2 多元伯努利抽樣

考慮一組有n個試驗、I個類別的多元伯努利分配(multiple Bernoulli with I categories)，令$Y_1, Y_2, \ldots, Y_n$分別表示第一個，第二個，...，第n個試驗的變數，並且令θ_i表示任一試驗結果來自第i個類別的機率，其中$i = 1, 2, \ldots, I$，表示為

$$Pr(Y_k = i) = \theta_i\text{，}\quad k = 1, 2, \ldots, n\text{，}$$

其中$\theta_+ = \sum_{i=1}^{I} \theta_i = 1$。

從這樣一組多元伯努利試驗中，很明顯的我們可以看出$\theta_i, i = 1, 2, \ldots, I$為$Y_k$的參數，而在貝氏統計中，我們將$\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_I)$視為隨機變數。考慮多元伯努利的共軛先驗(conjugate prior)Dirichlet分配，假設$\boldsymbol{\theta}$服從參數為$\boldsymbol{a} = (a_1, a_2, \ldots, a_I)$的Dirichlet分配，記作$\boldsymbol{\theta} \sim D(\boldsymbol{a}), \forall a_i > 0$，其機率密度函數為

$$f(\boldsymbol{\theta}; \boldsymbol{a}) \equiv B(\boldsymbol{a})^{-1} \cdot \prod_{i=1}^{I} \theta_i^{a_i - 1}\text{，}\quad \boldsymbol{\theta} \in S\text{，} \tag{2.1}$$

其中$B(\boldsymbol{a}) = \frac{\prod_{i=1}^{I} \Gamma(a_i)}{\Gamma(a_+)}$，$S = \{\boldsymbol{\theta} \mid \forall \theta_i > 0, \theta_+ = 1\}$。然而在實際的試驗中，所得到的試驗結果不見得是真實且完整的，受訪者可能因為某種原因而複選，因此我們將所有的試驗結果考慮進去，假設R_k表示第k個試驗所得到結果，$k = 1, 2, \ldots, n$，則R_k的觀察值是$\{1, 2, \ldots, I\}$的非空子集。在此，我們將不回答與全選的情況在統計上視為具有相同資訊，則試驗結果最多會有$2^I - 1$種。

假設試驗所得到的結果總共有$J (J \leq 2^I - 1)$種不同的報告，以λ_{ij}表示「在真實類別為第i類下，試驗結果為第j類的條件機率」。令$\boldsymbol{\Lambda}$是一個$I \times J$的條件機率矩陣

$$\boldsymbol{\Lambda} = \begin{bmatrix} \lambda_{11} & \lambda_{12} & \cdots & \lambda_{1J} \\ \lambda_{21} & \lambda_{22} & \cdots & \lambda_{2J} \\ \vdots & \vdots & & \vdots \\ \lambda_{I1} & \lambda_{I2} & \cdots & \lambda_{IJ} \end{bmatrix}\text{，}\quad \sum_{j=1}^{J} \lambda_{ij} = 1\text{，}\quad i = 1, 2, \ldots, I\text{，}$$

則得到試驗結果 $R_k = r_k$ 的機率爲

$$
\begin{aligned}
Pr(R_k = r_k \mid \boldsymbol{\theta}, \boldsymbol{\Lambda}) &= \sum_{i=1}^{I} \text{Pr[k-th report=}r_k \mid \text{true category=}i]\cdot \text{Pr[true category=}i] \\
&= \sum_{i=1}^{I} \lambda_{ir_k} \cdot \theta_i\text{，} \quad \text{for } k = 1, 2, \ldots, n\text{。} \qquad (2.2)
\end{aligned}
$$

Dickey, Jiang, and Kadane(1987) 考慮以下列三個假設來處理此貝氏問題：

1. 回答是誠實的。
2. 每個回答 $R_k = r_k$ 在試驗結果集合 r_k 中的類別項之間的差異是無價值的 (noninformative)。
3. 參數 $\boldsymbol{\theta}$ 和每一個 $\boldsymbol{\lambda_i}$ ，$i = 1, 2, \ldots, I$ 是彼此獨立的。

Jiang(1995) 取消了前兩個假設，並令 $I + 1$ 個獨立的隨機向量爲

$$\boldsymbol{\theta} \sim D(\boldsymbol{a})\text{，}$$

$$\boldsymbol{\lambda_i} \sim D(\boldsymbol{b_i})\text{，} \quad i = 1, 2, \ldots, I\text{，}$$

其中 $\boldsymbol{\lambda_i} = (\lambda_{i1}, \lambda_{i2}, \ldots, \lambda_{iJ})$，$\boldsymbol{b_i} = (b_{i1}, b_{i2}, \ldots, b_{iJ})$，則 $\boldsymbol{\theta}$ 和 $\boldsymbol{\Lambda}$ 聯合驗前機率密度函數如下型式

$$Pr(\boldsymbol{\theta}, \boldsymbol{\Lambda}) \propto \left(\prod_{i=1}^{I} \theta_i^{a_i - 1}\right) \cdot \left[\prod_{i=1}^{I} \left(\prod_{j=1}^{J} \lambda_{ij}^{b_{ij}-1}\right)\right]\text{。} \qquad (2.3)$$

當我們接收第一筆 $(k = 1)$ 資料後，可得聯合後驗機率密度函數爲以下型式

$$
\begin{aligned}
P_1(\boldsymbol{\theta}, \boldsymbol{\Lambda} \mid R_1 = r_1) &\propto \left(\prod_{i=1}^{I} \theta_i^{a_i-1}\right) \cdot \left[\prod_{i=1}^{I} \left(\prod_{j=1}^{J} \lambda_{ij}^{b_{ij}-1}\right)\right] \cdot \left(\sum_{m=1}^{I} \lambda_{mr_1} \cdot \theta_m\right) \\
&= \sum_{m=1}^{I} \left\{\left(\prod_{i=1}^{I} \theta_i^{a_i+\delta_i^m-1}\right) \cdot \left[\prod_{i=1}^{I} \left(\prod_{j=1}^{J} \lambda_{ij}^{b_{ij}+\delta_{ij}^{mr_1}-1}\right)\right]\right\}\text{，}
\end{aligned}
$$

其中

$$\delta_i^m = \begin{cases} 1 & ，i = m， \\ 0 & ，其他 \end{cases} \qquad 及 \quad \delta_{ij}^{mr_1} = \begin{cases} 1 & ，i = m \quad 且 \quad j = r_1， \\ 0 & ，其他。 \end{cases}$$

此時後驗機率密度函數亦可以表示為

$$P_1(\boldsymbol{\theta}, \boldsymbol{\Lambda} \mid R_1 = r_1) = \sum_{m=1}^{I} \left\{ \frac{A_m}{A_+} \left[\left(\prod_{i=1}^{I} \theta_i^{a_i + \delta_i^m - 1} \right) \cdot \left(\prod_{i=1}^{I} \prod_{j=1}^{J} \lambda_{ij}^{b_{ij} + \delta_{ij}^{mr_1} - 1} \right) \right] \Big/ A_m \right\}， \tag{2.4}$$

其中

$$A_m = B(\boldsymbol{a} + \boldsymbol{\delta}^m) \cdot \prod_{i=1}^{I} B(\boldsymbol{b}_{i*} + \boldsymbol{\delta}_{i*}^{mr_1})，$$

$$A_+ = \sum_{m=1}^{I} A_m，$$

$$\boldsymbol{a} = (a_1, a_2, \ldots, a_I)，$$

$$\boldsymbol{\delta}^m = (\delta_1^m, \delta_2^m, \ldots, \delta_I^m)，$$

$$\boldsymbol{b}_{i*} = (b_{i1}, b_{i2}, \ldots, b_{iJ})，$$

$$\boldsymbol{\delta}_{i*}^{mr_1} = (\delta_{i1}^{mr_1}, \delta_{i2}^{mr_1}, \ldots, \delta_{iJ}^{mr_1})。$$

同時$\boldsymbol{\theta}$的後驗分配均數、後驗二階動差與後驗變異數分別為

$$E(\theta_i | R_1 = r_1) = \sum_{m=1}^{I} \frac{A_m}{A_+} \frac{a_i + \delta_i^m}{a_+ + 1}， \qquad i = 1, 2, \ldots, I， \tag{2.5a}$$

$$E(\theta_i^2 | R_1 = r_1) = \sum_{m=1}^{I} \frac{A_m}{A_+} \frac{(a_i + \delta_i^m)(a_i + \delta_i^m + 1)}{(a_+ + 1)(a_+ + 2)}， \qquad i = 1, 2, \ldots, I， \tag{2.5b}$$

及

$$Var(\theta_i | R_1 = r_1) = E(\theta_i^2 | R_1 = r_1) - (E(\theta_i | R_1 = r_1))^2， \qquad i = 1, 2, \ldots, I。$$

又當我們接收到第二筆資料時，則將 (2.4) 式當作驗前分配與$k = 2$的 (2.2) 式相乘，得到更新後的後驗機率密度函數為

$$\begin{aligned} P_2(\boldsymbol{\theta}, \boldsymbol{\Lambda} \mid R_1 = r_1, R_2 = r_2) \;\; &\propto \;\; \left(\prod_{i=1}^{I} \theta_i^{a_i - 1} \right) \left[\prod_{i=1}^{I} \left(\prod_{j=1}^{J} \lambda_{ij}^{b_{ij} - 1} \right) \right] \left(\sum_{m_1=1}^{I} \lambda_{m_1 r_1} \cdot \theta_{m_1} \right) \left(\sum_{m_2=1}^{I} \lambda_{m_2 r_2} \cdot \theta_{m_2} \right) \\ &= \;\; \sum_{m_2=1}^{I} \sum_{m_1=1}^{I} \left\{ \left(\prod_{i=1}^{I} \theta_i^{a_i + \delta_i^{m_1} + \delta_i^{m_2} - 1} \right) \cdot \left[\prod_{i=1}^{I} \left(\prod_{j=1}^{J} \lambda_{ij}^{b_{ij} + \delta_{ij}^{m_1 r_1} + \delta_{ij}^{m_2 r_2} - 1} \right) \right] \right\}。 \end{aligned}$$

此時後驗機率密度函數可以表示爲

$$
\begin{aligned}
&P_2(\boldsymbol{\theta}, \boldsymbol{\Lambda} \mid R_1 = r_1, R_2 = r_2) \\
&= \sum_{m_2=1}^{I} \sum_{m_1=1}^{I} \left\{ \frac{A_{m_1 m_2}}{A_{++}} \left[\left(\prod_{i=1}^{I} \theta_i^{a_i + \delta_i^{m_1} + \delta_i^{m_2} - 1} \right) \cdot \left(\prod_{i=1}^{I} \prod_{j=1}^{J} \lambda_{ij}^{b_{ij} + \delta_{ij}^{m_1 r_1} + \delta_{ij}^{m_2 r_2} - 1} \right) \right] \Big/ A_{m_1 m_2} \right\} ,
\end{aligned}
$$

其中

$$
\begin{aligned}
A_{m_1 m_2} &= B(\boldsymbol{a} + \boldsymbol{\delta}^{m_1} + \boldsymbol{\delta}^{m_2}) \cdot \prod_{i=1}^{I} B(\boldsymbol{b}_{i*} + \boldsymbol{\delta}_{i*}^{m_1 r_1} + \boldsymbol{\delta}_{i*}^{m_2 r_2}) , \\
A_{++} &= \sum_{m_2=1}^{I} \sum_{m_1=1}^{I} A_{m_1 m_2} 。
\end{aligned}
$$

同時$\boldsymbol{\theta}$的後驗分配均數、後驗二階動差與後驗變異數分別爲

$$
E(\theta_i \mid R_1 = r_1, R_2 = r_2) = \sum_{m_2=1}^{I} \sum_{m_1=1}^{I} \frac{A_{m_1, m_2}}{A_{++}} \cdot \frac{a_i + \delta_i^{m_1} + \delta_i^{m_2}}{(a_+ + 2)} ,\ i = 1, 2, \ldots, I , \tag{2.6}
$$

$$
E(\theta_i^2 \mid R_1 = r_1, R_2 = r_2) = \sum_{m_2=1}^{I} \sum_{m_1=1}^{I} \frac{A_{m_1, m_2}}{A_{++}} \cdot \frac{(a_i + \delta_i^{m_1} + \delta_i^{m_2})(a_i + \delta_i^{m_1} + \delta_i^{m_2} + 1)}{(a_+ + 2)((a_+ + 2) + 1)} ,\ i = 1, 2, \ldots, I ,
$$

及

$$
Var(\theta_i \mid R_1 = r_1, R_2 = r_2) = E(\theta_i^2 \mid R_1 = r_1, R_2 = r_2) - (E(\theta_i | R_1 = r_1, R_2 = r_2))^2 ,\ i = 1, 2, \ldots, I 。
$$

此時，後驗機率密度函數爲I^2項的$I+1$個Dirichlet分配乘積的混合分配。以此類推，我們可以發現，當我們得到第n筆資料後，後驗機率密度函數將會是I^n項的混合分配；當樣本數增加時，後驗均數的項數會是冪次方成長，以致於我們在運算上會顯得困難且費時。