3 Dirichlet 過程參數 (α_0, G_0) 探討

由於 (α_0, G_0) 對我們估計 $\boldsymbol{\lambda}$ 有重要的影響，本節就 α_0、G_0 一些性質做探討，進一步給它們先驗分配，結合資料 $\boldsymbol{Y}$ 來計算 (α_0, G_0) 的後驗分配。

3.1 參數 α_0 的討論

如前所述，Dirichlet 過程的參數 α_0 代表 G 和 G_0 的相似程度，由 Polya urn 的抽樣結構知，如果 α_0 大，λ_i 從 G_0 抽出來的機率 $\alpha_0/(\alpha_0+i-1)$ 較高；反之，如果 α_0 小，λ_i 將有較高的機率等同於已存在的值 $\lambda_1,\ldots,\lambda_{i-1}$。若用 $\boldsymbol{\lambda_n}=(\lambda_1,\ldots,\lambda_n)$ 代表 Polya urn 從第 1 次抽到第 n 次的結果，當 α_0 很大，$\boldsymbol{\lambda_n}$ 約有 n 種相異的值；當 α_0 很小，約所有 $\boldsymbol{\lambda_n}$ 都爲同一種值。若定義隨機變數（Antoniak (1974)）

$$W_i=\begin{cases}1\text{，第 } i \text{ 次的 Polya urn 抽樣值爲一新值，}\\ 0\text{，第 } i \text{ 次的 Polya urn 抽樣值等於已存在的某一值，}\end{cases}\qquad i=1,\ldots,n\text{，}$$

則隨機變數 $K_n=\sum_{i=1}^{n}W_i$，代表 Polya urn 抽取至 n 次時，$\boldsymbol{\lambda_n}$ 有幾種值，也就是第二節所提到的群數，且 K_n 的期望值

$$E[K_n]=\sum_{i=1}^{n}\left[\frac{\alpha_0}{\alpha_0+i-1}\times 1+\frac{i-1}{\alpha_0+i-1}\times 0\right]=\sum_{i=1}^{n}\frac{\alpha_0}{\alpha_0+i-1}\text{。}\tag{3.1}$$

如此一來，只要給定 α_0 值，我們可以估算在此 α_0 的 Polya urn 結構下，抽樣 n 次的預期群數。爲了進一步瞭解 α_0 和群數的關係，我們仿照 Escobar (1994, p271) 探討 α_0 和群數的關係，固定 Polya urn 抽樣次數 n，設定 7 個不同的 α_0 值，利用 (3.1) 式分別估算抽樣 n 次的預期群數，列表如下：

表 1. 固定 $Polya\ urn$ 抽樣次數 n，在7個不同的 α_0 下，預期的群數 $E[K_n]$

n \ α_0 / $E[K_n]$	n^{-3}	n^{-2}	n^{-1}	n^{0}	n^{1}	n^{2}	n^{3}
15	1.0	1.0	1.2	3.3	10.7	14.6	15.0
30	1.0	1.0	1.1	4.0	21.0	29.5	30.0
50	1.0	1.0	1.1	4.5	34.9	49.5	50.0
100	1.0	1.0	1.1	5.2	69.6	99.5	100.0
500	1.0	1.0	1.0	6.8	346.8	499.5	500.0

由表1可知，當 $\alpha_0 \leq n^{-1}$，Polya urn 抽樣結果約爲一群，如果套用第2節坐圓桌例子的說法，此 n 位客人，大概都坐在同一餐桌；當 $\alpha_0 \geq n^2$，n 位客人約坐成了 n 張餐桌，而 $\alpha_0 = n^0$ 及 $\alpha_0 = n^1$ 分別表示約坐成了少數幾張餐桌及許多張餐桌。

有了 α_0 和群數關係認知後，我們在決定參數 α_0 時，便有了一些依據。當觀察值 $\boldsymbol{Y}$「很」集中，我們認爲這些觀察值 $\boldsymbol{Y}$ 皆來自於相同均數的卜瓦松分配，也就是認定 $\lambda_1, \ldots, \lambda_n$ 值都相同，則給定 $\alpha_0 = n^{-1}$；當觀察值 $\boldsymbol{Y}$「很」分散，我們認爲這些觀察值 $\boldsymbol{Y}$ 來自於 n 種不同均數的卜瓦松分配，則給定 $\alpha_0 = n^2$。但在大部分情況下，要從觀察值 $\boldsymbol{Y}$ 判斷出 $\boldsymbol{\lambda}$ 的群數並不容易，這導致我們很難選定一個最佳的 α_0 值。

Escobar (1994) 根據表1，設定四種 α_0 可能值 $\{n^{-1}, n^0, n^1, n^2\}$，分別表示選擇 $\boldsymbol{\lambda}$ 爲一群、少數幾群、許多群及 n 個群，然後在給定觀察值 $\boldsymbol{Y}$ 下，算出 α_0 在這四種可能值的機率，以幫助選擇 α_0 值。計算 α_0 後驗機率的方法將於3.3節介紹。

3.2 參數 G_0 的討論

Dirichlet 過程的參數 G_0 代表混合分配 G 最接近的分配型式。爲了實際計算上的可行性，我們常常使用伽瑪分配（卜瓦松的共軛先驗分配）或均勻分配當成 G_0 的型式。

在決定 Dirichlet 過程參數的過程中，如果利用資料 $\boldsymbol{Y}$ 來決定 G_0 的分配，其實就

是運用了EB法則。若進一步的將G_0設爲伽瑪分配的型式，再用資料$\boldsymbol{Y}$估計此伽瑪分配的參數，便是採用PEB法則；除了共軛先驗分配，G_0還有一種可選擇的型式，非正常先驗分配(improper prior)，即給定λ_i所有可能值相同權重來計算後驗分配。然而卜瓦松均數λ_i的範圍介於0和∞之間，採用非正常先驗分配將導致後驗分配積分上的困難。爲了避免此一困難，我們使用有限範圍的均勻分配來取代之，藉由$Y_1, \ldots, Y_n$的最大、最小值幫助決定均勻分配的範圍。

第4節的模擬研究中，我們將G_0設爲均勻分配的型式。若$u_1 = min(Y_1, \ldots, Y_n)$，$u_2 = max(Y_1, \ldots, Y_n)$，我們選定$U(u_1, u_2)$、$U(u_1 - 1, u_2 + 1)$、$U(u_1 - 2, u_2 + 2)$以及$U(u_1 - 3, u_2 + 3)$四種不同範圍的均勻分配，當成$G_0$的可能分配。

3.3 (α_0, G_0)的先驗和後驗分配

相互搭配3.1節的四種α_0及3.2節的四種G_0，我們得到下表16組參數(α_0, G_0)的可能值：

α_0 \ G_0	$U(u_1, u_2)$	$U(u_1 - 1, u_2 + 1)$	$U(u_1 - 2, u_2 + 2)$	$U(u_1 - 3, u_2 + 3)$
n^{-1}	(α_1, G_1)	(α_1, G_2)	(α_1, G_3)	(α_1, G_4)
n^0	(α_2, G_1)	(α_2, G_2)	(α_2, G_3)	(α_2, G_4)
n^1	(α_3, G_1)	(α_3, G_2)	(α_3, G_3)	(α_3, G_4)
n^2	(α_4, G_1)	(α_4, G_2)	(α_4, G_3)	(α_4, G_4)

其中$(\alpha_1, G_1) = (n^{-1}, U(u_1, u_2))$，$(\alpha_2, G_1) = (n^0, U(u_1, u_2))$，依此類推。

運用貝氏定理，我們給這16組(α_i, G_j)同樣的權重當成其先驗分配，即

$$g(\alpha_i, G_j) = \frac{1}{16}, \qquad i, j = 1, 2, 3, 4。 \tag{3.2}$$

將(3.2)式乘上(α_0, G_0)對觀察值$\boldsymbol{Y}$的概似函數

$$l\{(\alpha_0, G_0) \mid \boldsymbol{Y}\} = \int \prod_{i=1}^{n} e^{-\lambda_i} \frac{\lambda_i^{Y_i}}{Y_i} dF[\boldsymbol{\lambda} \mid (\alpha_0, G_0)], \tag{3.3}$$

其中，$dF[\boldsymbol{\lambda} \mid (\alpha_0, G_0)]$ 的定義與 (2.1) 式同。

如此一來，參數 (α_0, G_0) 的離散型後驗分配有以下的型式：

$$P\{(\alpha_i, G_j) \mid \boldsymbol{Y}\} = \frac{l\{(\alpha_0, G_0) = (\alpha_i, G_j) \mid \boldsymbol{Y}\} \cdot g(\alpha_i, G_j)}{\sum_{i=1}^{4} \sum_{j=1}^{4} l\{(\alpha_0, G_0) = (\alpha_i, G_j) \mid \boldsymbol{Y}\} \cdot g(\alpha_i, G_j)}\text{，} \qquad i, j = 1, 2, 3, 4\text{。} \tag{3.4}$$

在實際計算上，由於積分的困難，我們無法直接將16組 (α_i, G_j) 代入 (3.3) 式求值，也導致 (3.4) 式後驗分配計算上的困難。爲了解決積分問題，Escobar (1995) 提供了一個類似 Polaya urn 的重要抽樣方法，幫助完成 (3.3) 式的估算。

以 $(\alpha_0, G_0) = (\alpha_1, G_1)$ 爲例，依序抽取 $\lambda_1, \ldots, \lambda_n$ 如下：

$i = 1$

$$\lambda_i \sim F_{G_1}(\lambda_i \mid Y_i)\text{，}$$

$i = 2, \ldots, n$

$$\lambda_i \mid \lambda_1, \ldots, \lambda_{i-1} \begin{cases} \sim & F_{G_1}(\lambda_i \mid Y_i)\text{，} & \text{機率正比於} \quad q_{i,0}\text{，} \\ = & \lambda_1\text{，} & \text{機率正比於} \quad q_{i,1}\text{，} \\ = & \lambda_2\text{，} & \text{機率正比於} \quad q_{i,2}\text{，} \\ & \vdots & \vdots \\ = & \lambda_{i-1}\text{，} & \text{機率正比於} \quad q_{i,i-1}\text{，} \end{cases}$$

其中，$dF_{G_1}(\lambda_i \mid Y_i) \propto \lambda_i^{(Y_i+1)-1} e^{-\lambda_i} / \Gamma(Y_i + 1)$，$u_1 < \lambda_i < u_2$，表示先驗分配爲均勻分配 G_1 下，λ_i 的後驗分配，是一有界的伽碼分配；$q_{i0} = \alpha_0 \int_{u_1}^{u_2} [\lambda_i^{(Y_i+1)-1} e^{-\lambda_i} / \Gamma(Y_i + 1)] d\lambda_i$ 表示 α_0 乘上 $dF_{G_1}(\lambda_i \mid Y_i)$ 的標準化常數，且 $q_{ij} = \lambda_j^{Y_i} e^{-\lambda_j} / Y_i$，表示將觀察值 Y_i 帶入分別以 λ_j $(j = 1, 2, \ldots, i-1)$ 當均數之卜瓦松分配的機率值。

重複此抽樣方法 K 次後，得到 K 組 $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_n)$，記爲 $\boldsymbol{\lambda}_{(k)} = (\lambda_{1(k)}, \ldots, \lambda_{n(k)})$，$k = 1, \ldots, K$。Escobar (1995) 利用 $\boldsymbol{\lambda}_{(k)}$，提出 $l\{(\alpha_0, G_0) = (\alpha_1, G_1) \mid \boldsymbol{Y}\}$ 的估計式爲

$$\hat{l}\{(\alpha_0, G_0) = (\alpha_1, G_1) \mid \boldsymbol{Y}\} = \frac{1}{K} \sum_{k=1}^{K} \prod_{i=1}^{n} \frac{q_{i0} + \sum_{j=1}^{i-1} q_{ij}^{(k)}}{\alpha + i - 1}\text{，} \tag{3.5}$$

其中，q_{i0} 的定義如上；$q_{ij}^{(k)}$ 爲 q_{ij} 中的 λ_j 以第 k 組抽樣的 $\lambda_{j(k)}$ 代入的結果。

利用相同方法分別算出其餘15組$l\{(\alpha_i, G_j) \mid \boldsymbol{Y}\}$的估計值，配合(3.2)式相同權重的先驗分配，便可得到(3.4)式離散型後驗分配的估計。有了此離散型後驗分配，我們對第2節的吉氏取樣稍做調整，在執行步驟2.之前，先依照(3.4)式(α_0, G_0)的後驗機率分配抽取一組(α_0, G_0)，在此組(α_0, G_0)值下，再進行步驟2.的過程，最後得到一組$(\lambda_1, \ldots, \lambda_n)$。

直接由觀察值$\boldsymbol{Y}$將$\boldsymbol{\lambda}$分群並不容易，只給G_0唯一型式也非易事，之前我們想找一組最佳的(α_0, G_0)來執行吉氏取樣，顯得不太可行。(α_0, G_0)的離散型後驗分配讓此兩個重要參數在決定上有更多選擇，且幫我們對決定的參數做一檢視，當某一組(α_0, G_0)的後驗分配機率太小，甚至接近零，我們可猜想那是不適合的。

4 無母數經驗貝氏估計法和母數經驗貝氏估計法

本節將以往估計卜瓦松均數$\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_n)$的方法，做一整理介紹，在第5節的模擬研究，我們會拿這些估計方法和本文提出的Dirichlet過程做比較。

4.1 Robbins的估計法

Robbin's (1955)提供了一個在G不知道的情況下估計卜瓦松均數的方法，他將(1.2)式中的$m(Y_i)$、$m(Y_i+1)$，以$\boldsymbol{Y}$中等同於Y_i值及Y_i+1值的個數來做估計，即

$$\hat{\lambda}_i = (Y_i + 1)\frac{\sharp\{Y_k : Y_k = Y_i + 1, k = 1, \ldots, n\}}{\sharp\{Y_k : Y_k = Y_i, k = 1, \ldots, n\}}\text{。} \tag{4.1}$$

並證明，當$n \to \infty$，此估計式的貝氏風險會趨近G爲已知型式下的貝氏風險。然而，在下一節的模擬，我們可以看出，此估計法的估計並不理想。