行政院國家科學委員會專題研究計畫 成果報告

隨機函數分配
研究成果報告(精簡版)

計 畫 類 別 ： 個別型
計 畫 編 號 ： NSC 98-2118-M-004-002-
執 行 期 間 ： 98 年 08 月 01 日至 99 年 09 月 30 日
執 行 單 位 ： 國立政治大學應用數學學系

計 畫 主 持 人 ： 姜志銘

計畫參與人員： 碩士班研究生-兼任助理人員：曾琬甯
大專生-兼任助理人員：鄭琳揚

報 告 附 件 ： 出席國際會議研究心得報告及發表論文

處 理 方 式 ： 本計畫可公開查詢

中 華 民 國 99 年 12 月 28 日

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

On the random functional distributions

隨機函數分配

計畫類別：□個別型計畫　□整合型計畫
計畫編號：NSC－98－2118－M－004－002
執行期間：2009 年 8 月 1 日至 2010 年 9 月 30 日

執行機構及系所：國立政治大學應用數學系

計畫主持人：姜志銘
共同主持人：
計畫參與人員：曾琬甯、鄭琳陽

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

本計畫除繳交成果報告外，另須繳交以下出國心得報告：
□赴國外出差或研習心得報告
□赴大陸地區出差或研習心得報告
■出席國際學術會議心得報告
□國際合作研究計畫國外研究報告

處理方式：除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

中　華　民　國　99　年　12　月　27　日

On the random functional distributions

隨機函數分配

摘要

在貝式統計無母數模型所使用的隨機分配中,最常用的就是 Ferguson-Dirichlet 過程。首先我們提供 Jiang(1988) and Jiang, Dickey, and Kuo(2004)所介紹的 C-特徵函數的一些性質;接著,我們再提供直接尋找 Ferguson-Dirichlet 過程在三維球形實心的隨機均數的分配的方法;最後,我們找到也發現了一個 Ferguson-Dirichlet 過程在三維球體內部具母數測度的隨機均數的機率密度函數。這是一個從 Jiang(1991), Jiang, Dickey, and Kuo (2004) 以及 Jiang and Kuo(2008)所得到的結果之後的一個重要推廣。

關鍵字:Ferguson-Dirichlet 過程,C-特徵函數,隨機均數,球狀分配

Abstract

The Ferguson-Dirichlet process is the most widely used nonparametric model for random distributions in Bayesian statistics. We first give the properties of the c-characteristic functions, which are first given by Jiang (1988), and Jiang, Dickey, and Kuo (2004). Then, we give direct procedure for finding the explicit distribution of random mean of the Ferguson-Dirichlet process with parameter measure over three-dimensional spherical solid. Finally, we find the new probability density function of the random mean of the Ferguson-Dirichlet process with parameter measure over three-dimensional ball. This is an important generalization of results given by Jiang (1991), Jiang, Dickey, and Kuo (2004) and Jiang, and Kuo (2008)

Keywords: Ferguson-Dirichlet process, c-characteristic function, random means, spherical distribution

1 Introduction

The univariate c-characteristic function and multivariate c-characteristic function have been shown to be able to solve many problems that are difficult to manage using the traditional characteristic function. Examples can be seen in Jiang (1988, 1991) and Jiang, Dickey, and Kuo (2004). In this research, for simplicity, we will refer to both the univariate and multivariate c-characteristic functions as c-characteristic function.

We give a new property of the c-characteristic function for a spherical distribution in Section 2. This property is useful in determining whether the corresponding random variable (random vector) has a spherical distribution when its c-characteristic function is known. In addition, we can easily obtain the c-characteristic function for a marginal distribution when the multivariate c-characteristic function for a spherical distribution is known.

The Ferguson-Dirichlet process was first introduced and studied by Ferguson (1973) with regard to its applications to Bayesian nonparametric statistics. Since then, several researchers have studied random functional of the Ferguson-Dirichlet process. Recently, the issue of finding the exact distributions of functionals (means, for example) has drawn the attention of researchers. See, for example, Jiang and Kuo (2008), Hjort and Ongaro (2005), Jiang, Dickey, and Kuo (2004), Regazzini, Guglielmi, and Di Nunno (2002), and references therein. For some examples of applications of distributions of random functionals, see Dickey, Garthwaite, and Bian (1995) and Dickey and Jiang (1998). Only a few distributional results for a vector of functionals of a Ferguson-Dirichlet process are available; for instance, Jiang (1991) studied the distribution of the random mean of a Ferguson-Dirichlet process with parameter measure over a unit circle, and Jiang, Dickey, and Kuo (2004) further generalized the result to an ellipse. Jiang and Kuo (2008) further generalized the result to the surface of three-dimensional unit ball. However, the results of the above studies were all obtained indirectly. In this research, we provide a new approach, which gives direct procedure for finding the explicit distribution of random mean of Ferguson-
Dirichlet process with parameter measure over three-dimensional spherical solid. In Section 3, we first give the c-characteristic function expression of any bounded functional for Ferguson-Dirichlet process with parameter measure over Euclidean space. With this expression, we show that the random mean of spherically symmetric Ferguson-Dirichlet process with parameter measure over three-dimensional solid has spherical distributions on the three-dimensional ball. Using this result, we provide its exact probability density function, which is a new result. Finally, we give conclusions in Section 4.

2 Spherical properties of the c-characteristic function

First, we state the definition of the c-characteristic function.

Definition 2.1 (Jiang, Dickey, and Kuo, 2004) If $\mathbf{u}=(u_1,...,u_L)'$ is a random vector on a subset of $A=[-a_1,a_1]\times\cdots\times[-a_L,a_L]$, its c-characteristic function is defined as

$$g(\mathbf{t};\mathbf{u},c) = \underset{\mathbf{u}}{E}\left[(1 - i\mathbf{t}\cdot\mathbf{u})^{-c}\right], |\mathbf{t}| < |\mathbf{a}|^{-1},$$

where $c > 0, \mathbf{a} = (a_1,...,a_L)', \mathbf{t} = (t_1,...,t_L)', |\mathbf{t}| = \sqrt{\sum_{i=1}^{L} t_i^2}$, and $\mathbf{t}\cdot\mathbf{u} = \sum_{j=1}^{L} t_j u_j$, the inner product of **t** and **u**.

Some important properties, e.g., uniqueness and convergence theorems, of the c-characteristic function can be seen in Jiang, Dickey, and Kuo (2004). In the following, we give the definition of a spherically symmetric distribution.

Definition 2.2 A random vector **u** is said to be spherically symmetric if T**u** has the same distribution as **u** for every orthogonal matrix T.

When the c-characteristic function $g(\mathbf{t};\mathbf{u},c)$ is known, the following theorem provides an easy method for determining whether **u** has a spherically symmetric distribution.

Theorem 2.3 Suppose that **u** is an L-dimensional random vector and $g(\mathbf{t};\mathbf{u},c)$ is its c-characteristic function. Then **u** has a spherically symmetric distribution if and only if $g(\mathbf{t};\mathbf{u},c)$ is a function of $|\mathbf{t}|$ and c, only.

We will call **v** (k-dimensional vector) a margin of **u** (L-dimensional vector) and the distribution of **v** a marginal distribution of the distribution of **u** if k≤L and **v** is a linear projection of **u**, that is, the coordinates of **v** are a sublist of the coordinates of D**u**, for some matrix D of size L×L, where $\mathrm{DD}' = \mathrm{I}_\mathrm{L}$ (the identity matrix with dimensional L). The following corollary shows that a spherically symmetric distribution and its marginal distribution have the same c-characteristic function expression and that the c-characteristic function in lower dimension can be derived easily from that in higher dimension.

Corollary 2.4 Let **u** be random vector having a spherically symmetric distribution and let **v** be any margin of **u**. Suppose that $g(\mathbf{t};\mathbf{u},c) = h(|\mathbf{t}|,\mathrm{c})$ for some function h. Then

$$g(\mathbf{s};\mathbf{v},c) = h(|\mathbf{s}|,\mathrm{c}).$$

With Theorem 2.3 and Corollary 2.4, we see that any marginal distribution of a spherically symmetric distribution is also spherically symmetric distribution.

3 Random functional of a Ferguson-Dirichlet process with parameter measure over a three-dimensional ball

Let μ be a finite non-null measure on $(\Omega, \mathbf{B})$, where $\mathbf{B}$ is the σ- *field* of Borel subsets of Euclidean space Ω, and let U be a stochastic process indexed by elements of $\mathbf{B}$. We say U is a Ferguson-Dirichlet process with parameter μ, denoted by $\mathrm{U} \sim \mathrm{D}(\mu)$ on Ω, if for every finite measurable partition $\{\mathrm{B}_1,...,\mathrm{B}_m\}$ of Ω (i.e., B_i's are measurable, disjoint, and $\bigcup_{i=1}^m B_i = \Omega$), the random vector $\{\mathrm{U}(\mathrm{B}_1),...,\mathrm{U}(\mathrm{B}_m)\}$ has a Dirichlet distribution with parameter vector $\{\mu(\mathrm{B}_1),...,\mu(\mathrm{B}_m)\}$, where $\bigcup_{i=1}^m \mathrm{U}(B_i) = 1$.

Let $\mathbf{X}$ be a N-dimensional random vector defined as

$$\mathbf{X} = \int_\Omega \mathbf{h}(\mathbf{y}) dU(\mathbf{y}) = (\int_\Omega h_1(\mathbf{y}) dU(\mathbf{y}),..., \int_\Omega h_N(\mathbf{y}) dU(\mathbf{y}))', \qquad (1)$$

where $\mathrm{U} \sim \mathrm{D}(\mu)$ on Ω, $\mu(\Omega) = \mathrm{c}$, $\mathbf{y} = (y_1,..., y_L)'$, $\mathbf{h}(\mathbf{y}) = (h_1(\mathbf{y}),..., h_N(\mathbf{y}))'$, and the $h_n(\mathbf{y})$'s are bounded measurable real-value functions defined on Ω. The following lemma gives the c-characteristic function expression for $\mathbf{X}$.

Lemma 3.1 The c-characteristic function of $\mathbf{X}$, as in Eq. (1), can be expressed by

$$g(\mathbf{t}; \mathbf{X}, c) = \exp[-\int_\Omega \ln(1 - \mathrm{i}\mathbf{t} \cdot \mathbf{h}(\mathbf{y})) \mathrm{d}\mu(\mathbf{y})],$$

where $\mathbf{t} = (t_1,..., t_N)'$.

The following corollary can be obtained by applying the above Lemma 3.1 and Theorem 2.5 of Jiang, Dickey, and Kuo (2004).

Corollary 3.2 Let $\mathbf{X} = (\mathrm{X}_1,..., \mathrm{X}_\mathrm{N})'$ and $\mathbf{h}(\mathbf{y}) = (h_1(\mathbf{y}),..., h_N(\mathbf{y}))'$. Then $\mathrm{E}(X_n) = \frac{1}{c} \int_\Omega \mathbf{h}_n(\mathbf{y}) d\mu(\mathbf{y})$,

for $1 \le n \le N$,

and

$$\mathrm{Cov}(X_n, X_m) = \frac{1}{c+1} [\frac{1}{c} \int_\Omega h_n(\mathbf{y}) h_m(\mathbf{y}) d\mu(\mathbf{y}) - (\frac{1}{c} \int_\Omega h_n(\mathbf{y}) d\mu(\mathbf{y}))(\frac{1}{c} \int_\Omega h_m(\mathbf{y}) d\mu(\mathbf{y}))], \text{ for } 1 \le n, m \le N.$$

The following corollary is obtained immediately from Corollary 3.2.

Corollary 3.3 Let $\mathbf{Y}$ be a random vector associated with the probability measure μ / c. Then the mean vector and the variance-covariance matrix of $\mathbf{X}$, Eq. (1), can be expressed as $\mathrm{E}(\mathbf{X}) = \mathrm{E}(\mathbf{h}(\mathbf{Y}))$ and $\mathrm{Cov}(\mathbf{X}) = \frac{1}{\mathrm{c}+1} Cov(\mathbf{h}(\mathbf{Y}))$, respectively.

In the following, we want to study the distribution of

$$\tilde{\mathbf{B}}_r = \int_{B_r} \mathbf{y} dV_r(\mathbf{y}),\ V_r \sim \mathrm{D}(\nu_r) \qquad (2)$$

where $B_r = \{\mathbf{y} \in R^3 \mid |\mathbf{y}| \le r, r > 0\}$ is the three dimensional spherical ball, and where ν_r is the usual Lebesgue measure on B_r with total measure c.

Just as the parameter measure ν_r is spherically symmetric, the Ferguson-Dirichlet process $V_r \sim \mathrm{D}(\nu_r)$ will be said to be spherically symmetric. The c-characteristic function expression for $\tilde{\mathbf{B}}_r$ can be obtained as follows.

Theorem 3.4 The c-characteristic function of $\tilde{\mathbf{B}}_r$ in Eq (2) can be expressed as

$$g(\mathbf{t};\tilde{\mathbf{B}}_r, c) = \exp\{\sum_{k=1}^{\infty} \frac{\mathrm{c}(1/2,\mathrm{k})}{2\mathrm{k}(3/2+1,\mathrm{k})} [-\mathrm{r}^2(\mathrm{t}_1^2 + t_2^2 + t_3^2)]^{\mathrm{k}}\},$$

where $\mathbf{t} = (t_1, t_2, t_3)'$.

By Corollary 2.4, the c-characteristic function of any one-dimensional margin of $\tilde{\mathbf{B}}_r$, say $\tilde{\mathbf{B}}_r$, is expressed as

$$g(t;\tilde{\mathrm{B}}_r, c) = \exp\{\sum_{k=1}^{\infty} \frac{\mathrm{c}(1/2,\mathrm{k})}{2\mathrm{k}(3/2+1,\mathrm{k})} (-\mathrm{r}^2 t^2)^{\mathrm{k}}\}. \qquad (3)$$

Our aim now is to find the probability density functions of $\tilde{\mathbf{B}}_r$. Lord (1954) demonstrated that a spherically symmetric distribution can be determined by its marginal distribution. Therefore, if the probability density function of the marginal random variable $\tilde{\mathbf{B}}_r$ is known, then we can obtain the probability density function of the random vector $\tilde{\mathbf{B}}_r$. Now the important p. d. f. of the $\tilde{\mathbf{B}}_r$ can be shown in the following theorem.

Theorem 3.5 Let $h_{\tilde{\mathbf{B}}}(\mathbf{x}; r)$ be the probability density function of $\tilde{\mathbf{B}}_{\mathbf{r}} = \int_{B_r} \mathbf{y} dV_r(\mathbf{y})$ defined in Eq. (2), where $\mathbf{x} = (x_1, x_2, x_3)'$, and $\nu_{\mathrm{r}}(\mathrm{B}_{\mathrm{r}}) = 1$, then

$$h_{\tilde{\mathbf{B}}}(x_1, x_2, x_3; r) = \frac{-3e^{4/3-t^2/(2r^2)}}{8tr^3\pi^2}(r+t)^{-(2r-t)(r+t)^2/(4r^3)}(r-t)^{-(r-t)^2(2r+t)/(4r^3)}$$
$$\times[\pi(r^2-t^2)\cos\frac{\pi(2r-t)(r+t)^2}{4r^3} + (-2rt + (r^2-t^2)\ln\frac{r-t}{r+t}) \times \sin\frac{\pi(2r-t)(r+t)^2}{4r^3},$$

where $\mathrm{t} = \sqrt{x_1^2 + x_2^2 + x_3^2}$ and $x_1^2 + x_2^2 + x_3^2 < r^2$.

The above theorem is an important generalization of the results given by Jiang (1991) Jiang, Dickey, and Kuo (2004) and by Jiang and Kuo (2008).

4 Conclusions

Through the c-characteristic function, we have given a new approach for studying spherical distributions. The c-characteristic function expression of any functional of any Ferguson-Dirichlet process has also been given. With this expression, we can obtain properties, e.g. any moment, of the random functional. In addition, we have given the exact distribution in 3 dimension of the random mean of a symmetric Ferguson-Dirichlet process with parameter measure over a spherical solid.

References

Dickey, J.M., Garthwaite, P.H., and Bian, G., 1995. An elementary continuous-type nonparametric distribution estimate. Int. J. Math. Stat. Sci. 4, 193-247

Dickey, J.M. and Jiang, T.J., 1998. Filtered-variate prior distributions for histogram smoothing. J. Amer. Statist. Assoc. 93, 651-662.

Ferguson, T.S., 1973. A Bayesian analysis of some nonparametric problems. Ann. Statist.1 209-230.

Hjort, N.L. and Ongaro, A.,2005. Exact inference for random Dirichlet means. Stat. Inference Stoch. Process. 8, 227-254.

Jiang, J., 1988. Starlike functions and linear functions of a Dirichlet distributed vector. SIAM J. Math. Anal. 19, 390-397.

Jiang, T.J., 1991. Distribution of random functional of Dirichlet process on the unit disk. Statist. Probab. Lett. 12, 263-265.

Jiang, T.J., Dickey, J.M., and Kuo, K.-L., 2004. A new multivariate transform and the distribution of a random functional of Ferguson-Dirichlet process. Stochastic Process. Appl. 111, 77-95.

Jiang, T.J., and Kuo,K.-L., 2008. Distribution of random functional of a Ferguson-Dirichlet process over the unit circle. Electron. Comm. Probab. 13, 518-525.

Lord, R. D. 1954. The used of the Hankel transformations in statistics. I. General theory and example. Biometrika 41,44-55.

Regazzini, E., Guglielmi, A., and Di Nunno, G., 2002. Theory and numerical analysis for exact distributions of functionals of a Dirichlet process. Ann. Statist. 30, 1376-1411.

國科會補助專題研究計畫項下出席國際學術會議心得報告

日期：___年___月___日

計畫編號	NSC－98－2118－M－004－002		
計畫名稱	隨機函數分配		
出國人員姓名	姜志銘	服務機構及職稱	國立政治大學應用數學系教授
會議時間	2010 年 7 月 31 日 至 2010 年 8 月 5 日	會議地點	加拿大溫哥華
會議名稱	(中文)2010 聯合統計季會議 (英文)2010 Joint Statistical Meetings		
發表論文題目	(中文) (英文) On the compatibility of conditional normal distributions		

一、參加會議經過

每年由美國統計學會及其他多個國際知名的統計學會，如 IMS、ENAR、WNAR 及 SSC 聯合主辦的聯合統計會議，一直為全球最大型的統計會議，今年在加拿大溫哥華舉辦，吸引來自全球各地約 5~6 千名的統計專家學者們，共同為追求及交換新的統計理論與方法而來參加，本人亦抱持著這種態度，希望藉由這個機會與其他的統計學者專家作學術上的交流。

二、與會心得

我的演講題目為「On the compatibility of conditional normal distributions」。一般而言，一維的條件分配比二維的聯合分配容易建立，但兩個一維的條件分配是否相容(compatible)，就顯得非常重要，即使相容，他們對應的二維聯合分配是否唯一以及如何表示亦相當重要。此論文提供了在常態分配下，回答此問題的新方法。在實務上，即使兩條件分配相容，但因抽樣的關係，不太可能完全相容，所以本文也同時探討此種不相容度(degrees of incompatibility)的問題。演講中也引起與會者的興趣與發問。

跟往年一樣，2010 年聯合統計會議涵蓋的領域非常廣泛，從理論到應用，甚至目前大家所關心的區域性甚至國際性的議題，如溫室效應等等，都可以看到與會者所展現的最新研究成果。本次年會特別邀請相關領域的知名專家學者作大會特邀報告，包括星期一（8/2）ASA President's Invited Address，請 SAS 公司創辦人 Jim Goodnite 博士報告 "The Forecast for Predictive Analytic: Hot and Getting Hotter"，星期二（8/3）The Deming Lecture 則邀請 Institute for Health Care Delivery Research 的 Brent C. James 介紹 "Better: Dr. Deming consults on Quality for Sir William Osler"，以及星期三（8/4）COPSS Fisher Lecture 邀請賓州大學的 Bruce G. Lindsay 介紹 "Likelihood with Hidden Variables"，透過參與這些會議，以及會議期間與所遇到的部份學者、專家，互相交換最近一些研究領域方向的看法，使得參加這次會議得益不少。

最後，謝謝國科會給予這次機會參加這個有意義的會議。

三、考察參觀活動(無是項活動者略)

四、建議

五、攜回資料名稱及內容

六、其他

國科會補助計畫衍生研發成果推廣資料表

日期:2010/12/28

國科會補助計畫	計畫名稱: 隨機函數分配	
	計畫主持人: 姜志銘	
	計畫編號: 98-2118-M-004-002-	學門領域: 機率

無研發成果推廣資料

98 年度專題研究計畫研究成果彙整表

計畫主持人：姜志銘			計畫編號：98-2118-M-004-002-				
計畫名稱：隨機函數分配							
成果項目			量化			單位	備註（質化說明：如數個計畫共同成果、成果列為該期刊之封面故事...等）
			實際已達成數（被接受或已發表）	預期總達成數(含實際已達成數)	本計畫實際貢獻百分比		
國內	論文著作	期刊論文	0	0	100%	篇	
		研究報告/技術報告	0	0	100%		
		研討會論文	0	0	100%		
		專書	0	0	100%		
	專利	申請中件數	0	0	100%	件	
		已獲得件數	0	0	100%		
	技術移轉	件數	0	0	100%	件	
		權利金	0	0	100%	千元	
	參與計畫人力（本國籍）	碩士生	0	0	100%	人次	
		博士生	0	0	100%		
		博士後研究員	0	0	100%		
		專任助理	0	0	100%		
國外	論文著作	期刊論文	0	1	100%	篇	
		研究報告/技術報告	1	1	100%		
		研討會論文	1	1	100%		
		專書	0	0	100%	章/本	
	專利	申請中件數	0	0	100%	件	
		已獲得件數	0	0	100%		
	技術移轉	件數	0	0	100%	件	
		權利金	0	0	100%	千元	
	參與計畫人力（外國籍）	碩士生	1	1	100%	人次	
		博士生	0	0	100%		
		博士後研究員	0	0	100%		
		專任助理	0	0	100%		

其他成果 (無法以量化表達之成果如辦理學術活動、獲得獎項、重要國際合作、研究成果國際影響力及其他協助產業技術發展之具體效益事項等，請以文字敘述填列。)	無

	成果項目	量化	名稱或內容性質簡述
科教處計畫加填項目	測驗工具(含質性與量性)	0	
	課程/模組	0	
	電腦及網路系統或工具	0	
	教材	0	
	舉辦之活動/競賽	0	
	研討會/工作坊	0	
	電子報、網站	0	
	計畫成果推廣之參與（閱聽）人數	0	

國科會補助專題研究計畫成果報告自評表

請就研究內容與原計畫相符程度、達成預期目標情況、研究成果之學術或應用價值（簡要敘述成果所代表之意義、價值、影響或進一步發展之可能性）、是否適合在學術期刊發表或申請專利、主要發現或其他有關價值等，作一綜合評估。

1. 請就研究內容與原計畫相符程度、達成預期目標情況作一綜合評估

■達成目標

□未達成目標（請說明，以 100 字為限）

□實驗失敗

□因故實驗中斷

□其他原因

說明：

2. 研究成果在學術期刊發表或申請專利等情形：

論文：□已發表 ■未發表之文稿 □撰寫中 □無

專利：□已獲得 □申請中 ■無

技轉：□已技轉 □洽談中 ■無

其他：（以 100 字為限）

3. 請依學術成就、技術創新、社會影響等方面，評估研究成果之學術或應用價值（簡要敘述成果所代表之意義、價值、影響或進一步發展之可能性）（以 500 字為限）

在貝式統計無母數模型所使用的隨機分配中,最常用的就是 Ferguson-Dirichlet 過程。Jiang(1991)提供了 Ferguson-Dirichlet 過程在圓周上的隨機均數的機率密度函數,Jiang, Dickey, and Kuo(2004)將圓周擴大為橢圓,Jiang and Kuo(2008)進一步擴大為三維之球表面。本研究則再作了一個很重要的擴充,再度從表面擴充到內部,提供了 Ferguson-Dirichlet 過程在球體內部的隨機均數的完整機率密度函數的表示法。如同前述文章,本研究結果,預期將發表於重要期刊。