribution)。然而，當資料型態為二元資料時，此一共軛特性將不存在，而邊際分配將不再具有封閉形式 (closed form)。因此我們需要一些數值方法來計算其邊際機率。如果考慮透過蒙地卡羅積分法 (Monte Carlo integral) 來計算邊際後驗分配，參數的個數的多寡將會對計算中無論是執行時間與準確度有直接的影響。因此，考慮平均數模型的 MBMM 會比考慮效果模型的 BMM 更適合推廣到二元資料的分析。

本論文的架構如下：我們將在第二章簡單介紹 BMM 和 MBMM 兩種方法。第三章將說明如何把 MBMM 推廣到二元資料因子篩選實驗的分析並推導出所需的計算方法。在第四章中我們會利用第三章中所提出的方法分析一個實際的例子。最後第五章為結論。

第二章 BMM 與 MBMM

考慮一個 p 個因子 n 個實驗徑的因子篩選實驗。考慮所有活性因子可能的集合作為候選模型一共有 $q = 2^p$ 個候選模型。候選模型分別以 M_1, \dots, M_q 表示。

令 $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{ip})'$ 為第 i 次實驗徑的因子設定，其中 x_{ij} 為第 i 次實驗徑下第 j 個因子水準的設定。

根據貝氏理論，在給定資料 $\mathbf{y} = (y_1, \dots, y_n)'$ 下，模型 M_k 為真的後驗機率可以表示為：

$$p(M_k | \mathbf{y}) = \frac{p(M_k) f(\mathbf{y} | M_k)}{\sum_{h=1}^q p(M_h) f(\mathbf{y} | M_h)},$$

其中 $p(M_k)$ 為模型 M_k 為真的先驗機率， $f(\mathbf{y} | M_k)$ 為給定 M_k 下 \mathbf{y} 的邊際密度函數。在先驗機率方面，我們假設 $p(M_k) = \pi^{f_k} (1 - \pi)^{p - f_k}$ 。其中 f_k 為模型 M_k 中活性因子的個數。 π 為任何個別因子為活性因子的機率。通常在實驗中因子為活性因子的個數不會太多，因此 π 通常設定為 0.25。邊際密度函數的取得可以透過

$$f(\mathbf{y} | M_k) = \int f(\mathbf{y} | M_k, \boldsymbol{\theta}_k) f(\boldsymbol{\theta}_k | M_k) d\boldsymbol{\theta}_k$$

來計算，其中 $\boldsymbol{\theta}_k$ 為模型 M_k 中參數所形成的向量。在概似函數部分，BMM 使用效果模型並假設

$$\mathbf{y} | M_k, \boldsymbol{\beta}_k, \lambda \sim N\left(\mathbf{X}_k \boldsymbol{\beta}_k, \frac{1}{\lambda} \mathbf{I}\right),$$

其中 \mathbf{X}_k 為在模型 M_k 下活性因子產生的所有各類效果組成的模型矩陣， $\boldsymbol{\beta}_k$ 為模型 M_k 下對應迴歸係數的向量， λ 為精確度（precision）。在各參數的先驗分配方面，我們假設互相獨立的無資訊先驗分配（non-informative prior）。

BMM 與 MBMM 主要的不同在於概似函數的假設。MBMM 以平均數模型取代效果模型。我們先介紹 MBMM 中所需的符號，令 \mathbf{x}_i^k 為模型 M_k 下， \mathbf{x}_i 在活性因子所生成空間的投影。也就是說當第 j 個因子為活性因子時， \mathbf{x}_i^k 與 \mathbf{x}_i 的第 j 個元素相同；反之，在模型 M_k 下，第 j 個因子不為活性因子時， \mathbf{x}_i^k 的第 j 個元素為 0。令 $\mu(\mathbf{x}_i^k)$ 為在模型 M_k 下，第 i 個實驗徑下的平均數。例如：考慮一個三個因子 A、B、C 的 2^3 因子設計。在 A、B 因子為活性因子的候選模型下，兩個實驗徑 (+、+、+) 和 (+、+、-) 的 A、B 因子有相同水準，因此這兩個實驗徑在此候選模型有相同的平均數。所以我們只需要四個解釋變數 $\mu(+、+、0)$ 、 $\mu(+、-、0)$ 、 $\mu(-、+、0)$ 和 $\mu(-、-、0)$ 如下表 1 所示，作為八個實驗徑實驗的平均數。

表 1

Run	\mathbf{x}_i	\mathbf{x}_i^k	$\mu(\mathbf{x}_i^k)$
1	(-, -, -)	(-, -, 0)	$\mu(-, -, 0)$
2	(-, -, +)	(-, -, 0)	$\mu(-, -, 0)$
3	(-, +, -)	(-, +, 0)	$\mu(-, +, 0)$
4	(-, +, +)	(-, +, 0)	$\mu(-, +, 0)$
5	(+, -, -)	(+, -, 0)	$\mu(+, -, 0)$
6	(+, -, +)	(+, -, 0)	$\mu(+, -, 0)$
7	(+, +, -)	(+, +, 0)	$\mu(+, +, 0)$
8	(+, +, +)	(+, +, 0)	$\mu(+, +, 0)$

在給定 M_k 下，MBMM 假設 \mathbf{y} 的概似函數為

$$\mathbf{y} | M_k, \mu(\mathbf{x}^k), \lambda \sim \mathcal{N}\left(\mu(\mathbf{x}^k), \frac{1}{\lambda} \mathbf{I}_n\right),$$

(1)

其中 $\boldsymbol{\mu}(\mathbf{x}^k) = (\mu(\mathbf{x}_1^k), \dots, \mu(\mathbf{x}_n^k))'$ 。在參數先驗分配方面，我們假設

$$\boldsymbol{\mu}(\mathbf{x}^k) | M_k, \mu, \lambda \sim N\left(\mu \mathbf{1}, \frac{\Delta}{\lambda} \mathbf{R}^k\right), \quad (2)$$

其中 μ 是共同的平均值， $\mathbf{R}^k = (\mathbf{R}_{st}^k)$ 是單元平均數間的相關係數矩陣， Δ 為調整參數。

從第 (1) 和 (2) 式，我們可以得到

$$\mathbf{y} | M_k, \mu, \lambda \sim N\left(\mu \mathbf{1}, \frac{\Delta}{\lambda} (\mathbf{I}_n + \mathbf{R}^k)\right). \quad (3)$$

模型中 \mathbf{R}^k 的選擇方面，我們可以令 $\mathbf{R}^k = \mathbf{I}_n$ ，但是單元平均數中互相獨立的假設太強。一般而言，在模型 M_k 下，不同的兩個實驗徑之間活性因子的相同設定水準越多，可預期兩個對應的單元平均數有越多的關聯。而且隨著活性因子的個數增加，對活性因子的關聯性影響就越小。因此，我們假設在模型 M_k 下第 s 實驗徑與第 t 實驗徑的相關程度為

$$R_{st}^k = \frac{2^{m_{st}^k} - 1}{2^{f_k - 1}},$$

其中 m_{st}^k 為 \mathbf{x}_s^k 與 \mathbf{x}_t^k 的因子之間相同水準的個數。再給定先驗分配 $\pi(\mu) \propto 1$ ， $\pi(\lambda) \propto \frac{1}{\lambda}$ ，

我們可以得到邊際機率

$$f(\mathbf{y} | M_k, \Delta) = c \left(\mathbf{1}' (I + \Delta \mathbf{R}^k)^{-\frac{1}{2}} \right) |I + \Delta \mathbf{R}^k|^{-\frac{1}{2}} \left\{ \frac{2 \mathbf{1}' (I + \Delta \mathbf{R}^k)^{-1} \mathbf{1}}{(\mathbf{1}' (I + \Delta \mathbf{R}^k)^{-1} \mathbf{1})(\mathbf{y}' (I + \Delta \mathbf{R}^k)^{-1} \mathbf{y}) - (\mathbf{1}' (I + \Delta \mathbf{R}^k)^{-1} \mathbf{y})^2} \right\}^{\frac{n-1}{2}},$$

其中 $c = (2\pi)^{-\frac{n-1}{2}} \Gamma\left(\frac{n-1}{2}\right)$ 。最後我們利用因子 k 為活性因子的邊際後驗機率

$$p_k = \sum_{M_h: \text{factor } k \text{ is active}} p(M_h | \mathbf{y}), \quad (4)$$

來評估各因子的重要性， p_k 越大因子 k 為活性因子的可能性越大。

第三章 MBMM 在二元資料上之應用

當資料為二元資料時，我們假設

$$y_{ij} | M_k, \mu(x_i^k) \sim B\left(1, \Phi\left(\mu(x_i^k)\right)\right). \quad (5)$$

其中 $B(n, p)$ 為參數為 n 和 p 之二項式分配， $\Phi(\cdot)$ 是 $N(0, 1)$ 的累積分配函數。參數先驗分配方面，相同地我們假設

$$\mu(x^k) | M_k, \mu, \lambda \sim N\left(\mu \mathbf{1}, \frac{1}{\lambda} \mathbf{R}^k\right). \quad (6)$$

在 MBMM 方法中，概似函數和參數先驗為共軛的情況下，計算的過程中可以容易的取得 \mathbf{y} 的邊際密度函數。但是當資料為二元資料時，不同於前面之 MBMM 方法中的積分，我們可以利用蒙地卡羅積分法推導邊際密度函數。根據 (5) 和 (6)，在給定 λ 、 μ 之下我們可以得到

$$\begin{aligned} f(\mathbf{y} | M_k, \lambda, \mu) &= \int \dots \int f(\mathbf{y}, \mu^*(x^k) | M_k, \lambda, \mu) d\mu^*(x^k) \\ &= \int \dots \int f(\mathbf{y} | M_k, \mu^*(x^k), \lambda, \mu) f(\mu^*(x^k) | M_k, \lambda, \mu) d\mu^*(x^k) \\ &= \int \dots \int C \exp\left(-\frac{\lambda}{2} (\mu^*(x^k) - \mu \mathbf{1})' (\mathbf{R}^{k*})^{-1} (\mu^*(x^k) - \mu \mathbf{1})\right) \\ &\quad \times \prod_{i=1}^n \left(\Phi\left(\mu^*(x_i^k)\right)\right)^{y_i} \left(\Phi\left(1 - \mu^*(x_i^k)\right)\right)^{m - y_i} d\mu^*(x^k), \quad (7) \end{aligned}$$

其中 d 為在 M_k 下 n 個實驗徑中相異單元平均數的個數；

$\mu^*(x^k) = (\mu^*(x^k)_1, \dots, \mu^*(x^k)_d)'$ 為 $\mu(x_i^k)$ 中相異之元素所構成之向量， \mathbf{R}^{k*} 為 $\mu^*(x^k)$

之相關係數矩陣， $C = \lambda^{-\frac{d}{2}} |\mathbf{R}^{k*}|^{-\frac{1}{2}} (2\pi)^{-\frac{d}{2}}$ 。例如在前一章中 2^3 因子實驗的例子中，

在因子 A、B 為活性因子的候選模型下， $d = 4$ ，真正需要考慮的單元平均數

$$\boldsymbol{\mu}^*(\boldsymbol{x}^k) = \begin{pmatrix} \boldsymbol{\mu}^*(\boldsymbol{x}^k)_1 \\ \boldsymbol{\mu}^*(\boldsymbol{x}^k)_2 \\ \boldsymbol{\mu}^*(\boldsymbol{x}^k)_3 \\ \boldsymbol{\mu}^*(\boldsymbol{x}^k)_4 \end{pmatrix} = \begin{pmatrix} \boldsymbol{\mu}(\boldsymbol{x}_1^k) \\ \boldsymbol{\mu}(\boldsymbol{x}_3^k) \\ \boldsymbol{\mu}(\boldsymbol{x}_5^k) \\ \boldsymbol{\mu}(\boldsymbol{x}_7^k) \end{pmatrix}, \text{其中 } \boldsymbol{\mu}(\boldsymbol{x}_1^k) = \boldsymbol{\mu}(-, -, 0), \boldsymbol{\mu}(\boldsymbol{x}_3^k) = \boldsymbol{\mu}(-, +, 0),$$

$$\boldsymbol{\mu}(\boldsymbol{x}_5^k) = \boldsymbol{\mu}(+, -, 0), \boldsymbol{\mu}(\boldsymbol{x}_7^k) = \boldsymbol{\mu}(+, +, 0), \boldsymbol{R}^{k*} = \begin{bmatrix} 1 & \frac{1}{3} & \frac{1}{3} & 0 \\ \frac{1}{3} & 1 & 0 & \frac{1}{3} \\ \frac{1}{3} & 0 & 1 & \frac{1}{3} \\ 0 & \frac{1}{3} & \frac{1}{3} & 1 \end{bmatrix}.$$

為了把 (7) 式中的 $\boldsymbol{\mu}^*(\boldsymbol{x}^k)$ 給積分掉，我們令 $\Phi(\boldsymbol{\mu}^*(\boldsymbol{x}^k)_i) = \boldsymbol{u}_i$ ，所以 $d\boldsymbol{u}_i = \Phi(\boldsymbol{\mu}^*(\boldsymbol{x}^k)_i) d\boldsymbol{\mu}^*(\boldsymbol{x}^k)_i$ 。我們可以得到

$$f(\boldsymbol{y} | M_k, \lambda, \boldsymbol{\mu}) = \int \dots \int \mathcal{C} \exp\left(-\frac{\lambda}{2}(\Phi^{-1}(\boldsymbol{u}) - \boldsymbol{\mu}\mathbf{1})'(\boldsymbol{R}^{k*})^{-1}(\Phi^{-1}(\boldsymbol{u}) - \boldsymbol{\mu}\mathbf{1}))\right) \prod_{i=1}^d \frac{u_i^{a_i}(1-u_i)^{b_i}}{\phi(\Phi^{-1}(u_i))} d\boldsymbol{u}, \quad (8)$$

其中 $\boldsymbol{u} = (u_1, \dots, u_d)'$ ， $\Phi^{-1}(\boldsymbol{u}) = \Phi^{-1}((u_1, \dots, u_d)')$ 。 a_i 為在模型 M_k 下 $\boldsymbol{\mu}(\boldsymbol{x}_i^k)$ 有相同單元平均數對應之 y_i 總和， b_i 為在模型 M_k 下 $\boldsymbol{\mu}(\boldsymbol{x}_i^k)$ 有相同單元平均數對應之 $m - y_i$ 總和。例如前面 2^3 因子實驗的例子只需要四個解釋變數，有相同 $\boldsymbol{\mu}(+, +, 0)$ 的為 $\boldsymbol{\mu}(\boldsymbol{x}_1^k)$ 與 $\boldsymbol{\mu}(\boldsymbol{x}_2^k)$ ，因此 $\boldsymbol{\mu}(+, +, 0)$ 對應之 $a_1 = y_1 + y_2$ 、 $b_1 = 2m - a_1$ 。依此類推 $\boldsymbol{\mu}(+, -, 0)$ 、 $\boldsymbol{\mu}(-, +, 0)$ 和 $\boldsymbol{\mu}(-, -, 0)$ 分別對應之 $a_2 = y_3 + y_4$ 、 $a_3 = y_5 + y_6$ 、 $a_4 = y_7 + y_8$ ， $b_2 = 2m - a_2$ 、 $b_3 = 2m - a_3$ 、 $b_4 = 2m - a_4$ 。

由 (8) 式我們可以看 $f(\boldsymbol{y} | M_k, \lambda, \boldsymbol{\mu})$ 是 $\lambda^{-\frac{d}{2}} |\boldsymbol{R}^{k*}|^{-\frac{1}{2}} (2\pi)^{-\frac{d}{2}} \exp\left(-\frac{1}{2}\left(\lambda(\Phi^{-1}(\boldsymbol{u}) - \boldsymbol{\mu}\mathbf{1})'(\boldsymbol{R}^{k*})^{-1}(\Phi^{-1}(\boldsymbol{u}) - \boldsymbol{\mu}\mathbf{1}) - \Phi^{-1}(\boldsymbol{u})'\Phi^{-1}(\boldsymbol{u}))\right)\right) \prod_{i=1}^d \frac{\Gamma(a_i + b_i + 2)}{\Gamma(a_i + 1, b_i + 1)}$ 的期望值， u_i 是來自互相獨立之 $\text{Beta}(a_i + 1, b_i + 1)$ 分配。因此我們可以藉由蒙地卡羅積分法先隨機抽取多個 u 代入 $\lambda(\Phi^{-1}(\boldsymbol{u}) - \boldsymbol{\mu}\mathbf{1})'(\boldsymbol{R}^{k*})^{-1}(\Phi^{-1}(\boldsymbol{u}) - \boldsymbol{\mu}\mathbf{1}) - \Phi^{-1}(\boldsymbol{u})'\Phi^{-1}(\boldsymbol{u}))$ 中，利用樣本平均數去近似邊際分配。再以上方法中我們必須選定一個適當的方法選取超

參數 λ, μ 。在超參數的取得方面，我們考慮以下模型，

$$y_{ij} | \mu_i \sim B(1, \Phi(\mu_i))$$

$$\mu_i | \mu, \lambda \sim N\left(\mu, \frac{1}{\lambda}\right)$$

其中我們假設參數先驗分配為 $\pi(\mu) \propto 1$ 、 $\pi(\lambda) \propto \frac{1}{\lambda}$ 。為了簡化所需的計算過程，我們

引進一個潛在變數(*latent variable*) z_{ij} 。我們假設 z_{ij} 具有常態分配 $N(\mu_i, 1)$ ，

$y_{ij} = 1$ 如果 z_{ij} 大於0；相反地 $y_{ij} = 0$ 如果 z_{ij} 小於0。我們可以證明出以下之階層

形式模型與上述等價

$$\mu_i | \mu, \lambda \sim N_k\left(\mu, \frac{1}{\lambda}\right)$$

$$z_{ij} | \mu_i \sim N(\mu_i, 1)$$

$$y_{ij} = I[z_{ij} > 0],$$

因此， $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)'$ ， $\mathbf{z} = (z_{11}, \dots, z_{nm})'$ ， $\mathbf{y} = (y_{11}, \dots, y_{nm})'$ 假設我們可以得到

$$\begin{aligned} f(\lambda, \mu, \mathbf{z}, \mathbf{y}) &= \pi(\mu)\pi(\lambda)f(\boldsymbol{\mu}|\mu, \lambda)f(\mathbf{z}|\boldsymbol{\mu})f(\mathbf{y}|\mathbf{z}) \\ &= \lambda^{\frac{n}{2}-1} \exp\left(-\frac{\lambda}{2} \sum_i^n (\mu_i - \mu)^2\right) \exp\left(-\frac{1}{2} \sum_i^n \sum_j^m (z_{ij} - \mu_i)^2\right) \\ &\quad \times \prod_{i=1}^n \prod_{j=1}^m I(z_{ij}, y_{ij}), \end{aligned}$$

因此我們從聯合分配中得到以下各參數的完全條件分配

(full conditional distribution) 為

$$\mu | \cdot \sim N\left(\frac{\sum_{i=1}^n \mu_i}{n}, \frac{\lambda}{n}\right)$$

$$\lambda | \cdot \sim \Gamma\left(\frac{n}{2}, \frac{2}{\sum_{i=1}^n (\mu_i - \mu)^2}\right)$$

$$\mu_i | \cdot \sim N\left(\frac{\lambda\mu + \sum_j^m z_{ij}}{\lambda + m}, \frac{1}{\lambda + m}\right)$$

$$z_{ij} | \cdot \sim TN_{[a,b]}(\mu_i, 1)$$

其中 TN 表示為截斷式常態分配 (truncated normal)， $[a, b]$ 為 z_{ij} 之範圍，當 $y_{ij} = 1$ ，

$z_{ij} \in [0, \infty]$ ；當 $y_{ij} = 0$ ， $z_{ij} \in [-\infty, 0]$ 。因此我們可以利用一般的 Gibbs sampling

algorithm 抽取後驗分配樣本，而在 z_{ij} 的抽取方面我們可以先抽 $U \sim U(0,1)$ ，再代入以下式子

$$\mu_i + \Phi^{-1}[\Phi(a - \mu_i) + U(\Phi(b - \mu_i) - \Phi(a - \mu_i))]$$

得到截斷式常態分配之樣本。最後根據所抽取之後驗分配樣本，取平均得到 λ ， μ 之超參數估計值。



第四章 分析實例

我們考慮以下磁磚不良率測驗實驗，實驗中共有八個實驗徑（ $n=8$ ）用來研究七個兩水準的因子 A、B、C、D、E、F、G。每個實驗徑中都會測試 100 片磁磚（ $m=100$ ）。令 $y_i = \sum_{j=1}^m y_{ij}$ 為第 i 次實驗下不良的總個數。其設計矩陣與反應變數如下表 1 所示。資料可以在 Wu and Hamada（2009）獲得。

表 1 磁磚不良率測驗之實驗設計與反應結果

實驗徑	ABCDEFG	實驗不良個數
1	(-, -, -, -, -, -, -)	16
2	(-, -, -, +, +, +, +)	17
3	(-, +, +, -, -, +, +)	12
4	(-, +, +, +, +, -, -)	6
5	(+, -, +, -, +, -, +)	6
6	(+, -, +, +, -, +, -)	68
7	(+, +, -, -, +, +, -)	42
8	(+, +, -, +, -, -, +)	26

在資料分析方法中，我們首先考慮以傳統分析二元資料中的 probit 迴歸估計各因子的重要性。由於此實驗為一飽和設計，因此在模型中我們只考慮各因子的主效果。其估計值與顯著性結果如表 2 所示。

表 2 因子估計值與顯著水準

因子	估計值	顯著性
A	0.34326	9.84e-10
B	-0.06741	0.23001
C	-0.12787	0.02279
D	0.15519	0.00572
E	-0.24006	1.92e-05
F	0.36050	1.37e-10
G	-0.25548	5.39e-06

我們可以得到除了 B 因子外，其它的因子都有顯著的影響，其重要性依次為 F、A、G、E、D、C。

接著我們以二元資料的 MBMM 對此實驗做分析。在超參數的選擇方面，我們先用第三章節中的 Gibbs sampling algorithm 抽取後驗分配樣本，在扣除 100 個預燒 (burn-in) 後，我們保留 5000 個後驗分配樣本如圖 1 所示。取樣本平均數可以得到 $\hat{\lambda} = 1.9128$ 、 $\hat{\mu} = -0.8272$ 。將其代入 (7) 中計算邊際分配。在所需的蒙地卡羅積分法中，每一個變數，我們抽取 1000 個樣本。再將其邊際機率代入 (4) 中判斷各因子的重要性，其結果如圖 2 所示。

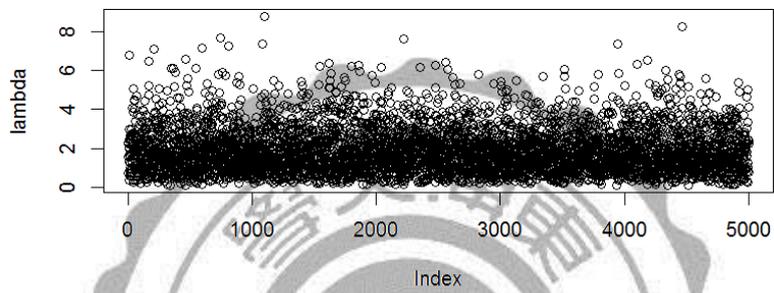
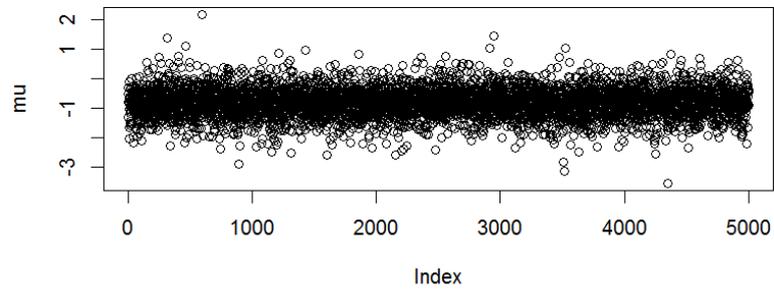


圖 1 λ 、 μ 驗後分配時間序列圖

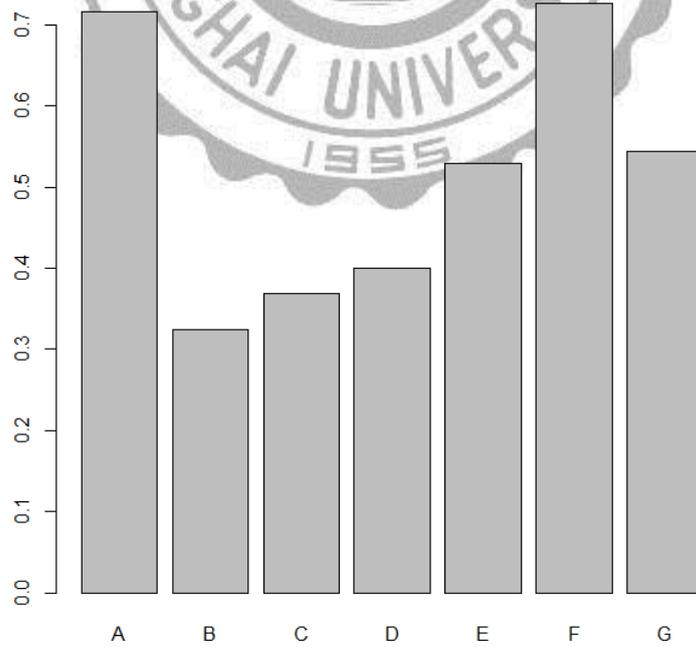


圖 2 活性因子後驗機率

我們可以得到，因子 A-F 為活性因子的後驗機率依次為 0.7138、0.3213、0.3757、0.4045、0.5212、0.7221、0.5451。其重要性依次為 F、A、G、E、D、C、B。與利用傳統 probit 迴歸模型得到的結果相同。



第五章 結論

在工業實驗中的因子篩選實驗，常常有實驗徑個數不多的情況，增加了分析篩選實驗的困難度。在貝氏模型平均法的結構下，**BMM** 被提出解決上面的問題。然而，**BMM** 在一些交互作用很複雜的情況下，如果最高階交互作用判斷錯誤，會錯失活性因子的判斷。為了解決此一問題，Yu (2013) 提出 **MBMM**，以平均模型取代效果模型。在本研究中，我們將 **MBMM** 推廣到分析二元資料的因子篩選實驗。過程中由於先驗分配與概似估計不具共軛性，我們無法得到一個具有封閉形式的邊際機率。為了解決這個問題，我們利用蒙地卡羅積分法來計算邊際機率。透過邊際機率我們能有效的評估各因子的重要性。由第四章結果顯示利用 **MBMM** 分析二元資料因子篩選實驗與傳統的二元資料迴歸所得到的結果一致。此外，**MBMM** 提拱了各因子重要性的順序，我們可以用來考慮在因子篩選實驗後的正式實驗中選擇因子順序的參考。

此一方法可以進一步推廣到其他非常態資料的因子篩選實驗分析，將可做為將來的研究方向。

參考文獻

1. Box, G. E.P. and Meyer, R. D. 1993. *Finding the active factors in fractionated screening experiments*. J. Qual. Technol., 25 : 94–105.
2. Cossari, A. 2008. *Applying Box–Meyer method for analyzing supersaturated designs*. Qual. Technol. Quant. Manage., 5 : 393–401.
3. Montgomery, D. C. 2005. *Design and Analysis of Experiments*, 6, Hoboken, NJ : John Wiley & Sons.
4. Raftery, A. E., Madigan, D. and Hoeting, J. A. 1997. *Bayesian model averaging for linear regression models*. J. Amer. Statist. Assoc., 92 : 179–191.
5. Wu, C. F.J. and Hamada, M. 2009. *Experiments : Planning, Analysis, and Parameter Design Optimization*, 2, Hoboken, NJ : John Wiley & Sons.
6. Yu. I-T. 2013, *A modification of the Box–Meyer method for finding the active factors in screening experiments*, j. appl. statist, 40 : 972-984.