4 平均變異數和(average variance sum)的介紹

4.1 平均變異數和在不完整多元伯努利上的應用

Kuroda, Geng and Niki(2001)提出利用平均變異數和，即average variance sum(AVS)，一種新的計算方法，來逼近後驗機率密度函數母數的估計值，並將此方法應用於貝氏逐樣學習的可分解之圖狀模型上的不完整訊息。我們首先嘗試將此估計方法應用於失去部分訊息的資料上。爲了和準貝氏法所使用的參數做區別，在本節中，我們將2.3式的參數符號由$\boldsymbol{a}$改成$\boldsymbol{\alpha}$，$\boldsymbol{b}$改成$\boldsymbol{\beta}$，以方便讀者閱讀。以下爲簡單的介紹。

由於$\boldsymbol{\theta}$和$\boldsymbol{\lambda}$的聯合先驗機率密度函數

$$Pr(\boldsymbol{\theta}, \boldsymbol{\Lambda}) \propto \left(\prod_{i=1}^{I} \theta_i^{\alpha_i^{(0)}-1} \right) \cdot \left[\prod_{i=1}^{I} \left(\prod_{j=1}^{J} \lambda_{ij}^{\beta_{ij}^{(0)}-1} \right) \right] \tag{4.1}$$

是一個Dirichlet分配乘積。而在接受到第一筆資料後得到的聯合後驗機率密度函數表示如下

$$P_1(\boldsymbol{\theta}, \boldsymbol{\Lambda} \mid R_1 = r_1) = \sum_{m=1}^{I} \left\{ \frac{A_m^{(0)}}{A_+^{(0)}} \left[\left(\prod_{i=1}^{I} \theta_i^{\alpha_i^{(0)}+\delta_i^m-1} \right) \cdot \left(\prod_{i=1}^{I} \prod_{j=1}^{J} \lambda_{ij}^{\beta_{ij}^{(0)}+\delta_{ij}^{mr_1}-1} \right) \right] \Big/ A_m^{(0)} \right\} 。 \tag{4.2}$$

此時上式也是一個Dirichlet分配乘積的混合分配。因此$\boldsymbol{\theta}$的後驗分配均數，後驗二階動差及後驗變異數分別利用2.5a及2.5b可進一步表示爲

$$\begin{aligned} E(\theta_i | R_1 = r_1) &= \sum_{m=1}^{I} \frac{A_m^{(0)}}{A_+^{(0)}} \cdot \frac{\alpha_i^{(0)} + \delta_i^m}{\alpha_+^{(0)} + 1} \\ &= \frac{1}{A_+^{(0)}(\alpha_+^{(0)} + 1)} \cdot \left(\sum_{m=1}^{I} A_m^{(0)} \alpha_i^{(0)} + \sum_{m=1}^{I} A_m^{(0)} \delta_i^m \right) \\ &= \frac{1}{A_+^{(0)}(\alpha_+^{(0)} + 1)} \cdot \left(A_+^{(0)} \alpha_i^{(0)} + A_i^{(0)} \right) \\ &= \frac{\alpha_i^{(0)}}{\alpha_+^{(0)} + 1} + \left(\frac{A_i^{(0)}}{A_+^{(0)}} \right) \cdot \frac{1}{\alpha_+^{(0)} + 1} , \quad i = 1, 2, \ldots, I , \end{aligned} \tag{4.3}$$

$$
\begin{aligned}
&E(\theta_i^2|R_1=r_1)\\
&=\sum_{m=1}^{I}\frac{A_m^{(0)}}{A_+^{(0)}}\frac{(\alpha_i^{(0)}+\delta_i^m)(\alpha_i^{(0)}+\delta_i^m+1)}{(\alpha_+^{(0)}+1)(\alpha_+^{(0)}+2)}\\
&=\frac{1}{A_+^{(0)}(\alpha_+^{(0)}+1)(\alpha_+^{(0)}+2)}\left(\alpha_i^{(0)}(\alpha_i^{(0)}+1)A_+^{(0)}+2\alpha_i^{(0)}\sum_{m=1}^{I}A_m^{(0)}\delta_i^m+\sum_{m=1}^{I}A_m^{(0)}(\delta_i^m)^2+\sum_{m=1}^{I}A_m^{(0)}\delta_i^m\right)\\
&=\frac{1}{A_+^{(0)}(\alpha_+^{(0)}+1)(\alpha_+^{(0)}+2)}\left(\alpha_i^{(0)}(\alpha_i^{(0)}+1)A_+^{(0)}+2(\alpha_i^{(0)}+1)A_i^{(0)}\right)\\
&=\frac{\alpha_i^{(0)}(\alpha_i^{(0)}+1)}{(\alpha_+^{(0)}+1)(\alpha_+^{(0)}+2)}+\left(\frac{A_i^{(0)}}{A_+^{(0)}}\right)\frac{2(\alpha_i^{(0)}+1)}{(\alpha_+^{(0)}+1)(\alpha_+^{(0)}+2)}，\qquad i=1,2,\ldots,I，
\end{aligned}
\tag{4.4}
$$

及

$$
Var(\theta_i|R_1=r_1)=E(\theta_i^2|R_1=r_1)-(E(\theta_i|R_1=r_1))^2，\quad i=1,2,\ldots,I。\tag{4.5}
$$

同理$\boldsymbol{\lambda}$的後驗分配均數，後驗二階動差及後驗變異數亦可分別表示爲

$$
\begin{aligned}
E(\lambda_{ij}|R_1=r_1) &= \sum_{m=1}^{I}\frac{A_m^{(0)}}{A_+^{(0)}}\cdot\frac{\beta_{ij}^{(0)}+\delta_{ij}^{mr_1}}{\sum_{t=1}^{J}\beta_{it}^{(0)}+\delta_{it}^{mr_1}}\\
&= \sum_{m=1}^{I}\frac{A_m^{(0)}}{A_+^{(0)}}\cdot\frac{\beta_{ij}^{(0)}+\delta_{ij}^{mr_1}}{\beta_{i+}^{(0)}+\delta_{ir_1}^{mr_1}}，\quad i=1,2,\ldots,I，\quad j=1,2,\ldots,J，
\end{aligned}
\tag{4.6}
$$

$$
\begin{aligned}
E(\lambda_{ij}^2|R_1=r_1) &= \sum_{m=1}^{I}\frac{A_m^{(0)}}{A_+^{(0)}}\cdot\frac{(\beta_{ij}^{(0)}+\delta_{ij}^{mr_1})(\beta_{ij}^{(0)}+\delta_{ij}^{mr_1}+1)}{[\sum_{t=1}^{J}\beta_{it}^{(0)}+\delta_{it}^{mr_1}][(\sum_{t=1}^{J}\beta_{it}^{(0)}+\delta_{it}^{mr_1})+1]}\\
&= \sum_{m=1}^{I}\frac{A_m^{(0)}}{A_+^{(0)}}\cdot\frac{(\beta_{ij}^{(0)}+\delta_{ij}^{mr_1})(\beta_{ij}^{(0)}+\delta_{ij}^{mr_1}+1)}{(\beta_{i+}^{(0)}+\delta_{ir_1}^{mr_1})(\beta_{ir_1}^{(0)}+\delta_{ir_1}^{mr_1}+1)}，\quad i=1,2,\ldots,I，\quad j=1,2,\ldots,J，
\end{aligned}
\tag{4.7}
$$

及

$$
Var(\lambda_{ij}|R_1=r_1)=E(\lambda_{ij}^2|R_1=r_1)-(E(\lambda_{ij}|R_1=r_1))^2，\quad i=1,2,\ldots,I，\quad j=1,2,\ldots,J。\tag{4.8}
$$

在此先定義$\boldsymbol{\theta}$和$\boldsymbol{\lambda}_{i*}$的平均變異數和(AVS)分別爲：

$$
\begin{aligned}
V_{\boldsymbol{\theta}} &= \sum_{i=1}^{I}E(\theta_i)Var(\theta_i)，\\
V_{\boldsymbol{\lambda}_{i*}} &= \sum_{J=1}^{J}E(\lambda_{ij})Var(\lambda_{ij})，\quad i=1,2,\ldots,I。
\end{aligned}
$$

而在本節我們主要的目的其實是利用一個比較簡單的Dirichlet分配的乘積(single hyper Dirichlet distribution)

$$\hat{p}_1(\theta,\Lambda)=\left[\frac{1}{B(\boldsymbol{\alpha}^{(1)})}\cdot\prod_{i=1}^{I}\theta_i^{\alpha_i^{(1)}-1}\right]\cdot\left[\prod_{i=1}^{I}\left(\frac{1}{B(\boldsymbol{\beta}_{i*}^{(1)})}\prod_{j=1}^{J}\lambda_{ij}^{\beta_{ij}^{(1)}-1}\right)\right] \tag{4.9}$$

去逼近真實的後驗分配(4.2)，並且分別用$\hat{p}_1(\theta,\Lambda)$的參數$\boldsymbol{\alpha}^{(1)}$、$\boldsymbol{\beta}^{(1)}$去估計真實後驗分配的參數$\boldsymbol{\alpha}^{(0)}$和$\boldsymbol{\beta}^{(0)}$，也就是用$\hat{p}_1(\theta,\Lambda)$的估計值去逼近真實後驗分配的估計值。此時$\boldsymbol{\theta}$和$\boldsymbol{\lambda}$的期望值及變異數分別爲

$$E_{\hat{p}_1}(\theta_i)=\frac{\alpha_i^{(1)}}{\alpha_+^{(1)}}\text{，} \tag{4.10a}$$

$$Var_{\hat{p}_1}(\theta_i)=\frac{\alpha_i^{(1)}(\alpha_+^{(1)}-\alpha_i^{(1)})}{(\alpha_+^{(1)})^2(\alpha_+^{(1)}+1)}\text{，} \tag{4.10b}$$

$$E_{\hat{p}_1}(\lambda_{ij})=\frac{\beta_{ij}^{(1)}}{\beta_{i+}^{(1)}}\text{，} \tag{4.11a}$$

$$Var_{\hat{p}_1}(\lambda_{ij})=\frac{\beta_{ij}^{(1)}(\beta_{i+}^{(1)}-\beta_{ij}^{(1)})}{(\beta_{i+}^{(1)})^2(\beta_{i+}^{(1)}+1)}\text{。} \tag{4.11b}$$

同時$\boldsymbol{\theta}$及$\boldsymbol{\lambda}_{i*}$的平均變異數和(AVS)可分別得到如下

$$V_{\boldsymbol{\theta}}^{(1)}=\sum_{i=1}^{I}E_{\hat{p}_1}(\theta_i)\cdot Var_{\hat{p}_1}(\theta_i)=\sum_{i=1}^{I}(\frac{\alpha_i^{(1)}}{\alpha_+^{(1)}})^2(1-\frac{\alpha_i^{(1)}}{\alpha_+^{(1)}})\frac{1}{\alpha_+^{(1)}+1}\text{，} \tag{4.12a}$$

$$V_{\boldsymbol{\lambda_i}}^{(1)}=\sum_{j=1}^{J}E_{\hat{p}_1}(\lambda_{ij})\cdot Var_{\hat{p}_1}(\lambda_{ij})=\sum_{j=1}^{J}(\frac{\beta_{ij}^{(1)}}{\beta_{i+}^{(1)}})^2(1-\frac{\beta_{ij}^{(1)}}{\beta_{i+}^{(1)}})\frac{1}{\beta_{i+}^{(1)}+1}\text{。} \tag{4.12b}$$

AVS的計算方法在於利用簡單的分配來近似原本複雜的分配，亦即尋找(4.9)式中的估計參數$\boldsymbol{\alpha}^{(1)}$及$\boldsymbol{\beta}_{i*}^{(1)}$，使得後驗均數等於後驗近似均數而且後驗平均變異數和等於後驗近似平均變異數和，亦即

$$\begin{cases} E_{\hat{p}_1}(\theta_i)\equiv E(\theta_i|R_1=r_1)\text{，} & \text{for } i=1,\cdots,I\text{，} \\ V_{\boldsymbol{\theta}}^{(1)}\equiv\sum_{i=1}^{I}E(\theta_i|R_1=r_1)\cdot Var(\theta_i|R_1=r_1)\text{，} & \end{cases} \tag{4.13}$$

且對所有i

$$\begin{cases} E_{\hat{p}_1}(\lambda_{ij}) \equiv E(\lambda_{ij}|R_1=r_1)\text{，} & \text{for } j=1,\cdots,J\text{，} \\ V_{\boldsymbol{\lambda_i}}^{(1)} \equiv \sum_{i=1}^{I} E(\lambda_{ij}|R_1=r_1)\cdot Var(\lambda_{ij}|R_1=r_1)\text{。} & \end{cases} \tag{4.14}$$

由(4.10a)、(4.12a)及(4.13)可得到

$$\begin{cases} \frac{\alpha_i^{(1)}}{\alpha_+^{(1)}} \equiv E(\theta_i|R_1=r_1)\text{，} & \text{for } i=1,\cdots,I\text{，} \\ \sum_{i=1}^{I}(\frac{\alpha_i^{(1)}}{\alpha_+^{(1)}})^2(1-\frac{\alpha_i^{(1)}}{\alpha_+^{(1)}})\frac{1}{\alpha_+^{(1)}+1} \equiv \sum_{i=1}^{I} E(\theta_i|R_1=r_1)\cdot Var(\theta_i|R_1=r_1)\text{，} & \end{cases}$$

且由(4.11a)、(4.12b)及(4.14)可得到對所有i，

$$\begin{cases} \frac{\beta_{ij}^{(1)}}{\beta_{i+}^{(1)}} \equiv E(\lambda_{ij}|R_1=r_1)\text{，} & \text{for } j=1,2,\cdots,J\text{，} \\ \sum_{j=1}^{J}(\frac{\beta_{ij}^{(1)}}{\beta_{i+}^{(1)}})^2(1-\frac{\beta_{ij}^{(1)}}{\beta_{i+}^{(1)}})\frac{1}{\beta_{i+}^{(1)}+1} \equiv \sum_{i=1}^{I} E(\lambda_{ij}|R_1=r_1)\cdot Var(\lambda_{ij}|R_1=r_1)\text{。} & \end{cases}$$

解以上的聯立方程式，可以得出觀察到第一筆資料後的估計參數為

$$\begin{cases} \alpha_+^{(1)} = \frac{\sum_{i=1}^{I} E(\theta_i|R_1=r_1)^2\cdot[1-E(\theta_i|R_1=r_1)]}{\sum_{i=1}^{I} E(\theta_i|R_1=r_1)\cdot Var(\theta_i|R_1=r_1)} - 1\text{，} & \\ \alpha_i^{(1)} = \alpha_+^{(1)}\cdot E(\theta_i|R_1=r_1)\text{，} & \text{for } i=1,\cdots,I\text{，} \end{cases} \tag{4.15}$$

且對所有i

$$\begin{cases} \beta_{i+}^{(1)} = \frac{\sum_{j=1}^{J} E(\lambda_{ij}|R_1=r_1)^2\cdot[1-E(\lambda_{ij}|R_1=r_1)]}{\sum_{i=1}^{I} E(\lambda_{ij}|R_1=r_1)\cdot Var(\lambda_{ij}|R_1=r_1)} - 1\text{，} & \\ \beta_{ij}^{(1)} = \beta_{i+}^{(1)}\cdot E(\lambda_{ij}|R_1=r_1)\text{，} & \text{for } j=1,2,\cdots,J\text{。} \end{cases} \tag{4.16}$$

因此(4.2)便可利用以下式子來估計

$$\hat{p}_1(\boldsymbol{\theta},\boldsymbol{\Lambda} \mid R_1=r_1) = \left[\frac{1}{B(\boldsymbol{\alpha}^{(1)})}\cdot\prod_{i=1}^{I}\theta_i^{\alpha_i^{(1)}-1}\right]\cdot\left[\prod_{i=1}^{I}\left(\frac{1}{B(\boldsymbol{\beta}_{i*}^{(1)})}\prod_{j=1}^{J}\lambda_{ij}^{\beta_{ij}^{(1)}-1}\right)\right]\text{。} \tag{4.17}$$

前式中$\boldsymbol{\alpha}^{(1)}$和$\boldsymbol{\beta}_{i*}^{(1)}$是接收第一筆資料之後所更新的估計參數，其中$\boldsymbol{\alpha}^{(1)}$和$\boldsymbol{\beta}_{i*}^{(1)}$中的元素分別由(4.15)和(4.16)式所算得。此時後驗分配可以(4.17)式的Dirichlet分配乘積來代替，並作為接收到第二筆資料前的近似先驗分配，同時$\boldsymbol{\theta}$與$\boldsymbol{\lambda}$的後驗近似均數(approximated posterior mean)為

$$\begin{aligned} \hat{\theta}_i^{(1)} &= \frac{\alpha_i^{(1)}}{\alpha_+^{(1)}} = \frac{\alpha_i^{(0)}}{\alpha_+^{(0)}+1} + \left(\frac{A_i^{(0)}}{A_+^{(0)}}\right)\cdot\frac{1}{\alpha_+^{(0)}+1}\text{，} \\ \hat{\lambda}_{ij}^{(1)} &= \frac{\beta_{ij}^{(1)}}{\beta_{i+}^{(1)}} = \sum_{m=1}^{I}\frac{A_m^{(0)}}{A_+^{(0)}}\cdot\frac{\beta_{ij}^{(0)}+\delta_{ij}^{mr_1}}{\beta_{i+}^{(0)}+\delta_{ir_1}^{mr_1}}\text{。} \end{aligned}$$

同理，以 (4.17) 式爲近似先驗分配，當接收到第二筆資料 $R_2 = r_2$ 後，後驗機率密度函數便可利用以下式子來估計

$$\hat{p}_2(\boldsymbol{\theta}, \boldsymbol{\Lambda} \mid R_1 = r_1, R_2 = r_2) \;=\; \left[\frac{1}{B(\boldsymbol{\alpha}^{(2)})} \cdot \prod_{i=1}^{I} \theta_i^{\alpha_i^{(2)}-1}\right] \cdot \left[\prod_{i=1}^{I}\left(\frac{1}{B(\boldsymbol{\beta}_{i*}^{(2)})} \prod_{j=1}^{J} \lambda_{ij}^{\beta_{ij}^{(2)}-1}\right)\right] 。$$

前式中 $\boldsymbol{\alpha}^{(2)}$ 和 $\boldsymbol{\beta}_{i*}^{(2)}$ 是接收第二筆資料之後所更新的估計參數，其中

$$\begin{cases} \alpha_+^{(2)} = \frac{\sum_{i=1}^{I} E(\theta_i|R_2=r_2)^2 \cdot [1-E(\theta_i|R_2=r_2)]}{\sum_{i=1}^{I} E(\theta_i|R_2=r_2)\cdot Var(\theta_i|R_2=r_2)} - 1 \text{，} \\ \alpha_i^{(2)} = \alpha_+^{(2)} \cdot E(\theta_i|R_2 = r_2) \text{，} \qquad\qquad \text{for } i = 1, 2, \cdots, I \text{，} \end{cases}$$

且對所有 i

$$\begin{cases} \beta_{i+}^{(2)} = \frac{\sum_{j=1}^{J} E(\lambda_{ij}|R_2=r_2)^2 \cdot [1-E(\lambda_{ij}|R_2=r_2)]}{\sum_{i=1}^{I} E(\lambda_{ij}|R_2=r_2)\cdot Var(\lambda_{ij}|R_2=r_2)} - 1 \text{，} \\ \beta_{ij}^{(2)} = \beta_{i+}^{(2)} \cdot E(\lambda_{ij}|R_2 = r_2) \text{，} \qquad\qquad \text{for } j = 1, 2, \cdots, J \text{，} \end{cases}$$

而期望值及變異數係指對近似先驗分配 (4.17) 而得。

依此類推，當接收到第 n 筆資料 $R_n = r_n$ 後，可得後驗機率分配爲

$$\hat{p}_n(\boldsymbol{\theta}, \boldsymbol{\Lambda} \mid R_1 = r_1, R_2 = r_2, \cdots, R_n = r_n) \;=\; \left[\frac{1}{B(\boldsymbol{\alpha}^{(n)})} \cdot \prod_{i=1}^{I} \theta_i^{\alpha_i^{(n)}-1}\right] \cdot \left[\prod_{i=1}^{I}\left(\frac{1}{B(\boldsymbol{\beta}_{i*}^{(n)})} \prod_{j=1}^{J} \lambda_{ij}^{\beta_{ij}^{(n)}-1}\right)\right] 。$$

此時 $\boldsymbol{\theta}$ 與 $\boldsymbol{\lambda}$ 的後驗近似均數爲

$$\begin{aligned} \hat{\theta}_i^{(n)} &= \frac{\alpha_i^{(n)}}{\alpha_+^{(n)}} = \frac{\alpha_i^{(n-1)}}{\alpha_+^{(n-1)}+1} + \left(\frac{A_i^{(n-1)}}{A_+^{(n-1)}}\right)\frac{1}{\alpha_+^{(n-1)}+1} \text{，} \\ \hat{\lambda}_{ij}^{(n)} &= \frac{\beta_{ij}^{(n)}}{\beta_{i+}^{(n)}} = \sum_{m=1}^{I} \frac{A_m^{(n-1)}}{A_+^{(n-1)}} \cdot \frac{\beta_{ij}^{(n-1)} + \delta_{ij}^{mr_n}}{\beta_{i+}^{(n-1)} + \delta_{ir_n}^{mr_n}} 。 \end{aligned}$$

而此近似解即本文所謂利用平均變異數和的計算方法所求出的近似估計值。

4.2 平均變異數和的性質

定理 **4.1** 考慮有 I 個類別，J 種不同的試驗結果，及 n 個觀察值，假設先驗分配 $\boldsymbol{\theta} \sim D(\boldsymbol{\alpha}^{(0)})$，$\boldsymbol{\lambda}_{i*} \sim D(\boldsymbol{\beta}_{i*}^{(0)})$，$i = 1, 2, \ldots, I$ 是獨立的 *Dirichlet* 分配，其中 $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_I)$，

$\boldsymbol{\alpha}^{(0)} = (\alpha_1^{(0)}, \ldots, \alpha_I^{(0)})$，$\boldsymbol{\lambda}_{i*} = (\lambda_{i1}, \lambda_{i2}, \ldots, \lambda_{iJ})$，$\boldsymbol{\beta}_{i*}^{(0)} = (\beta_{i1}^{(0)}, \beta_{i2}^{(0)}, \ldots, \beta_{iJ}^{(0)})$。則當$[\beta_{ij}^{(0)}]$的每一列成比例，即$\beta_{i*}^{(0)} = t_i \beta_{1*}^{(0)}$，$t_i \in \mathbb{R}$，$i = 2, 3, \ldots, I$時，對任意的$n$個觀察值利用平均變異數和的方法經過$n$次疊代後的估計參數仍會和起始值一樣，即對所有$n \in \mathbb{N}$，$\boldsymbol{\alpha}^{(n)} = \boldsymbol{\alpha}^{(0)}$。

證明: 我們將使用數學歸納法來證明此式子。我們先證明$n = 1$式子成立。

爲了滿足以上的假設，我們令

$$[\beta_{ij}^{(0)}] = \begin{pmatrix} \beta_{11}^{(0)} & \beta_{12}^{(0)} & \beta_{13}^{(0)} & \cdots & \beta_{1J}^{(0)} \\ \beta_{21}^{(0)} & \beta_{22}^{(0)} & \beta_{23}^{(0)} & \cdots & \beta_{2J}^{(0)} \\ \vdots & \vdots & \vdots & & \vdots \\ \beta_{I1}^{(0)} & \beta_{I2}^{(0)} & \beta_{I3}^{(0)} & \cdots & \beta_{IJ}^{(0)} \end{pmatrix} = \begin{pmatrix} \beta_{11}^{(0)} & \beta_{12}^{(0)} & \beta_{13}^{(0)} & \cdots & \beta_{1J}^{(0)} \\ t_2\beta_{11}^{(0)} & t_2\beta_{12}^{(0)} & t_2\beta_{13}^{(0)} & \cdots & t_2\beta_{1J}^{(0)} \\ \vdots & \vdots & \vdots & & \vdots \\ t_I\beta_{11}^{(0)} & t_I\beta_{12}^{(0)} & t_I\beta_{13}^{(0)} & \cdots & t_I\beta_{1J}^{(0)} \end{pmatrix}。$$

所以此時$\beta_{i+}^{(0)} = t_i \beta_{1+}^{(0)}$，對所有$i = 2, 3, \ldots, I$。

根據(4.2)式，可算得

$$A_m^{(0)} = B(\boldsymbol{\alpha}^{(0)} + \boldsymbol{\delta}^m) \cdot \prod_{i=1}^{I} B(\boldsymbol{\beta}_{i*}^{(0)} + \boldsymbol{\delta}_{i*}^{mr}) = \frac{\alpha_m^{(0)} \cdot \beta_{mr}^{(0)}}{\alpha_+^{(0)} \cdot \beta_{m+}^{(0)}} \cdot B(\boldsymbol{\alpha}^{(0)}) \cdot \prod_{i=1}^{I} B(\boldsymbol{\beta}_{i*}^{(0)})。$$

因此，

$$\begin{aligned} \frac{A_i^{(0)}}{A_+^{(0)}} &= \frac{\frac{\alpha_i^{(0)}\beta_{ir}^{(0)}}{\alpha_+^{(0)}\beta_{i+}^{(0)}}}{\sum_{m=1}^{I} \frac{\alpha_m^{(0)}\beta_{mr}^{(0)}}{\alpha_+^{(0)}\beta_{m+}^{(0)}}} = \frac{\frac{\alpha_i^{(0)}\beta_{ir}^{(0)}}{\beta_{i+}^{(0)}}}{\frac{\alpha_1^{(0)}\beta_{1r}^{(0)}}{\beta_{1+}^{(0)}} + \frac{\alpha_2^{(0)}\beta_{2r}^{(0)}}{\beta_{2+}^{(0)}} + \cdots + \frac{\alpha_I^{(0)}\beta_{Ir}^{(0)}}{\beta_{I+}^{(0)}}} \\ &= \frac{\frac{t_i\alpha_i^{(0)}\beta_{1r}^{(0)}}{t_i\beta_{1+}^{(0)}}}{\frac{\alpha_1^{(0)}\beta_{1r}^{(0)}}{\beta_{1+}^{(0)}} + \frac{t_2\alpha_2^{(0)}\beta_{1r}^{(0)}}{t_2\beta_{1+}^{(0)}} + \cdots + \frac{t_I\alpha_I^{(0)}\beta_{1r}^{(0)}}{t_I\beta_{1+}^{(0)}}} = \frac{\alpha_i^{(0)}}{\alpha_1^{(0)} + \alpha_2^{(0)} + \cdots + \alpha_I^{(0)}} \\ &= \frac{\alpha_i^{(0)}}{\alpha_+^{(0)}}, \quad i = 1, 2, \ldots, I。 \end{aligned}$$

又依(4.3)式，可得$\boldsymbol{\theta}$的後驗均數爲

$$\begin{aligned}
E(\theta_i|R=r) &= \frac{\alpha_i^{(0)}}{\alpha_+^{(0)}+1}+\left(\frac{A_i^{(0)}}{A_+^{(0)}}\right)\cdot\frac{1}{\alpha_+^{(0)}+1} \\
&= \frac{\alpha_i^{(0)}}{\alpha_+^{(0)}+1}+\left(\frac{\alpha_i^{(0)}}{\alpha_+^{(0)}}\right)\cdot\frac{1}{\alpha_+^{(0)}+1} \\
&= \frac{\alpha_i^{(0)}}{\alpha_+^{(0)}+1}\cdot\left(1+\frac{1}{\alpha_+^{(0)}}\right) \\
&= \frac{\alpha_i^{(0)}}{\alpha_+^{(0)}}\text{，}\quad i=1,2,\ldots,I\text{。}
\end{aligned}$$

依(4.4)和(4.5)式，可得$\boldsymbol{\theta}$的後驗二階動差及後驗變異數分別爲

$$\begin{aligned}
E(\theta_i^2|R=r) &= \frac{\alpha_i^{(0)}(\alpha^{(0)}+1)}{(\alpha_+^{(0)}+1)(\alpha_+^{(0)}+2)}+\left(\frac{A_i^{(0)}}{A_+^{(0)}}\right)\cdot\frac{2(\alpha_i^{(0)}+1)}{(\alpha_+^{(0)}+1)(\alpha_+^{(0)}+2)} \\
&= \frac{\alpha_i^{(0)}+1}{(\alpha_+^{(0)}+1)(\alpha_+^{(0)}+2)}(\alpha_i^{(0)}+\frac{2\alpha_i^{(0)}}{\alpha_+^{(0)}}) \\
&= \frac{\alpha_i^{(0)}(\alpha_i^{(0)}+1)}{(\alpha_+^{(0)}+1)(\alpha_+^{(0)}+2)}(1+\frac{2}{\alpha_+^{(0)}}) \\
&= \frac{\alpha_i^{(0)}(\alpha_i^{(0)}+1)}{\alpha_+^{(0)}(\alpha_+^{(0)}+1)}\text{，}\quad i=1,2,\ldots,I\text{，}
\end{aligned}$$

及

$$Var(\theta_i|R=r) = \frac{\alpha_i^{(0)}(\alpha_i^{(0)}+1)}{\alpha_+^{(0)}(\alpha_+^{(0)}+1)}-(\frac{\alpha_i^{(0)}}{\alpha_+^{(0)}})^2=\frac{\alpha_i^{(0)}(\alpha_+^{(0)}-\alpha_i^{(0)})}{(\alpha_+^{(0)})^2(\alpha_+^{(0)}+1)}\text{，}\quad i=1,2,\ldots,I\text{。}$$

因此可得$\boldsymbol{\theta}$的平均變異數和爲

$$\sum_{i=1}^{I}E(\theta_i|R=r)Var(\theta_i|R=r) = \sum_{i=1}^{I}\frac{(\alpha_i^{(0)})^2(\alpha_+^{(0)}-\alpha_i^{(0)})}{(\alpha_+^{(0)})^3(\alpha_+^{(0)}+1)}\text{。}$$

則由(4.15)式，當接收到資料r之後的估計參數爲

$$\begin{aligned}
\alpha_+^{(1)} &= \frac{\sum_{i=1}^{I}(\frac{\alpha_i^{(0)}}{\alpha_+^{(0)}})^2 \cdot [1-(\frac{\alpha_i^{(0)}}{\alpha_+^{(0)}})]}{\sum_{i=1}^{I}\frac{(\alpha_i^{(0)})^2(\alpha_+^{(0)}-\alpha_i^{(0)})}{(\alpha_+^{(0)})^3(\alpha_+^{(0)}+1)}} - 1 \\
&= \frac{\frac{1}{(\alpha_+^{(0)})^3}\sum_{i=1}^{I}(\alpha_i^{(0)})^2(\alpha_+^{(0)}-\alpha_i^{(0)})}{\frac{1}{(\alpha_+^{(0)})^3(\alpha_+^{(0)}+1)}\sum_{i=1}^{I}(\alpha_i^{(0)})^2(\alpha_+^{(0)}-\alpha_i^{(0)})} - 1 \\
&= \alpha_+^{(0)}\text{，}
\end{aligned}$$

及

$$\alpha_i^{(1)} = \alpha_+^{(1)} \cdot E(\theta_i|R=r) = \alpha_+^{(0)} \cdot \frac{\alpha_i^{(0)}}{\alpha_+^{(0)}} = \alpha_i^{(0)}\text{，}\quad i=1,2,\ldots,I\text{。}$$

令$n=k$時成立，即$\alpha_i^{(k)}=\alpha_i^{(0)}$，　$i=1,2,\ldots,I$。

同$n=1$時的證明我們可以知道

$$\alpha_i^{(k+1)} = \alpha_i^{(0)}\text{，}\quad i=1,2,\ldots,I\text{。}$$

因此根據數學歸納法我們可以證得，對任意的n個觀察值利用AVS的方法經過n次疊代後的估計參數仍會和起始值一樣，即對所有$n\in\mathbb{N}$，$\boldsymbol{\alpha}^{(n)}=\boldsymbol{\alpha}^{(0)}$。 ■

所以當先驗分配的參數如上述的型式時要特別注意，此時若用AVS的方法去估計真實後驗分配的參數，疊代的時候將不會受到觀察值的影響，估計參數$\boldsymbol{\alpha}^{(n)}$不會被更新，也就是我們的估計值完全由先驗分配的參數$\boldsymbol{\alpha}^{(0)}$來決定。

系理**4.2** 在滿足定理*4.1*的條件下，如果進一步假設$\alpha_1^{(0)}=\cdots=\alpha_I^{(0)}$時，則對任意的$n$個觀察值利用平均變異數和的方法經過$n$次疊代後計算出任意$i$個類別的後驗均數都為$1/I$。

由於Jiang and Dickey(2006)證明出類似的情況下用準貝氏法(q-B)計算出的後驗近似均數也會等於$1/I$，且等於用貝氏法計算出來的結果，所以在上述的情況下我們亦可以用平均變異數和(AVS)來代替貝氏法。