行政院國家科學委員會補助專題研究計畫 ■成果報告 □期中進度報告

一個新的多變量特徵函數的貝氏應用

Bayesian application on a new multivariate characteristic function

計畫類別：■個別型計畫　□整合型計畫
計畫編號：NSC 92-2118-M-004-001
執行期間：92年8月1日至93年7月31日

計畫主持人：姜志銘
計畫參與人員：郭錕霖　林其緯　羅文宜　陳妙津

成果報告類型(依經費核定清單規定繳交)：■精簡報告　□完整報告

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

處理方式：除產學合作研究計畫、提升產業技術及人才培育研究計畫、列管計畫及下列情形者外，得立即公開查詢
□涉及專利或其他智慧財產權，□一年■二年後可公開查詢

執行單位：國立政治大學應用數學系

中　華　民　國　九　十　三　年　十　月　二　十　六　日

行政院國家科學委員會專題研究計畫成果報告

一個新的多變量特徵函數的貝氏應用
Bayesian application on a new multivariate characteristic function

計畫編號：NSC 92-2118-M-004-001
執行期限：92年8月1日至93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　林其緯　羅文宜　陳妙津
國立政治大學應用數學系

中文摘要

對很多不容易使用傳統特徵函數的分配，Jiang, Dickey, and Kuo (2004) 提供這些分配一種非常有用的新的多變量特徵函數，其中如在 Dicky and Jiang (1998) 所提出且被廣泛應用於平滑直方圖問題的過濾變量 Dirichlet 分配，可被證明有這種新的多變量特徵函數的完整式子。本研究中，我們探討及發展利用這種新的多變量特徵函數來評估在平滑直方圖問題相當有用的過濾變量 Dirichlet 先驗分配的簡易方法。

關鍵詞：貝氏推論，過濾變量的 Dirichlet 分配，貝氏區域平滑性，準貝氏計算方法

Abstract

Jiang, Dickey, and Kuo (2004) give a new multivariate characteristic function which is very useful for many distributions that are difficult to use traditional characteristic function. The filtered-variate Dirichlrt distribution, which has been extensively used for histogram smoothing problems (Dickey and Jiang, 1998), can be shown to have closed form for this new characteristic function. In this research, we study and develop easy methods to use this new multivariate characteristic function to assess the filtered-variate Dirichlet prior, which is very useful for histogram smoothing problems.

Keywords: Bayesian inference, filtered-variate Dirichlet distribution, Bayesian local smoothness, quasi-Bayesian computational method

1 Introduction

How can a joint prior distribution be chosen to give high prior probability to the event that the sampling probabilities are "smooth," that neighboring categories have probabilities close in value? In its extreme form, it is the problem of Bayesian nonparametric inference, a major embarrassment to Bayesian (L. J. Savage, who is a Bayesian pioneer). Hence, it is very important to study the prior smoothness. Dickey and Jiang (1998) give a pioneered work in such area and introduced the filtered-variate Dirichlet distribution as prior.

Jiang, Dickey, and Kuo (2004) give a new multivariate c-characteristic function and show its properties, which are very important in those cases when the traditional characteristic functions are difficulty to apply. In this paper, we shall give further properties of the multivariate c-characteristic function in Section 2. With these results, we give in Section 3 a prior assessment method. Conclusions are given in Section 4.

2 Properties of the multivariate c-characteristic function

In this section, we shall give some further properties of the multivariate c-characteristic function which is first given by Jiang, Dickey, and Kuo (2004). We shall restate its definition as follows:

Definition 2.1 *If $\boldsymbol{u} = (u_1, u_2, \dots, u_L)'$ is a random vector on a subset S of $A = [-a_1, a_1] \times [-a_2, a_2] \times \cdots \times [-a_L, a_L]$, its multivariate c-characteristic function is defined as*

$$g(\boldsymbol{t}; \boldsymbol{u}, c) = \mathop{\mathrm{E}}_{\boldsymbol{u}}\left[(1 - i\boldsymbol{t} \cdot \boldsymbol{u})^{-c}\right], \; |\boldsymbol{t}| < a^{-1}, \tag{2.1}$$

where c is a positive real number, $a = \sqrt{\sum_{i=1}^{L} a_i^2}$, $\boldsymbol{t}' = (t_1, t_2, \dots, t_L)$, $|\boldsymbol{t}| = \sqrt{\sum_{i=1}^{L} t_i^2}$, and $\boldsymbol{t} \cdot \boldsymbol{u}$ is the inner product of two vectors (i.e., $\boldsymbol{t} \cdot \boldsymbol{u} = \sum_{i=1}^{L} t_i u_i$).

The following theorem can be proved by Corollary 2.6 of Jiang, Dickey, and Kuo (2004) and the fact that $g(\boldsymbol{0}; \boldsymbol{u}, c) = 1$.

Theorem 2.2 *Let $\boldsymbol{u}$ be an $L \times 1$ random vector in a subset S of $A = [-a_1, a_1] \times [-a_2, a_2] \times \cdots \times [-a_L, a_L]$, and its corresponding multivariate c-characteristic function is $g(\boldsymbol{t}; \boldsymbol{u}, c)$ as defined in Definition 2.1. Then, we have*

$$\mathrm{E}(u_j) = \frac{1}{ci} \left.\frac{\partial \ln g(\boldsymbol{t}; \boldsymbol{u}, c)}{\partial t_j}\right|_{\boldsymbol{t}=\boldsymbol{0}}, \; \textit{for } j = 1, 2, \dots, L, \tag{2.2a}$$

$$\mathrm{Cov}(u_j, u_k) = -\frac{1}{c(c+1)}\left[\left.\frac{\partial^2 \ln g(\boldsymbol{t}; \boldsymbol{u}, c)}{\partial t_j \partial t_k}\right|_{\boldsymbol{t}=\boldsymbol{0}} - \frac{1}{c}\left(\left.\frac{\partial \ln g(\boldsymbol{t}; \boldsymbol{u}, c)}{\partial t_j}\right|_{\boldsymbol{t}=\boldsymbol{0}}\right)\left(\left.\frac{\partial \ln g(\boldsymbol{t}; \boldsymbol{u}, c)}{\partial t_k}\right|_{\boldsymbol{t}=\boldsymbol{0}}\right)\right]. \tag{2.2b}$$

Dickey and Jiang (1998) defined a random probability vector $\boldsymbol{v} = (v_1, v_2, \dots, v_M)'$, where $\sum_{i=1}^{M} v_i = 1$, to be filtered-variate Dirichlet distributed, $\boldsymbol{v} \sim F_G D(\boldsymbol{b})$, with $\boldsymbol{v} = G\boldsymbol{u}$ for a constant matrix $G_{M \times L}$ and an $L \times 1$ parameter vector $\boldsymbol{b}$, where the $L \times 1$ random vector $\boldsymbol{u}$ has a Dirichlet distribution with parameter $\boldsymbol{b}$. Jiang, Dickey, and Kuo (2004) give the multivariate c-characteristic function of the filtered-variate distribution $F_G D(\boldsymbol{b})$, which is

$$g(\boldsymbol{s}; \boldsymbol{v}, c) = \prod_{j=1}^{L}\left[1 - i\left(\sum_{m=1}^{M} g_{mj} s_m\right)\right]^{-b_j}, \tag{2.3}$$

where $c = \sum_{j=1}^{L} b_j$.

With the applications of Theorem 2.2, the first two central moments of the filtered-variate Dirichlet distribution can now be expressed in the following theorem.

Theorem 2.3 *Let $\boldsymbol{v} = (v_1, v_2, \ldots, v_M)' \sim F_G D(\boldsymbol{b})$, where G is an $M \times L$ matrix and $\boldsymbol{b} = (b_1, b_2, \ldots, b_L)'$ is an $L \times 1$ random vector. That is, $\boldsymbol{v} = G\boldsymbol{u}$, where $\boldsymbol{u} \sim D(\boldsymbol{b})$. Then, we have*

$$\mathrm{E}(v_l) = \frac{1}{c}\sum_{j=1}^{L} b_j g_{lj}, \ \forall l = 1, 2, \ldots, M, \tag{2.4a}$$

$$\mathrm{Cov}(v_l, v_{l'}) = \frac{1}{c(c+1)}\left[\sum_{j=1}^{L} b_j g_{lj} g_{l'j} - \frac{1}{c}\left(\sum_{j=1}^{L} b_j g_{lj}\right)\left(\sum_{j=1}^{L} b_j g_{l'j}\right)\right], \ \forall l, l' = 1, 2, \ldots, M, \tag{2.4b}$$

where $c = \sum_{j=1}^{L} b_j$.

Proof. From (2.3), the multivariate c-characteristic function of $\boldsymbol{v}$ is

$$g(\boldsymbol{s}; \boldsymbol{v}, c) = \prod_{j=1}^{L}\left[1 - i\left(\sum_{m=1}^{M} g_{mj} s_m\right)\right]^{-b_j}.$$

Hence,

$$\ln g(\boldsymbol{s}; \boldsymbol{v}, c) = -\sum_{j=1}^{L} b_j \ln\left[1 - i\left(\sum_{m=1}^{M} g_{mj} s_m\right)\right].$$

By taking the first two partial derivatives, we have

$$\frac{\partial \ln g(\boldsymbol{s}; \boldsymbol{v}, c)}{\partial s_l} = \sum_{j=1}^{L} \frac{i b_j g_{lj}}{1 - i\left(\sum_{m=1}^{M} g_{mj} s_m\right)},$$

and

$$\frac{\partial^2 \ln g(\boldsymbol{s}; \boldsymbol{v}, c)}{\partial s_l \partial s_{l'}} = -\sum_{j=1}^{L} \frac{b_j g_{lj} g_{l'j}}{\left[1 - i\left(\sum_{m=1}^{M} g_{mj} s_m\right)\right]^2}.$$

Therefore,

$$\left.\frac{\partial \ln g(\boldsymbol{s}; \boldsymbol{v}, c)}{\partial s_l}\right|_{\boldsymbol{s}=\boldsymbol{0}} = i\sum_{j=1}^{L} b_j g_{lj},$$

and

$$\left.\frac{\partial^2 \ln g(\boldsymbol{s}; \boldsymbol{v}, c)}{\partial s_l \partial s_{l'}}\right|_{\boldsymbol{s}=\boldsymbol{0}} = -\sum_{j=1}^{L} b_j g_{lj} g_{l'j}.$$

By equations (2.2a) and (2.2b), the results of equations (2.4a) and (2.4b) can then be established. ❑

The equation (2.4a) can be expressed in terms of matrix expression

$$\mathrm{E}(\boldsymbol{v}) = \frac{1}{c} G\boldsymbol{b}. \tag{2.5a}$$

Similarly, the equation (2.4b) can be expressed in terms of matrix expression

$$\mathrm{Cov}(\boldsymbol{v}) = \frac{1}{c(c+1)} G D_b G' - \frac{1}{c^2(c+1)} G \boldsymbol{b}\boldsymbol{b}' G', \tag{2.5b}$$

where $D_b = \mathrm{diag}(b_1, b_2, \ldots, b_L)$.

3 Prior assessment

First, we shall give the relation between the first two moments of $\boldsymbol{v}$ and $\boldsymbol{u}$ in matrix notation. Note that $\mathrm{E}(\boldsymbol{u}) = \frac{1}{c}\boldsymbol{b}$, and $\mathrm{Cov}(\boldsymbol{u}) = \frac{1}{c(c+1)}D_b - \frac{1}{c^2(c+1)}\boldsymbol{b}\boldsymbol{b}'$. By the equations (2.5a) and (2.5b), it can be seen that $\mathrm{E}(\boldsymbol{v}) = G\mathrm{E}(\boldsymbol{u})$ and $\mathrm{Cov}(\boldsymbol{v}) = G\mathrm{Cov}(\boldsymbol{u})G'$.

Second, since $u_J = 1 - \sum_{j=1}^{J-1} u_j$ and $v_I = 1 - \sum_{i=1}^{I-1} v_i$, we shall consider the reduced random vectors $\boldsymbol{u}_r$ and $\boldsymbol{v}_r$, where $\boldsymbol{u}_r$ has only the first $J-1$ components of $\boldsymbol{u}$ and $\boldsymbol{v}_r$ is the first $I-1$ components of $\boldsymbol{v}$. It can be shown that $\mathrm{E}(\boldsymbol{u}_r) = \frac{1}{c}\boldsymbol{b}_r$ and $\mathrm{Cov}(\boldsymbol{u}_r) = \frac{1}{c^2(c+1)}\begin{bmatrix} b_1(c-b_1) & -b_1b_2 & \cdots & -b_1b_{J-1} \\ -b_1b_2 & b_2(c-b_2) & \cdots & -b_2b_{J-1} \\ \vdots & \vdots & \ddots & \vdots \\ -b_1b_{J-1} & -b_2b_{J-1} & \cdots & b_{J-1}(c-b_{J-1}) \end{bmatrix}$. Since $\mathrm{Cov}(\boldsymbol{u}_r)$ is a positive definite matrix, its square-root matrix, say F_r, by spectral decomposition, can be expressed as $\sum_{j=1}^{J-1}\sqrt{\lambda_j}\boldsymbol{e}_j\boldsymbol{e}_j'$, where $(\lambda_1, \boldsymbol{e}_1), (\lambda_2, \boldsymbol{e}_2), \ldots, (\lambda_{J-1}, \boldsymbol{e}_{J-1})$ are pairs of eigenvalues and eigenvectors of $\mathrm{Cov}(\boldsymbol{u}_r)$, and $\boldsymbol{e}_j'\boldsymbol{e}_j = 1$, $\boldsymbol{e}_j'\boldsymbol{e}_{j'} = 0$, for all $j, j' = 1, 2, \ldots, J-1$. For example, when $J = 3$, it can be shown that

$$\lambda_1 = \frac{b_1(b_2+b_3) + b_2(b_1+b_3) + \sqrt{4b_1^2b_2^2 + b_3^2(b_1-b_2)^2}}{2c^2(c+1)},$$

$$\boldsymbol{e}_1' = \frac{1}{d_1}\left(2b_1b_2, b_3(b_1-b_2) - \sqrt{4b_1^2b_2^2 + b_3^2(b_1-b_2)^2}\right),$$

and

$$\lambda_2 = \frac{b_1(b_2+b_3) + b_2(b_1+b_3) - \sqrt{4b_1^2b_2^2 + b_3^2(b_1-b_2)^2}}{2c^2(c+1)},$$

$$\boldsymbol{e}_2' = \frac{1}{d_1}\left(2b_1b_2, b_3(b_1-b_2) + \sqrt{4b_1^2b_2^2 + b_3^2(b_1-b_2)^2}\right)$$

where d_1 and d_2 are chosen so that $\boldsymbol{e}_1$ and $\boldsymbol{e}_2$ are normalized vectors.

Assume that an expert is familiar with means and variances/covariances of the underlying prior population. In particular, he/she is quite certain with his/her knowledge about the variances/covariances. With the assumption that the prior distribution family is the filtered-variate Dirichlet, i.e., $\boldsymbol{v} \sim F_G D(\boldsymbol{b})$, later we shall assess the prior parameters G and $\boldsymbol{b}$ based on the expert's prior mean and prior variance-covariance with a restriction on $\boldsymbol{b}$.

Now, we discuss how to assess the transformation matrix G in general. Assume that a positive definite prior covariance matrix $H_r = \mathrm{Cov}(\boldsymbol{v}_r)$ and prior mean vector $\boldsymbol{h}_r$ are given by an expert. As we did before, by spectral decomposition, H_r can be decomposed as, say E_rE_r'. Since we need to have $H_r = G_r\mathrm{Cov}(\boldsymbol{u}_r)G_r'$, we may define $G_r = E_rF_r^{-1}$, where $F_r^{-1} = \sum_{j=1}^{J-1}\frac{1}{\sqrt{\lambda_j}}\boldsymbol{e}_j\boldsymbol{e}_j'$. From the expert's prior means and prior variances/covariances, F_r is still not available yet. We shall make a restriction on $\boldsymbol{b}$ by assuming each component of b_i is the same in the following assessment procedure.

An assessment procedure for the prior distribution $\boldsymbol{v} \sim F_G D(\boldsymbol{b})$, where each component $b_i = b$ of $\boldsymbol{b}$ is restricted to be constant and $I = J$, can be as follows:

1. Elicit the prior covariance matrix of $\boldsymbol{v}_r$, say H_r.
2. Compute the square-root matrix of H_r, say E_r.
3. The covariance matrix of $\boldsymbol{u}_r$, when $b = 1$, is $\mathrm{Cov}(\boldsymbol{u}_r) = \frac{1}{I^2(I+1)} \begin{bmatrix} I-1 & -1 & \cdots & -1 \\ -1 & I-1 & \cdots & -1 \\ \vdots & \vdots & \ddots & \vdots \\ -1 & -1 & \cdots & I-1 \end{bmatrix}$.

   We then compute the square-root matrix of $\mathrm{Cov}(\boldsymbol{u}_r)$, say $F_r^{(1)}$.
4. Define $G_r^{(1)} = E_r F_r^{(1)}$, and $\boldsymbol{e}_r = \frac{1}{I} G_r^{(1)} \mathbf{1}$, where $\mathbf{1}$ is a vector with each component 1.
5. Elicit the prior mean vector of $\boldsymbol{v}_r$, say $\boldsymbol{d}_r$.
6. Compute $s = \frac{\sum_{i=1}^{I-1} s_i}{I-1}$, where the i-th component ratio is $s_i = \frac{d_i}{e_i}$, for all $i = 1, 2, \ldots, I-1$.
7. Let $G_r = s G_r^{(1)}$ and $\boldsymbol{b} = m\mathbf{1}$, where $m = s^2 + \frac{s^2-1}{I}$.

If we define the $I \times I$ matrix G so that the (i, j)-th element of G,

$$g_{ij} = \begin{cases} g_{ij}^{(r)}, & i \neq I \text{ and } j \neq I, \\ 1 - \sum_{i=1}^{I-1} g_{ij}^{(r)}, & i = I \text{ and } j \neq I, \\ 0, & i \neq I \text{ and } j = I, \\ 1, & i = I \text{ and } j = I, \end{cases} \quad \text{i.e., } G = \left[\begin{array}{ccc:c} & G_r & & 0 \\ \hdashline 1 - \sum_{i=1}^{I-1} g_{i1}^{(r)} & \cdots & 1 - \sum_{i=1}^{I-1} g_{i,I-1}^{(r)} & 1 \end{array} \right],$$

where $g_{ij}^{(r)}$ is the (i, j)-th element of G_r, we have the assessed prior as $\boldsymbol{v} \sim F_G D(\boldsymbol{b})$, where $\boldsymbol{b} = m\mathbf{1}$.

The assessed prior distribution $\boldsymbol{v}$ would have the same covariance matrix and the close mean vector as those given by the expert. Notice that the computations needed in the above procedure can be carried out easily by computer.

4 Conclusions

In this research, we give additional properties of the Jiang, Dickey and Kuo (2004)'s multivariate c-characteristic function. We also provide an alternative prior assessment procedure when we are using the filtered-variate Dirichlet prior distribution family. The computations of this procedure can be carried out easily.

References

Dickey, J. M. and Jiang, T. J. (1998), "Filtered-variate prior distributions for historgram smoothing," J. Amer. Statist. Assoc. 93, 651–662.

Jiang, T. J., Dickey, J. M., and Kuo, K.-L. (2004), "A new multivariate transform and the distribution of a random functional of a Ferguson-Dirichlet process," Stochastic Processes and their Applications, 111, 77–95.