3 準貝氏法 (quasi-Bayes) 在不完整多元伯努利上應用的介紹

爲了解決貝氏法在計算上的困難，Dr. Jiang(1995) 提出了準貝氏法，即 quasi-Bayes(q-B)，來估計後驗機率密度函數的期望值，以下爲簡單的介紹。

爲了方便說明，我們將先驗機率密度函數 (2.3) 簡化表示爲

$$P_0(\boldsymbol{\theta}, \boldsymbol{\Lambda}) = Q(\boldsymbol{a}^{(0)}) \cdot \prod_{i=1}^{I} Q(\boldsymbol{b}_{i*}^{(0)}) \text{，} \tag{3.1}$$

其中 $Q(\boldsymbol{a}^{(0)}) \equiv f(\boldsymbol{\theta}; \boldsymbol{a}^{(0)})$ 代表參數爲 $\boldsymbol{a}^{(0)}$ 之 Dirichlet 分配的機率密度函數，$Q(\boldsymbol{b}_{i*}^{(0)}) \equiv f(\boldsymbol{\lambda}_{i*}; \boldsymbol{b}_{i*}^{(0)})$ 代表參數爲 $\boldsymbol{b}_{i*}^{(0)}$ 之 Dirichlet 分配的機率密度函數。則我們可以將接收到第一筆資料後的後驗機率密度函數簡化表示爲

$$P_1(\boldsymbol{\theta}, \boldsymbol{\Lambda} \mid R_1 = r_1) = \sum_{i=1}^{I} \frac{A_m}{A_+} \cdot Q(\boldsymbol{a}^{(0)} + \boldsymbol{\delta}^m) \cdot \prod_{i=1}^{I} Q(\boldsymbol{b}_{i*}^{(0)} + \boldsymbol{\delta}_{i*}^{mr_1}) \text{。} \tag{3.2}$$

假若我們知道第一筆資料的真實類別，則後驗機率密度函數將爲

$$Q(\boldsymbol{a}^{(0)} + \boldsymbol{c}^{(1)}) \cdot \prod_{i=1}^{I} Q(\boldsymbol{b}_{i*}^{(0)} + c_i^{(1)} \cdot \boldsymbol{\delta}_{i*}^{ir_1}) \text{，} \tag{3.3}$$

其中 $\boldsymbol{c} = (c_1^{(1)}, c_2^{(1)}, \ldots, c_I^{(1)})$，且第一筆資料的真實類別爲第 i 類時，則 $c_i^{(1)} = 1$ 以及 $c_j^{(1)} = 0$，當 $j \neq i$。

但是事實上我們不可能知道受訪者的真實類別。此時，我們假設 $C_i^{(1)}$ 爲一個隨機變數，表示第一位受訪者的真實類別爲 i 類的機率，則 $C_i^{(1)} = 0$ 或 1。我們便利用條件期望值 $d_i^{(1)} = E[C_i^{(1)} \mid R_1 = r_1] = Pr(C_i^{(1)} = 1 \mid r_1)$ 來估計 $c_i^{(1)}$。由貝氏定理可知，

$$\begin{aligned}
d_i^{(1)} &= E[C_i^{(1)} \mid R_1 = r_1] \\
&= Pr(C_i^{(1)} = 1 \mid r_1) \\
&= \frac{Pr(R_1 = r_1 \mid C_i^{(1)} = 1) \cdot Pr(C_i^{(1)} = 1)}{\sum_{m=1}^{I} [Pr(R_1 = r_1 \mid C_m^{(1)} = 1) \cdot Pr(C_m^{(1)} = 1)]} \text{。}
\end{aligned}$$

令 $\hat{d}_i^{(1)} = [\hat{\lambda}_{ir_1}^{(0)} \cdot \hat{\theta}_i^{(0)}] \Big/ \sum_{m=1}^{I} [\hat{\lambda}_{mr_1}^{(0)} \cdot \hat{\theta}_m^{(0)}]$，其中 $\hat{\lambda}_{ir_1}^{(0)}$ 和 $\hat{\theta}_i^{(0)}$ 分別代表 λ_{ir_1} 和 θ_i 的先驗期望值，我們再利用這樣的 $\hat{d}_i^{(1)}$ 來估計 $d_i^{(1)}$。此時 (3.2) 式便可利用 (3.3) 式來估計，得到以下的式子

$$\begin{aligned} \hat{p}_1(\boldsymbol{\theta}, \boldsymbol{\Lambda} \mid R_1 = r_1) &= Q(\boldsymbol{a}^{(0)} + \hat{\boldsymbol{d}}^{(1)}) \cdot \prod_{i=1}^{I} Q(\boldsymbol{b}_{i*}^{(0)} + \hat{d}_i^{(1)} \cdot \boldsymbol{\delta}_{i*}^{ir_1}) \\ &= Q(\boldsymbol{a}^{(1)}) \cdot \prod_{i=1}^{I} Q(\boldsymbol{b}_{i*}^{(1)})\text{，} \end{aligned} \tag{3.4}$$

其中 $\boldsymbol{a}^{(1)}$ 和 $\boldsymbol{b}_{i*}^{(1)}$ 是接收第一筆資料之後所更新的參數。

比較 (3.1) 式與 (3.4) 式，我們發現兩式具有相同的形式，不同處只在於 (3.1) 式中的 $\boldsymbol{a}^{(0)}$ 與 $\boldsymbol{b}_{i*}^{(0)}$ 在 (3.4) 式中我們以 $\boldsymbol{a}^{(1)}$ 與 $\boldsymbol{b}_{i*}^{(1)}$ 代替。此時 θ_i 與 λ_{ij} 的後驗近似均數 (approximated posterior mean) 爲

$$\hat{\theta}_i^{(1)} = \frac{a_i^{(1)}}{a_+^{(1)}}\text{，}$$

$$\hat{\lambda}_{ij}^{(1)} = \frac{b_{ij}^{(1)}}{b_{i+}^{(1)}}\text{。}$$

當接收到第二筆資料 $R_2 = r_2$ 後，後驗機率密度函數則爲

$$\begin{aligned} \hat{p}_2(\boldsymbol{\theta}, \boldsymbol{\Lambda} \mid R_1 = r_1, R_2 = r_2) &= Q(\boldsymbol{a}^{(1)} + \hat{\boldsymbol{d}}^{(2)}) \cdot \prod_{i=1}^{I} Q(\boldsymbol{b}_{i*}^{(1)} + \hat{d}_i^{(2)} \cdot \boldsymbol{\delta}_{i*}^{ir_2}) \\ &= Q(\boldsymbol{a}^{(2)}) \cdot \prod_{i=1}^{I} Q(\boldsymbol{b}_{i*}^{(2)})\text{，} \end{aligned} \tag{3.5}$$

其中 $\boldsymbol{a}^{(2)} = \boldsymbol{a}^{(1)} + \hat{\boldsymbol{d}}^{(2)}$，$\boldsymbol{b}_{i*}^{(2)} = \boldsymbol{b}_{i*}^{(1)} + \hat{d}_i^{(2)} \cdot \boldsymbol{\delta}_{i*}^{ir_2}$，$\hat{d}_i^{(2)} = [\hat{\lambda}_{ir_2}^{(1)} \cdot \hat{\theta}_i^{(1)}] \Big/ \sum_{m=1}^{I} [\hat{\lambda}_{mr_2}^{(1)} \cdot \hat{\theta}_m^{(1)}]$，$\hat{\boldsymbol{d}}^{(2)} = (\hat{d}_1^{(2)}, \hat{d}_2^{(2)}, \ldots, \hat{d}_I^{(2)})$。以此類推到接收第 n 筆資料 $R_n = r_n$ 後，可得後驗機率分配爲

$$\begin{aligned} \hat{p}_n(\boldsymbol{\theta}, \boldsymbol{\Lambda} \mid R_1 = r_1, R_2 = r_2, \ldots, R_n = r_n) &= Q(\boldsymbol{a}^{(n-1)} + \hat{\boldsymbol{d}}^{(n)}) \cdot \prod_{i=1}^{I} Q(\boldsymbol{b}_{i*}^{(n-1)} + \hat{d}_i^{(n)} \cdot \boldsymbol{\delta}_{i*}^{ir_n}) \\ &= Q(\boldsymbol{a}^{(n)}) \cdot \prod_{i=1}^{I} Q(\boldsymbol{b}_{i*}^{(n)})\text{，} \end{aligned}$$

此時後驗近似均數爲

$$
\begin{aligned}
\hat{\theta}_i^{(n)} &= \frac{a_i^{(n)}}{a_+^{(n)}} = \frac{a_i^{(n-1)} + \hat{P}r(C_i^{(n)} = 1 \mid R_n = r_n)}{a_+^{(0)} + n} = \frac{a_i^{(n-1)} + \hat{d}_i^{(n)}}{a_+^{(0)} + n} \text{ ，} \\
\hat{\lambda}_{ij}^{(n)} &= \frac{b_{ij}^{(n)}}{b_{i+}^{(n)}} \equiv \begin{cases} \dfrac{b_{ij}^{(n-1)} + \hat{P}r(C_i^{(n)} = 1 \mid R_n = r_n)}{b_{i+}^{(n)}} = \dfrac{b_{ij}^{(n-1)} + \hat{d}_i^{(n)}}{b_{i+}^{(n)}} \text{ ，} & \text{for} j = r_n \text{ ，} \\ \dfrac{b_{ij}^{(n-1)}}{b_{i+}^{(n)}} \text{ ，} & \text{for} j \neq r_n \text{ ，} \end{cases}
\end{aligned}
$$

且後驗二階動差爲

$$
\begin{aligned}
\hat{\theta_i^2}^{(n)} &= \frac{a_i^{(n)}(a_i^{(n)} + 1)}{a_+^{(n)}(a_+^{(n)} + 1)} = \frac{[a_i^{(n-1)} + \hat{d}_i^{(n)}][a_i^{(n-1)} + \hat{d}_i^{(n)} + 1]}{(a_+^{(0)} + n)(a_+^{(0)} + n + 1)} \text{ ，} \\
\hat{\lambda^2}_{ij}^{(n)} &= \frac{b_{ij}^{(n)}(b_{ij}^{(n)} + 1)}{b_{i+}^{(n)}(b_{i+}^{(n)} + 1)} \equiv \begin{cases} \dfrac{[b_{ij}^{(n-1)} + \hat{d}_i^{(n)}][b_{ij}^{(n-1)} + \hat{d}_i^{(n)} + 1]}{b_{i+}^{(n)}(b_{i+}^{(n)} + 1)} \text{ ，} & \text{for} j = r_n \text{ ，} \\ \dfrac{b_{ij}^{(n-1)}(b_{ij}^{(n-1)} + 1)}{b_{i+}^{(n)}(b_{i+}^{(n)} + 1)} \text{ ，} & \text{for} j \neq r_n \text{ ，} \end{cases}
\end{aligned}
$$

其中 $a_+^{(0)} = \sum_{i=1}^{I} a_i^{(0)}$，而此近似解即爲本文所謂的準貝氏解。