# 2 計算方法的介紹

爲了方便理解，我們利用一個簡單的例子來做說明。假設一個選舉中有5個候選人，我們的目的就在於預測此5個候選人的得票率。首先，我們先隨機抽樣1000受訪者做調查，假若我們允許受訪者可以複選，並假設得到的結果爲

$$R_1 = r_1, R_2 = r_2, \ldots, R_{1000} = r_{1000}$$

其中$R_i, i = 1, 2, \ldots, 1000,$ 爲$1, 2, \ldots, 5$的非空子集且爲隨機變數，因爲我們允許受訪者可以複選，因此這些回答$r_i, i = 1, 2, \ldots, 1000,$ 是以集合的型態呈現。舉個例子來說，當$r_5 = \{1, 3\}$，代表第五個受訪者可能投給1號或3號候選人，以此類推。

## 2.0 貝氏法在不完整多元伯努利上的應用

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$\theta_i$表示任一試驗結果來自第$i$個類別的機率，其中$i = 1, 2, \ldots, I$，表示爲

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

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

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

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

其中$B(\mathbf{a}) = \frac{\prod_{i=1}^{I} \Gamma(a_i)}{\Gamma(a_+)}$，$S = \{\theta \mid \forall \theta_i > 0, \theta_+ = 1\}$。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，受訪者可能因爲某種原因而複選，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$R_k$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爲具有相

同資訊，則試驗結果最多會有$2^I-1$種。假設試驗結果總共有$J$種不同的報告，我們將這$J$種不同$\{1,2,\ldots,I\}$的非空子集一一對應到不同的$j, j=1,2,\ldots,J$。舉例來說，如果我們的類別數爲3，而得到的試驗有$\{1\}$、$\{2\}$、$\{3\}$及$\{1,2,3\}$這4種不同的結果$(J=4)$，則我們可以令$j=1$時代表試驗結果爲$\{1\}$，$j=2$時代表試驗結果爲$\{2\}$，$j=3$時代表試驗結果爲$\{3\}$，$j=4$時則代表試驗結果爲$\{1,2,3\}$的情況。

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$\lambda_{ij}$表示「在真實類別爲第$i$類下，試驗結果爲第$j$類的條件機率」。令$\mathbf{\Lambda}$是一個$I\times J$的條件機率矩陣

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

則得到試驗結果$R_1=r_1, R_2=r_2,\ldots,R_n=r_n$後的獨立邊際機率分配爲

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

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$r$中的類別項之間的差異是無價值的(noninformative)。
3. 參數$\theta$和$\Lambda$是彼此獨立的。

Jiang(1995) 取消了前兩個假設，並取先驗分配

$$\theta\sim D(\mathbf{a})$$

$$\lambda_{\mathbf{i}*}\sim D(\mathbf{b}_{\mathbf{i}*}) \quad , i=1,2,\ldots,I$$

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

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

則在我們接收第一筆$(k=1)$資料後，可得聯合驗後機率密度函數爲以下型式

$$\begin{aligned} P_1(\theta, \mathbf{\Lambda} \mid R_1 = r_1) &\propto \left(\prod_{i=1}^{I} \theta_i^{a_i-1}\right) \cdot \left[\prod_{i=1}^{I}\left(\prod_{j=1}^{J} \lambda_{ij}^{b_{ij}-1}\right)\right] \cdot \left(\sum_{i=1}^{I} \lambda_{ir_1} \cdot \theta_i\right) \\ &= \sum_{m=1}^{I}\left\{\left(\prod_{i=1}^{I} \theta_i^{a_i+\delta_i^m-1}\right) \cdot \left[\prod_{i=1}^{I}\left(\prod_{j=1}^{J} \lambda_{ij}^{b_{ij}+\delta_{ij}^{mr_1}-1}\right)\right]\right\} \end{aligned}$$

其中

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

此時驗後機率密度函數可以表示爲

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

其中

$$\begin{aligned} &A_m = B(\mathbf{a} + \delta^m) \cdot \prod_{i=1}^{I} B(\mathbf{b}_{i*} + \delta_{i*}^{mr_1}) \\ &A_+ = \sum_{m=1}^{I} A_m \\ &\mathbf{a} = (a_1, a_2, \ldots, a_I) \\ &\delta^m = (\delta_1^m, \delta_2^m, \ldots, \delta_I^m) \\ &\mathbf{b}_{i*} = (b_{i1}, b_{i2}, \ldots, b_{iJ}) \\ &\delta_{i*}^{mr_1} = (\delta_{i1}^{mr_1}, \delta_{i2}^{mr_1}, \ldots, \delta_{iJ}^{mr_1}) \end{aligned}$$

而此時$\theta$的驗後分配均數及後驗二階動差分別爲

$$E\theta_i = \sum_{m=1}^{I} \frac{A_m}{A_+} \cdot \frac{a_i + \delta_i^m}{\sum_{j=1}^{I} a_j + \delta_j^m} \qquad i = 1, 2, \ldots, I \tag{2.5}$$

$$E\theta_i^2 = \sum_{m=1}^{I} \frac{A_m}{A_+} \frac{(a_i + \delta_i^m)(a_i + \delta_i^m + 1)}{[\sum_{j=1}^{I}(a_j + \delta_j^m)][(\sum_{j=1}^{I}(a_j + \delta_j^m)) + 1]}$$

$$Cov(\theta_i \mid R_1 = r_1) = E\theta_i^2 - (E\theta_i)^2$$

當我們在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，得到更新後的驗後機率密度函數爲

$$\begin{aligned} P_2(\theta, \mathbf{\Lambda} \mid R_1 = r_1, R_2 = r_2) \;\; &\propto \;\; \left(\prod_{i=1}^{I} \theta_i^{a_i-1}\right) \cdot \left[\prod_{i=1}^{I}\left(\prod_{j=1}^{J} \lambda_{ij}^{b_{ij}-1}\right)\right] \cdot \left(\sum_{i=1}^{I} \lambda_{ir_1} \cdot \theta_1\right)\left(\sum_{i=1}^{I} \lambda_{ir_2} \cdot \theta_2\right) \\ &= \sum_{m_2=1}^{I}\sum_{m_1=1}^{I}\left\{\left(\prod_{i=1}^{I} \theta_i^{a_i+\delta_i^{m_1}+\delta_i^{m_2}-1}\right) \cdot \left[\prod_{i=1}^{I}\left(\prod_{j=1}^{J} \lambda_{ij}^{b_{ij}+\delta_{ij}^{m_1 r_1}+\delta_{ij}^{m_2 r_2}-1}\right)\right]\right\} \end{aligned}$$

此時驗後機率密度函數可以表示爲

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

其中

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

$\theta$ 的驗後均數及後驗二階動差分別爲

$$E\theta_i = \sum_{m_2=1}^{I}\sum_{m_1=1}^{I} \frac{A_{m_1,m_2}}{A_{++}} \cdot \frac{a_i + \delta_i^{m_1} + \delta_i^{m_2}}{\sum_{j=1}^{I} a_j + \delta_j^{m_1} + \delta_j^{m_2}} \qquad i = 1, 2, \ldots, I \tag{2.6}$$

$$E\theta_i^2 = \sum_{m_2=1}^{I}\sum_{m_1=1}^{I} \frac{A_{m_1,m_2}}{A_{++}} \cdot \frac{(a_i + \delta_i^{m_1} + \delta_i^{m_2})(a_i + \delta_i^{m_1} + \delta_i^{m_2} + 1)}{(\sum_{j=1}^{I} a_j + \delta_j^{m_1} + \delta_j^{m_2})[(\sum_{j=1}^{I} a_j + \delta_j^{m_1} + \delta_j^{m_2}) + 1]} \qquad i = 1, 2, \ldots, I$$

$$Var(\theta_i) = E\theta_i^2 - (E\theta_i)^2$$

此時，驗後機率密度函數將是 $I^2$ 項的 Dirichlet 分配的混合乘積。

以此類推，我們可以發現，當我們得到第 $n$ 筆資料後，驗後機率密度函數將會是 $I^n$ 項的 Dirichlet 分配的混合乘積；當樣本數增加時，驗後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。

## 2.1 準貝氏法在不完整多元伯努利上的應用

爲了解決貝氏方法在計算上的困難，Dr.Jiang 提出了準貝式法來估計驗後機率密度函數的期望值，以下爲簡單的介紹。

爲了方便說明，我們將驗前機率密度函數 (2.3) 簡化表示爲

$$P_0(\theta, \mathbf{\Lambda}) = Q(\mathbf{a}) \cdot \prod_{i=1}^{I} Q(\mathbf{b}_{i*}) \tag{2.7}$$

其中 $Q(\mathbf{a}) \equiv f(\theta; \mathbf{a})$ 代表參數爲 $\mathbf{a}$ 之 Dirichlet 分配的機率密度函數，$Q(\mathbf{b}_{i*}) \equiv f(\lambda_{\mathbf{i}*}; \mathbf{b}_{\mathbf{i}*})$ 代表參數爲 $\mathbf{b}_{\mathbf{i}*}$ 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驗後機率密度函數簡化表示爲

$$P_1(\theta, \mathbf{\Lambda} \mid R_1 = r_1) = \sum_{i=1}^{I} \frac{A_m}{A_+} \cdot Q(\mathbf{a} + \delta^m) \cdot \prod_{i=1}^{I} Q(\mathbf{b}_{i*} + \delta_{i*}^{mr_1}) \tag{2.8}$$

假若我們知道第一筆資料的真實類別，則驗後機率密度函數將爲

$$Q(\mathbf{a} + \mathbf{c}^{(1)}) \cdot \prod_{i=1}^{I} Q(\mathbf{b}_{i*} + c_i^{(1)} \cdot \delta_{i*}^{ir_1}) \tag{2.9}$$

其中 $\mathbf{c} = (c_1^{(1)}, c_2^{(1)}, \ldots, c_I^{(1)})$，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 $c_i^{(1)} = 0, 1$。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，

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

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 $\lambda_{ir_1}$ 和 $\theta_i$ 的驗前期望值，我們再利用這樣的 $\hat{d}_i^{(0)}$ 來估計 $d_i^{(0)}$。

此時(2.8)式便可利用(2.9)式來估計，得到以下的式子

$$\begin{aligned}\hat{p}_1(\theta, \mathbf{\Lambda} \mid R_1 = r_1) &= Q(\mathbf{a} + \hat{\mathbf{d}}^{(1)}) \cdot \prod_{i=1}^{I} Q(\mathbf{b}_{i*} + \hat{d}_i^{(1)} \cdot \delta_{i*}^{ir_1}) \\ &= Q(\mathbf{a}^{(1)}) \cdot \prod_{i=1}^{I} Q(\mathbf{b}_{i*}^{(1)}) \qquad (2.10)\end{aligned}$$

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

比較(2.7)式與(2.10)式，我們發現兩式具有相同的形式，不同處只在於(2.7)式中的$\mathbf{a}$與$\mathbf{b}_{i*}$在(2.10)式中我們以$\mathbf{a}^{(1)}$與$\mathbf{b}_{i*}^{(1)}$代替。此時$\theta_i$與$\lambda_{ij}$的驗後近似均數(approximated posterior mean)將爲

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

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

當接收到第二筆資料$R_2 = r_2$後，驗後機率密度函數則爲

$$\begin{aligned}\hat{p}_2(\theta, \mathbf{\Lambda} \mid R_1 = r_1, R_2 = r_2) &= Q(\mathbf{a}^{(1)} + \hat{\mathbf{d}}^{(2)}) \cdot \prod_{i=1}^{I} Q(\mathbf{b}_{i*}^{(1)} + \hat{d}_i^{(2)} \cdot \delta_{i*}^{ir_2}) \\ &= Q(\mathbf{a}^{(2)}) \cdot \prod_{i=1}^{I} Q(\mathbf{b}_{i*}^{(2)}) \qquad (2.11)\end{aligned}$$

其中$\mathbf{a}^{(2)} = \mathbf{a}^{(1)} + \hat{\mathbf{d}}^{(2)}$，$\mathbf{b}_{i*}^{(2)} = \mathbf{b}_{i*}^{(1)} + \hat{d}_i^{(2)} \cdot \delta_{i*}^{ir_2}$，$\hat{d}_i^{(2)} = [\hat{\lambda}_{ir_2}^{(1)} \cdot \hat{\theta}_i^{(1)}] \Big/ \sum_{m=1}^{I} [\hat{\lambda}_{mr_2}^{(1)} \cdot \hat{\theta}_m^{(1)}]$，$\hat{\mathbf{d}}^{(2)} = (\hat{d}_1^{(2)}, \hat{d}_2^{(2)}, \ldots, \hat{d}_I^{(2)})$。

以此類推到接收第$n$筆資料$R_n = r_n$後，可得驗後機率分配爲

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

此時驗後近似均數爲

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

且

$$
\hat{\theta_i^2}^{(n)} = \frac{a_i^{(n)}(a_i^{(n)} + 1)}{a_+^{(n)}(a_+^{(n)} + 1)} = \frac{[a_i^{(n-1)} + \hat{d}_i^{(n)}][a_i^{(n-1)} + \hat{d}_i^{(n)} + 1]}{(a_+^{(0)} + n)(a_+^{(0)} + n + 1)}
$$

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

## 2.2 吉氏取樣器

### 2.2.1 吉氏取樣器的介紹

吉氏取樣器是一種利用條件分配間接得到邊際分配的工具，它避免掉了在算邊際分配時所遇到的複雜的積分問題，取而代之的是一個以較簡單的計算就可以得到的數列，以下就針對吉氏取樣器做簡單的介紹。

假設我們給定一個聯合密度函數$f(x_1, x_2, \ldots, x_p)$，且已知條件機率密度函數$f(x_s \mid x_r, \forall r \neq s), s = 1, 2, \ldots, p$。如果我們想要得到它的邊際分配，最一般的作法即是利用下列的式子

$$f(x_s) = \int \ldots \int f(x_1, x_2, \ldots, x_p) dx_1 dx_2 \ldots dx_{s-1} d_{s+1} \ldots dx_p \tag{2.12}$$

將邊際分配$f(x_s)$算出。但實際上，在許多的情況下，(2.12) 式是很難積分出來的。此時，我們便可以利用吉氏取樣器來幫助我們間接的得到邊際分配$f(x_s)$。

吉氏取樣器的過程如下

1. 先給定任意的初始值$x_1^{(0)}, x_2^{(0)}, \ldots, x_{p-1}^{(0)}$。
   從$f(X_p \mid X_1 = x_1^{(0)}, \ldots, X_{p-1} = x_{p-1}^{(0)})$分配中隨機取出$x_p^{(0)}$
2. 從$f(X_1 \mid X_2 = x_2^{(0)}, \ldots, X_p = x_p^{(0)})$分配中隨機取出$x_1^{(1)}$，
   再從$f(X_2 \mid X_1 = x_1^{(1)}, X_3 = x_3^{(0)}, \ldots, X_p = x_p^{(0)})$分配中隨機取出$x_2^{(1)}$，
   以此類推，直到從$f(X_p \mid X_1 = x_1^{(1)}, \ldots, X_{p-1} = x_{p-1}^{(1)})$分配中隨機取出$x_p^{(1)}$，此時我們稱完成一次疊代。
3. 在$k$次疊代之後，我們可以得到$(x_1^{(k)}, x_2^{(k)}, \ldots, x_p^{(k)})$。

Geman and Geman(1984) 指出，在適當的條件下，

$$X_s^{(k)} \xrightarrow{d} X_s, \text{ as } k \to \infty$$

其中$X_s$服從$f(x_s)$的分配。所以當$k$夠大時，我們將$x_s^{(k)}$視為$X_s$的模擬觀察值。

吉氏取樣器讓我們避免掉在積分計算上的困難，只需要利用簡單的計算生成一組序列，經過足夠次數的疊代後，就能得到模擬邊際分配的觀察值。再利用電腦模

擬出大量的樣本後，我們便可以利用這樣一組樣本來估計邊際分配的期望值或是變異數等等與邊際相關的資訊。

以兩個隨機變數$(X,Y)$爲例，吉氏取樣器經由交替地從條件分配$f(x \mid y)$和$f(y \mid x)$產生樣本，可以生成一個$X$的樣本，其中$X$服從$f(x)$的分配。過程中形成一個由隨機變數所組成的"Gibbs sequence" 如下

$$Y^{(0)}, X^{(0)}, Y^{(1)}, X^{(1)}, Y^{(2)}, X^{(2)}, \ldots, Y^{(k)}, X^{(k)} \tag{2.13}$$

其中除了$Y^{(0)}$是隨意給定的初始值外，其餘皆以下列規則產生

$$\begin{aligned} X^{(j)} &\sim f(x \mid Y = y^{(j)}) \\ Y^{(j+1)} &\sim f(y \mid X = x^{(j)}) \end{aligned} \tag{2.14}$$

且當$k \to \infty$，則$X^{(k)}$的分配會分配收斂到$X$真正的邊際分配$f(x)$。

### 2.2.2 簡單的收斂說明

設$X,Y$分別爲Bernoulli隨機分配，其聯合機率密度函數如下

$$\begin{bmatrix} f_{xy}(0,0) & f_{xy}(1,0) \\ f_{xy}(0,1) & f_{xy}(1,1) \end{bmatrix} = \begin{bmatrix} p_1 & p_2 \\ p_3 & p_4 \end{bmatrix}$$

對於此分配，我們可以看出$x$的邊際分配

$$f_x = [f_x(0), f_x(1)] = [p_1+p_3, p_2+p_4]$$

是一個成功機率爲$p_2+p_4$的Bernoulli分配。

則我們可以很直接就得到$X \mid Y=y$和$Y \mid X=x$兩個條件機率密度函數

$$A_{y|x} = \begin{bmatrix} \frac{p_1}{p_1+p_3} & \frac{p_3}{p_1+p_3} \\ \frac{p_2}{p_2+p_4} & \frac{p_4}{p_2+p_4} \end{bmatrix} = \begin{bmatrix} P(Y=0 \mid X=0) & P(Y=1 \mid X=0) \\ P(Y=0 \mid X=1) & P(Y=1 \mid X=1) \end{bmatrix}$$

及

$$A_{x|y} = \begin{bmatrix} \frac{p_1}{p_1+p_2} & \frac{p_2}{p_1+p_2} \\ \frac{p_3}{p_3+p_4} & \frac{p_4}{p_3+p_4} \end{bmatrix} = \begin{bmatrix} P(X=0 \mid Y=0) & P(X=1 \mid Y=0) \\ P(X=0 \mid Y=1) & P(X=1 \mid Y=1) \end{bmatrix}$$

$P(X=x \mid Y=y)$表示由$y$轉換到$x$的機率，所以我們可以將矩陣$A_{x|y}$視爲一轉換矩陣，給定由$y$的狀況轉換至$x$的狀況的機率。$A_{y|x}$反之。

若我們只想知道$X$的邊際分配，則在(2.13)的數列中，我們可以很清楚的看出，從$X^{(0)} \to X^{(1)}$的過程中，必須經過$Y^{(1)}$，這樣的過程形成了一個馬可夫鏈，轉換機率爲

$$P(X=x^{(1)} \mid X=x^{(0)}) = \sum_y P(X=x^{(1)} \mid Y=y) \times P(Y=y \mid X=x^{(0)})$$

則$X$數列的轉換矩陣爲

$$A'_{x|x} = A'_{x|y}A'_{y|x} \quad (\text{i.e. } A_{x|x} = A_{y|x}A_{x|y})$$

我們便能簡單的計算出數列中$X^{(k)}$的機率分配。也就是說，$P(X=x^{(k)} \mid X=x^{(0)})$的轉換矩陣爲$(A_{x|x})^k$。若我們以$f_k = [f_k(0) \quad f_k(1)]$來表示$X^{(k)}$的邊際機率密度函數，則對於所有$k$，

$$f_k = f_0 A_{x|x}^{k-1} A_{x|x} = f_{k-1} A_{x|x}$$

Hoel, Port, and Stone(1972) 提到，只要$A$的元素全爲正數，則當$k \to \infty$，$f_k$會收斂到唯一的分配$f$，並滿足

$$fA_{x|x} = f \tag{2.15}$$

所以，若Gibbs sequence收斂，則滿足(2.15)式中的$f$必爲$X$的邊際分配。

當然，在實際操作上，我們無法將疊代次數$k$取到無窮大，那$k$要取如何選取才適當，目前尚未有一肯定答案。在此，我們採用Gelfand and Smith(1990)提出的方法，獨立生成$m$個長度爲$k$的Gibbs sequences，然後取每個數列最後一個$X^{(k)}$。只要$k$夠大，則將$X^{(k)}$視爲$X$獨立的模擬樣本，其中$X$服從$f(x)$的分配。Gelfand and Smith(1990)所處理的$k \leq 50$，$m \leq 1000$，而且他們認爲$k$不需要大於50。

### 2.2.3 吉氏取樣器在不完整多元伯努利上的應用

現在將吉氏取樣器應用於失去部分訊息資料(censored data)上。

假設我們收到$n$筆資料，令爲$\mathbf{r}=(r_1,r_2,\ldots,r_n)$。$\theta\sim D(\mathbf{a})$及$\lambda_{i*}\sim D(b_{i*})$，其中$i=1,2,\ldots,I$爲先驗分配。$\mathbf{w}=\theta\mathbf{\Lambda}$，其中第$j$個元素$w_j$表示報告爲$j$的機率，$j=1,2,\ldots,J$。再假設這$n$筆報告的受訪者真實類別爲$\mathbf{Y}$，$\mathbf{Y}$是一個$1\times n$的隨機向量，其中$Y_k$表示爲第$k$位受訪者的真實類別。在此，我們須先隨機產生出$\mathbf{Y}$的起始值，表示爲$\mathbf{y}^{(0)}=(y_1^{(0)},y_2^{(0)},\ldots,y_n^{(0)})$。我們所取的起始值的方法是假設對第$k$個觀察值而言，在收到$R_k=r_k$的情況下，其真實類別$Y_k$會服從多元伯努利分配，其中各類別機率向量如下

$$\left(\frac{\theta_1^{(0)}\cdot\lambda_{1r_k}^{(0)}}{\sum_{m=1}^{I}\theta_m^{(0)}\cdot\lambda_{mr_k}^{(0)}},\frac{\theta_2^{(0)}\cdot\lambda_{2r_k}^{(0)}}{\sum_{m=1}^{I}\theta_m^{(0)}\cdot\lambda_{mr_k}^{(0)}},\ldots,\frac{\theta_I^{(0)}\cdot\lambda_{Ir_k}^{(0)}}{\sum_{m=1}^{I}\theta_m^{(0)}\cdot\lambda_{mr_k}^{(0)}}\right)$$

利用此機率分配，我們可以生成$y_k^{(0)}$，$k=1,2,\ldots,n$。

接著我們依照下列步驟進行

步驟一、首先對$\mathbf{y}^{(0)}$，$\mathbf{r}$作一個統計，定義下列矩陣

$$\mathbf{x}=(x_1,x_2,\ldots,x_I)\ \text{，}\ \mathbf{z}=\begin{bmatrix} z_{11} & z_{12} & \ldots & z_{1J} \\ z_{21} & z_{22} & \ldots & z_{2J} \\ \vdots & \vdots & & \vdots \\ z_{I1} & z_{I2} & \ldots & z_{IJ} \end{bmatrix}$$

其中$x_i$表示$\mathbf{y}=(y_1,y_2,\ldots,y_n)$中屬於類別$i$的數量。$z_{ij}$表示$n$個配對$[y_k,r_k]$(即[真實類別，報告值])中$y_k=i,r_k=j$的數量，$k=1,2,\ldots,n$。因此第一個步驟後我們可以得到$\mathbf{x}^{(0)}$，$\mathbf{z}^{(0)}$。

步驟二、利用$\mathbf{x}^{(0)}$，$\mathbf{z}^{(0)}$來計算下列兩數

$$\begin{aligned}\mathbf{a}^{(1)} &= \mathbf{a}^{(0)}+\mathbf{x}^{(0)}=(a_1^{(0)}+x_1^{(0)},\ldots,a_I^{(0)}+x_I^{(0)})=(a_1^{(1)},a_2^{(1)},\ldots a_I^{(1)}) \\ \mathbf{b}^{(1)} &= \mathbf{b}^{(0)}+\mathbf{z}^{(0)}=[(b_{ij}^{(0)}+z_{ij}^{(0)})]_{I\times J}=[(b_{ij}^{(1)})]_{I\times J}\end{aligned}$$

步驟三、利用Jiang, Kadane, and Dickey(1992)所提出的Monte Carlo方法來模擬$\theta$以及$\lambda$的觀察值，方法如下

1. 利用自由度為$2a_i^{(1)}$的卡方分配來生成$I$個隨機變數$u_i^{(1)}$，$i=1,2,\ldots,I$。
2. 令$\theta_i^{(1)}=\frac{u_i^{(1)}}{\sum_{m=1}^{I}u_m^{(1)}}$，此時可將$\theta^{(1)}$視為參數為$\mathbf{a}^{(1)}$之Dirichlet分配的觀察值。

同樣的，對每個$i$，$i=1,2,\ldots,I$，當$i$固定的情況下，

1. 利用自由度為$2b_{ij}^{(1)}$的卡方分配來生成$J$個隨機變數$g_{ij}^{(1)}$，$j=1,2,\ldots,J$。
2. 令$\lambda_{ij}^{(1)}=\frac{g_{ij}^{(1)}}{\sum_{m=1}^{J}g_{im}^{(1)}}$，此時可將$\lambda_{i*}^{(1)}$視為參數為$\mathbf{b}_{i*}^{(1)}$之Dirichlet分配的觀察值。

步驟四、此一步驟又分成$n$個小步驟。對第一個觀察值而言，在收到$R_1=r_1$的情況下，真實類別$\mathbf{Y}_1$會服從多元伯努利分配，其中各類別機率向量如下

$$\left(\frac{\theta_1^{(1)}\cdot\lambda_{1r_1}^{(1)}}{\sum_{m=1}^{I}\theta_m^{(1)}\cdot\lambda_{mr_1}^{(1)}},\frac{\theta_2^{(1)}\cdot\lambda_{2r_1}^{(1)}}{\sum_{m=1}^{I}\theta_m^{(1)}\cdot\lambda_{mr_1}^{(1)}},\ldots,\frac{\theta_I^{(1)}\cdot\lambda_{Ir_1}^{(1)}}{\sum_{m=1}^{I}\theta_m^{(1)}\cdot\lambda_{mr_1}^{(1)}}\right)$$

利用此機率分配，我們可生成$y_1^{(1)}$。同樣的，對第2個觀察值而言，在收到$R_2=r_2$的情況下，真實類別$\mathbf{Y}_2$會服從多元伯努利分配，其中各類別機率向量如下

$$\left(\frac{\theta_1^{(1)}\cdot\lambda_{1r_2}^{(1)}}{\sum_{m=1}^{I}\theta_m^{(1)}\cdot\lambda_{mr_2}^{(1)}},\frac{\theta_2^{(1)}\cdot\lambda_{2r_2}^{(1)}}{\sum_{m=1}^{I}\theta_m^{(1)}\cdot\lambda_{mr_2}^{(1)}},\ldots,\frac{\theta_I^{(1)}\cdot\lambda_{Ir_2}^{(1)}}{\sum_{m=1}^{I}\theta_m^{(1)}\cdot\lambda_{mr_2}^{(1)}}\right)$$

利用此機率分配，我們再生成$y_2^{(1)}$。以此類推，生成$y_k^{(1)}$，$k=1,2,\ldots,n$之後，得到$\mathbf{y}^{(1)}=(y_1^{(1)},y_2^{(1)},\ldots,y_n^{(1)})$。如此便完成一次疊代。

接下來進行第二次疊代，重複上列步驟，即

1. 利用$\mathbf{y}^{(1)}$計算出$\mathbf{x}^{(1)}$、$\mathbf{r}^{(1)}$。
2. 把$\mathbf{x}^{(1)}$、$\mathbf{r}^{(1)}$分別與$\mathbf{a}^{(1)}$、$\mathbf{b}^{(1)}$相加形成$\mathbf{a}^{(2)}$、$\mathbf{b}^{(2)}$。
3. 將$2a_i^{(2)}$、$2b_{ij}^{(2)}$當成自由度以Monte Carlo生成$\theta^{(2)}$、$\lambda_{i*}^{(2)}$。
4. 用$\theta^{(2)}$、$\lambda_{i*}^{(2)}$計算出第$k$個報告其多元伯努利分配中各類別的機率，再生成$y_k^{(2)}$，$k=1,2,\ldots,n$，完成第二次疊代。

依照同樣步驟繼續疊代下去，經過$k$次疊代後，若將$\theta$，$\lambda$視為(2.13)中的$X$，真實類別$Y$視為(2.13)中的$Y$，則可形成數列如下

$$\mathbf{y}^{(0)}\to\theta^{(1)},\lambda^{(1)}\to\mathbf{y}^{(1)}\to\theta^{(2)},\lambda^{(2)}\to\cdots\to\theta^{(k)},\lambda^{(k)}$$

當$k \to \infty$時，$\theta^{(k)}, \lambda^{(k)}$會趨近於真實的機率。

依照Gelfand and Smith(1990)所提出的方法，因此我們採用疊代$k$次，取最後一次疊代的值$\theta^{(k)}$，$\lambda^{(k)}$。然後重新給定一起始值，再疊代$k$次，同樣也取最後一次的估計值。我們總共給$m$次起始值，所以也會有$m$個不同的$\theta^{(k)}$，$\lambda^{(k)}$，將這些數的平均做爲Gibbs對$\theta$，$\lambda$的估計值。例如：

$$E(\theta) \doteq \frac{\sum_{i=1}^{m} \theta_i}{m}$$

$$E(\theta^2) \doteq \frac{\sum_{i=1}^{m} \theta_i^2}{m}$$

依照柯力文(2003)對選取$k$值的大小所做的模擬結果，發現$k$值取到20時差異已經不大，所以在本文我們沿用柯力文(2003)對$k$與$m$的選取，取$k = 20$，$m = 500$來做模擬。