行政院國家科學委員會補助專題研究計畫成果報告

※※※※※※※※※※※※※※※※※※※※※
※　　　　　　　　　　　　　　　　　　　※
※　過濾變量 Dirichlet 分配的研究　※
※　**A study on the filtered-variate Dirichlet**　※
※　**distributions**　※
※　　　　　　　　　　　　　　　　　　　※
※※※※※※※※※※※※※※※※※※※※※

計畫類別：☑個別型計畫　　□整合型計畫
計畫編號：NSC 89-2118-M-004-011
NSC 90-2118-M-004-014
執行期間：89 年 8 月 1 日至 91 年 7 月 31 日

計畫主持人：姜志銘
計畫參與人員：郭錕霖
楊勝斌
陳力揚
柯力文

本成果報告包括以下應繳交之附件：

□赴國外出差或研習心得報告一份
□赴大陸地區出差或研習心得報告一份
□出席國際學術會議心得報告及發表之論文各一份
□國際合作研究計畫國外研究報告書一份

執行單位：國立政治大學應用數學系

中華民國九十一年十月二十九日

行政院國家科學委員會專題研究計畫成果報告

過濾變量 Dirichlet 分配的研究

A study on the filtered-variate Dirichlet distributions

計畫編號：NSC 89-2118-M-004-011
NSC 90-2118-M-004-014
執行期限：89年8月1日至91年7月31日

主持人：姜志銘　國立政治大學應用數學系
jiangt@math.nccu.edu.tw
計畫參與人員：郭錕霖　楊勝斌　陳力揚　柯力文
國立政治大學應用數學系

一、中文摘要

過濾變量 Dirichlet 先驗分配已經被廣泛地解決應用於已困擾貝氏統計學者幾十年的平滑直方圖的問題（見 Dickey 及主持人，1998），因此這些分配的評估在貝氏應用上就顯得很重要，本文中，我們將發展並提供對這些分配的簡易評估方法。同時，對應於直方圖的隨機向量的粗糙性常跟過濾（廣義）Dirichlet 分配有關，我們也將研究這粗糙性的分配，並給予簡易的方法來計算它的期望值。

關鍵詞：貝氏推論，粗糙性，過濾變量（廣義）Dirichlet，先驗評估

Abstract

The filtered-variate Dirichlet distributions prior have been extensively used for histogram smoothing problems, which had embarrassed Bayesians for decades (see Dickey and Jiang, 1998). Therefore, the assessments of these distributions are very important in Bayesian applications. In this paper, we develop and give easy assessment methods for these distributions. In addition, the roughness of a random vector corresponding to histogram probabilities is usually related to the filtered-variate (generalized) Dirichlet distributions. We study the distribution of the roughness and give easy methods to compute its expectation.

Keywords: Bayesian inference, roughness, filtered-variate (generalized) Dirichlet, prior assessment

§1 Introduction

It is well known among Bayesian statisticians that the class of Dirichlet distributions is a natural conjugate prior family for the multinomial sampling. However, in practice, it is likely to encounter sampling probabilities with adjacent categories close in value. Hence, it is highly desirable to have "smooth" prior distributions in such cases. Unfortunately, the Dirichlet distribution have the property where corresponding adjacent random quantities have negative correlations, which are in violation to the smoothness assumption. Dickey and Jiang (1998) successfully resolve this decades' problem by replacing Dirichlet distribution with the filtered-variate Dirichlet distribution.

Although there are many methods of the assessment of the filtered-variate Dirichlet distribution available and they seem to be promising, some non-statistician users (experts) may have difficulty in providing required information for prior assessments. For example, providing some typical smooth probability vectors so that the empirical moments of these vectors having prior means and variances may not be an easy task for some users (experts). In Section 2, we propose alternative assessment methods that are easy to elicit prior smoothing information.

Achieving an accurate estimation is usually more important than portraying the true smoothness. This is why most of researchers focus on the study of the posterior mean (or the mode) when the posterior distribution must be summarized in the form of a vector estimate. However, there may be occasions when it is reasonable to quote an estimate that exhibits a roughness equal to the posterior expected roughness. In Section 3, we shall also investigate this interesting problem of the roughness. Finally, we give conclusions in Section 4.

§2 Prior assessment

Let $v_1, v_2, \ldots, v_I$ be the unknown cell properties for multinomial sampling. It is well known that the corresponding natural conjugate family is the Dirichlet distribution. The random quantities v_i's are nearly prior independent, with a slight negative association, because of the constraint on their sum. However, it is frequently the case in practice that probabilities corresponding to adjacent cells are subjectively positively correlated, that is, the cell properties are expected to be smooth.

Here, we shall perform a linear transformation of a Dirichlet vector so that the new probability vector is smooth. Let's consider the filtered-variate Dirichlet vector $\boldsymbol{v} \sim F_A D(\boldsymbol{b})$, where $\boldsymbol{b} = (b_1, b_2, \ldots, b_I)'$ and $I \times I$ circular matrix A with the first $k(k < I)$ entries having the value $1/k$ and the remaining entries having the value 0 for the first row, with the second to the k plus first entries having the value $1/k$ and the remaining entries having the value 0 for the second row, and so on.

The major issue now is then the prior assessment. An assessment procedure for the prior distribution, setting $b_+ = \sum_{i=1}^{I} b_i$ and $\boldsymbol{e} = kA\boldsymbol{b}$, can be as follows:

Step 1. Elicit prior mean for the i-th categorical probability, say m_i. We would have

$$m_i = \frac{1}{k} \cdot \frac{e_i}{b_+}, \qquad i = 1, \ldots, I.$$

Step 2. Elicit prior variance for the first categorical probability, say s_1^2. We then have

$$b_+ = \frac{m_1(1 - m_1 k)}{k \cdot s_1^2} - 1.$$

Step 3. Compute e_i,

$$e_i = k \cdot b_+ \cdot m_i, \qquad i = 1, 2, \ldots, I.$$

Step 4. Compute $\boldsymbol{b}$,

$$\boldsymbol{b} = \frac{1}{k} \cdot A^{-1} \cdot \boldsymbol{e}.$$

Note that an inverse matrix of a circular matrix A can be computed easily (see Marcus and Minc (1964)). Hence, the prior distribution can be easily assessed.

§3 Roughness

Define the r-roughness of a vector v by the average of its squared r-distant difference,

$$R_r(\boldsymbol{v}) = \sum_{i=1}^{I-r} \frac{(v_i - v_{i+r})^2}{I - r},$$

with special interest in adjacent differences, $r = 1$. That is,

$$R_1(\boldsymbol{v}) = \sum_{i=1}^{I-1} \frac{(v_i - v_{i+1})^2}{I - 1}.$$

If we further define $\boldsymbol{d} = (d_1, d_2, \ldots, d_I)'$, where $d_i = v_i - v_{i+1}$, for $i = 1, 2, \ldots, I - 1$, and $d_I = v_I - v_1$, then $R_1(\boldsymbol{v}) = \sum_{i=1}^{I-1} d_i^2/(I - 1)$. It can be shown that $\boldsymbol{d} = B\boldsymbol{v}$, where B is an $I \times I$ matrix and

$$B = \begin{bmatrix} 1 & -1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & -1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & -1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 & -1 \\ -1 & 0 & 0 & 0 & \cdots & 0 & 1 \end{bmatrix}.$$

Consider the filtered-variate generalized Dirichlet distribution $\boldsymbol{v} \sim F_A D(\boldsymbol{b}, H, \boldsymbol{c})$, where $\boldsymbol{v} = A\boldsymbol{u}$, A is an $I \times I$ matrix, $\boldsymbol{u}$ is an $I \times 1$ random vector, and $\boldsymbol{u} \sim D(\boldsymbol{b}, H, \boldsymbol{c})$. Note that such generalized Dirichlet distribution $\boldsymbol{u}$ was first defined by Dickey (1983). In addition, $D(\boldsymbol{b}, H, \boldsymbol{c}) \sim D(\boldsymbol{b})$ if $H = 0$ or $\boldsymbol{c} = \boldsymbol{0}$. Since $\boldsymbol{d} = B\boldsymbol{v}$, hence $\boldsymbol{d} = BA\boldsymbol{u}$. Note that $\boldsymbol{d}$ is not a probability vector and that $\boldsymbol{d}$ is not a regular filtered-variate generalized Dirichlet distribution. However, the moments of $\boldsymbol{d}$ can be expressed similarly as those of the regular filtered-variate generalized Dirichlet distribution.

Before studying the distribution of R_1, we shall reexpress R_1 first. Let D be a submatrix of B by removing the last row of B. Hence, D is an $I-1$ by I matrix. R_1 is then expressed as $[1/(I-1)](\boldsymbol{u}'A'D'DA\boldsymbol{u})$. Since $A'D'$ is the transpose matrix of DA, hence $A'D'DA$, denoted by $C(I\times I)$, is a symmetric matrix. Therefore,

$$R_1 = \frac{1}{I-1}(\boldsymbol{u}'C\boldsymbol{u}),$$

where $\boldsymbol{u}$ is an $I\times 1$ random vector, C is a $I\times I$ symmetric constant matrix. Now, by Theorem 1.7 of Seber (1977, p. 13), we have

$$E(R_1) = \frac{1}{I-1}\left[\operatorname{tr}(C\Sigma) + E(\boldsymbol{u})'CE(\boldsymbol{u})\right], \quad (1)$$

where Σ is the variance-covariance matrix of random vector $\boldsymbol{u}$. Equation (1) gives us the expected roughness.

To apply equation (1) in detail, we shall consider the special case when $I=5$ and $\boldsymbol{u}\sim D(\boldsymbol{b})$. If A is a 5×5 matrix and $k=4$, then, $\boldsymbol{d}=BA\boldsymbol{u}$, where $\boldsymbol{u}\sim D(\boldsymbol{b})$ and B is a 5×5 matrix. Since D is a submatrix of B by removing the last row of matrix B, $D = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 & -1 \end{bmatrix}$.

It can be shown that

$$C = A'D'DA = \frac{1}{16}\begin{bmatrix} 2 & -1 & 0 & 0 & -1 \\ -1 & 2 & -1 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & -1 & 1 & 0 \\ -1 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Hence, $R_1=(1/4)(\boldsymbol{u}'C\boldsymbol{u})$. If we define $G=16C$, then the expected 1-roughness $E(R_1)$ can be computed by equation (1), that is

$$E(R_1) = \frac{1}{4^3}\left[\operatorname{tr}(G\Sigma) + E(\boldsymbol{u})'GE(\boldsymbol{u})\right],$$

where Σ is the variance-covariance matrix of $\boldsymbol{u}$. It can be shown that

$$\operatorname{tr}(G\Sigma) = \frac{1}{b_+^2(b_++1)}(6b_1b_2 + 4b_1b_3 + 3b_1b_4 + 5b_1b_5 + 6b_2b_3 + 3b_2b_4 + 3b_2b_5 + 5b_3b_4 + 3b_3b_5 + 2b_4b_5).$$

In addition,

$$E(\boldsymbol{u})'GE(\boldsymbol{u}) = \frac{1}{b_+^2}(2b_1^2 + 2b_2^2 + 2b_3^2 + b_4^2 + b_5^2 - 2b_1b_2 - 2b_2b_3 - 2b_3b_4).$$

Therefore, we have

$$E(R_1) = \frac{1}{4^3 b_+(b_++1)}\left\{(2b_1 + 2b_2 + 2b_3 + b_4 + b_5) + (2b_1^2 + 2b_2^2 + 2b_3^2 + b_4^2 + b_5^2 - 2b_1b_2 - 2b_1b_5 - 2b_2b_3 - 2b_3b_4)\right\}.$$

If $\boldsymbol{u}$ is a generalized Dirichlet distribution, (i.e., $\boldsymbol{u}\sim D(\boldsymbol{b},H,\boldsymbol{c})$), then the probability density function of $\boldsymbol{u}$ is

$$f(\boldsymbol{u}) = \frac{\frac{1}{B(\boldsymbol{b})}\left(\prod_{i=1}^{I} u_i^{b_i-1}\right)\left[\prod_{j=1}^{J}\left(\sum_{i=1}^{I} u_i h_{ij}\right)^{c_j}\right]}{\mathcal{R}(\boldsymbol{b},H,-\boldsymbol{c})}, \quad (2)$$

where the normalized constant $\mathcal{R}$ is the Carlson's (see Carlson (1977)) multiple hypergeometric function. It can be seen that the moments of $\boldsymbol{u}$ are the ratios of $\mathcal{R}$'s. These ratios can be computed by the methods given by Jiang, Kadane, and Dickey (1992). We can also approximate moments by the quasi-Bayes method. Note the numerator of the right-hand side of equation (2) can be regarded as the product of a Dirichlet prior probability density function and a likelihood of the censored data $\left[\prod_{j=1}^{J}\left(\sum_{i=1}^{I} u_i h_{ij}\right)^{c_j}\right]$, where c_j is the number of observations reported as the j-th report set. The total number of observations is then $n\equiv c_+=\sum_{j=1}^{J} c_j$. For convenience, we shall assume that these data are received sequentially. Therefore, the posterior probability density function, after receipt of the first report $R_1=r_1$, has the form,

$$f(\boldsymbol{u}) = \sum_{m=1}^{I}\frac{B_m^{(1)}h_{mr_1}}{B_+^{(1)}}\left\{\left(\prod_{i=1}^{I} u_i^{b_i+\delta_i^m-1}\right)\Big/B_m^{(1)}\right\}, \quad (3)$$

where $B_m^{(1)}=B(\boldsymbol{b}+\boldsymbol{\delta}^m)$, $\boldsymbol{\delta}^m=(\delta_1^m,\ldots,\delta_I^m)$, and $B_+^{(1)}=\sum_{m=1}^{I}B_m^{(1)}h_{mr_1}$. If we were further informed the true category of the first subject, then the posterior density (3) would become

$$\frac{1}{B(\boldsymbol{b}+\boldsymbol{c}^{(1)})}\left(\prod_{i=1}^{I} u_i^{b_i+c_i^{(1)}-1}\right), \qquad (4)$$

where $\boldsymbol{c}^{(1)} = (c_1^{(1)}, c_2^{(1)}, \ldots, c_I^{(1)})$ and each $c_i^{(1)}$ is 1 if the true category of the first subject is i, and is 0 otherwise. However, we are not further informed the true category of the first subject. Instead, we base our decision upon the quasi-datum $d_i^{(1)} = E[C_i^{(1)} \mid R_1 = r_1] = \Pr(C_i^{(1)} = 1 \mid r_1)$, the expected value of $C_i^{(1)}$ posterior to the datum r_1. Now, we have

$$d_i^{(1)} = \frac{\Pr(R_1 = r_1 \mid C_i^{(1)} = 1)\Pr(C_i^{(1)} = 1)}{\sum_{k=1}^{I}[\Pr(R_1 = r_1 \mid C_k^{(1)} = 1)\Pr(C_k^{(1)} = 1)]}.$$

Let

$$\hat{d}_i^{(1)} = \frac{h_{ir_1}\hat{u}_i^{(0)}}{\sum_{k=1}^{I} h_{kr_1}\hat{u}_k^{(0)}},$$

where $\hat{u}_i^{(0)}$ is the prior mean of u_i. We use $\hat{d}_i^{(1)}$ to estimate $d_i^{(1)}$. This provides the key to the procedure. We approximate (3) by the Dirichlet density (4),

$$\begin{aligned}\hat{f}_1(\boldsymbol{u}) &= \frac{1}{B(\boldsymbol{b}+\hat{\boldsymbol{d}}^{(1)})}\prod_{i=1}^{I} u_i^{b_i+\hat{d}_i^{(1)}-1} \\ &= \frac{1}{B(\boldsymbol{b}^{(1)})}\prod_{i=1}^{I} u_i^{b_i^{(1)}-1} \qquad (5)\end{aligned}$$

where $\boldsymbol{b}^{(1)}$ is the updated parameter vector at the first step. That is, $\boldsymbol{b}_i^{(1)} = b_i + \hat{d}_i^{(1)}$ for all i. This approximate posterior distribution is within the Dirichlet distribution family. Subsequent updating proceeds in the identical manner. For the n-th step, after receiving the n-th report $R_n = r_n$, our approximate posterior distribution of $\boldsymbol{u}$ is

$$\begin{aligned}&\hat{f}_n(\boldsymbol{u} \mid R_1 = r_1, \ldots, R_n = r_n) \\ &= \frac{1}{B(\boldsymbol{b}^{(n-1)}+\hat{\boldsymbol{d}}^{(n)})}\prod_{i=1}^{I} u_i^{b_i^{(n-1)}+\hat{d}_i^{(n)}-1},\end{aligned}$$

where $\boldsymbol{b}^{(n-1)}$ is the updated parameter vector at the $(n-1)^{\text{st}}$ step. The approximate general posterior moment of u_i's is

$$\hat{E}\left(\prod_{i=1}^{I} u_i^{b_i'}\right) = B(\boldsymbol{b}^{(n)}+\boldsymbol{b}')\big/B(\boldsymbol{b}^{(n)}).$$

Hence, for example, the posterior mean of θ_i, based on r_1, r_2, $\ldots$, r_n, can be approximated by

$$\hat{u}_i^{(n)} = \frac{b_i^{(n)}}{b_+^{(n)}} = \frac{b_i^{(n-1)} + \widehat{\Pr}(C_i^{(n)} = 1 \mid R_n = r_n)}{b_+^{(0)} + n}.$$

§4 Conclusions

The filtered-variate Dirichlet distribution family is important for the problems of the Bayesian local smoothness. We now provide a new prior assessment method for the filtered-variate Dirichlet distribution. It is easy enough for even non-statisticians to apply. Although we are usually more interested in an accurate estimation than the true smoothness, there are occasions that the roughness is a concern. We also provide an easy method to compute the expected value of roughness for the filtered-variate (generalized) Dirichlet distribution.

References

[1] B. C. Carlson (1977), *Special Functions of Applied Mathematics*, Academic Press, New York.

[2] J. M. Dickey (1983), "Multiple hypergeometric functions: Probabilistic interpretations and statistical uses," Journal of the American Statistical Association, 78, 628–637.

[3] Dickey, J. M. and Jiang, T. J. (1998), "Filtered-Variate Prior Distributions for Histogram Smoothing," Journal of the American Statistical Association, 93, 651–662.

[4] T. J. Jiang, J. B. Kadane, and J. M. Dickey (1992), "Computation of Carlson's multiple hypergeometric function $\mathcal{R}$ for Bayesian applications," Journal of Computational and Graphical Statistics, 1, 231–251.

[5] Marcus, Marvin and Minc, Henryk (1964), *A Survey of Matrix Theory and Matrix Inequalities*, Allyn and Bacon, Boston.

[6] Seber, G. A. F. (1977),*Linear Regression Analysis*, Wiley, New York.