ELECTRONIC COMMUNICATIONS in PROBABILITY

DISTRIBUTION OF A RANDOM FUNCTIONAL OF A FERGUSON-DIRICHLET PROCESS OVER THE UNIT SPHERE

THOMAS J. JIANG

Department of Mathematical Sciences, National Chengchi University, Taipei 11605, Taiwan

email: jiangt@nccu.edu.tw

KUN-LIN KUO

Institute of Statistical Science, Academia Sinica, Taipei 11529, Taiwan

email: KunLin.Kuo@gmail.com

Submitted October 15, 2007, accepted in final form September 9, 2008

AMS 2000 Subject classification: 60E10, 62E15

Keywords: Ferguson-Dirichlet process, *c*-characteristic function

Abstract

Jiang, Dickey, and Kuo [12] gave the multivariate c-characteristic function and showed that it has properties similar to those of the multivariate Fourier transformation. We first give the multivariate c-characteristic function of a random functional of a Ferguson-Dirichlet process over the unit sphere. We then find out its probability density function using properties of the multivariate c-characteristic function. This new result would generalize that given by [11].

1 Introduction

Ferguson [5] introduced the Ferguson-Dirichlet process and studied its applications to nonparametric Bayesian inference. He also showed that when the prior distribution is a Ferguson-Dirichlet process with parameter μ , then the posterior distribution, given the sample s_1, s_2, \ldots, s_n , is also a Ferguson-Dirichlet process having parameter $\mu + \sum_{j=1}^n \delta_{s_j}$, where δ_{s_j} denotes point mass at s_j . The most natural use of random functionals of a Ferguson-Dirichlet process is to make Bayesian inferences concerning the parameters of a statistical population. Hence, the expression for the probability density function of any random functional of a Ferguson-Dirichlet process can be employed both for prior and posterior Bayesian analyses. Further applications related to the random functional can be seen in [3] and other references. For example, random means and random variances of a Ferguson-Dirichlet process can be used for smooth Bayesian nonparametric density estimation (see [15]) and for quality control problems (see [4] for further discussions), respectively.

Research on the distribution of a random functional of a Ferguson-Dirichlet process has been ongoing for decades. A partial list of papers in this area are [2, 3, 8, 9, 11, 12, 14, 16, 17]. In particular, [11] gave the distribution of a random functional of a Ferguson-Dirichlet process over the unit circle. In this paper, we shall use the multivariate *c*-characteristic function, a tool given

by [12], to generalize the result to the case over the unit sphere in three-dimension.

In Section 2, we first review the definition of the multivariate *c*-characteristic function and some of its properties. We then compute a multivariate *c*-characteristic function of an interesting distribution. The multivariate *c*-characteristic function of the random mean of a Ferguson-Dirichlet process over the unit sphere is given in Section 3. Using the uniqueness property of the multivariate *c*-characteristic function, we then determine the distribution of the random mean of a Ferguson-Dirichlet process over the unit sphere. Conclusions are given in Section 4.

2 Multivariate *c*-characteristic function

Jiang [10] first gave a univariate c-characteristic function. Jiang, Dickey, and Kuo [12] generalized it to a multivariate c-characteristic function, which can be very useful when a distribution is difficult to deal with by traditional characteristic function. See [12] for detailed results. First, we state the definition of the multivariate c-characteristic function.

Definition 1. If $\mathbf{u} = (u_1, \dots, u_L)'$ is a random vector on a subset S of $A = [-a_1, a_1] \times \dots \times [-a_L, a_L]$, its multivariate c-characteristic function is defined as

$$g(t; u, c) = E[(1 - it \cdot u)^{-c}], |t| < a^{-1},$$

where c > 0, $a = \sqrt{\sum_{i=1}^{L} a_i^2}$, $\mathbf{t}' = (t_1, \dots, t_L)$, $|\mathbf{t}| = \sqrt{\sum_{i=1}^{L} t_i^2}$, and $\mathbf{t} \cdot \mathbf{u}$ is the inner product of \mathbf{t} and \mathbf{u} .

The above assumptions that c is positive and u has a bounded support are needed in [12, Lemma 2.2], which shows that, for any positive c, there is a one-to-one correspondence between g(t; u, c) and the distribution of u.

Next, we give the multivariate c-characteristic function of an interesting distribution in the next lemma.

Lemma 2. Let $\mathbf{u} = (u_1, u_2, u_3)'$ be a distribution on the inside of a unit ball, i.e., $\{(u_1, u_2, u_3) \mid u_1^2 + u_2^2 + u_3^2 < 1\}$, with the probability density function

$$f(u) = \frac{-e}{4\pi^2 r} (1+r)^{-(1+r)/2} (1-r)^{-(1-r)/2} \left(-\pi \sin \frac{\pi r}{2} + \ln \frac{1-r}{1+r} \cos \frac{\pi r}{2} \right),$$

where $r = |\mathbf{u}|$. Then the multivariate 1-characteristic function of \mathbf{u} is

$$g(t; u, 1) = \exp\left(\sum_{n=1}^{\infty} \frac{(-t_1^2 - t_2^2 - t_3^2)^n}{2n(2n+1)}\right).$$
 (1)

Proof. Let $C = \{(u_1, u_2, u_3) \mid u_1^2 + u_2^2 + u_3^2 < 1\}$. Eq. (1) is equivalent to the following identity

$$\int_{C} (1 - i\mathbf{t} \cdot \mathbf{u})^{-1} f(\mathbf{u}) d\mathbf{u} = \exp\left(\sum_{n=1}^{\infty} \frac{(-t_1^2 - t_2^2 - t_3^2)^n}{2n(2n+1)}\right).$$

To prove the above identity, we establish the following four equations first. From [7, p. 105], we have

$$\int_0^{2\pi} (a\cos\alpha + b\sin\alpha)^n d\alpha = \begin{cases} \frac{(1/2, n/2)2(a^2 + b^2)^{n/2}\pi}{(n/2)!}, & n \text{ is even,} \\ 0, & n \text{ is odd,} \end{cases}$$
(2)

where a and b are real numbers and $(a,k) = a(a+1)\cdots(a+k-1)$. We also can obtain the following equation from [6, Eq. 3.621.5],

$$\int_0^{\pi} \sin^{a-1} x \cos^{b-1} x \, dx = \begin{cases} \frac{B(a/2, b/2)}{2}, & \text{Re } a > 0, b > 0 \text{ is odd,} \\ 0, & \text{Re } a > 0, b > 0 \text{ is even.} \end{cases}$$
 (3)

Using integration by parts, we have the following identity,

$$\int_{0}^{1} r^{2n+1} (1+r)^{-(1+r)/2} (1-r)^{-(1-r)/2} \left(-\pi \sin \frac{\pi r}{2}\right) dr$$

$$= -\int_{0}^{1} r^{2n+1} (1+r)^{-(1+r)/2} (1-r)^{-(1-r)/2} \ln \frac{1-r}{1+r} \cos \frac{\pi r}{2} dr$$

$$-\int_{0}^{1} 2(2n+1)r^{2n} (1+r)^{-(1+r)/2} (1-r)^{-(1-r)/2} \cos \frac{\pi r}{2}.$$
(4)

Using [13, Lemma 8 and Example 2], we can obtain the following equality:

$$\exp\left(-\int_{-1}^{1}\ln(1-itx)\frac{1}{2}\,dx\right) = \int_{-1}^{1}(1-itx)^{-1}\frac{e}{\pi}(x+1)^{-(x+1)/2}(1-x)^{-(1-x)/2}\cos\frac{\pi x}{2}\,dx.$$

Since

$$\exp\left(-\int_{-1}^{1} \ln(1-itx)\frac{1}{2} dx\right) = \exp\left(\sum_{n=1}^{\infty} \frac{(-t^2)^n}{2n(2n+1)}\right)$$

and

$$\int_{-1}^{1} (1 - itx)^{-1} \frac{e}{\pi} (x+1)^{-(x+1)/2} (1-x)^{-(1-x)/2} \cos \frac{\pi x}{2} dx$$

$$= \sum_{n=0}^{\infty} \int_{-1}^{1} \frac{ei^{n} t^{n}}{\pi} x^{n} (x+1)^{-(x+1)/2} (1-x)^{-(1-x)/2} \cos \frac{\pi x}{2} dx,$$

and by the fact that the function $(x+1)^{-(x+1)/2}(1-x)^{-(1-x)/2}\cos\frac{\pi x}{2}$ is symmetric at x=0, we have

$$\exp\left(\sum_{n=1}^{\infty} \frac{(-t^2)^n}{2n(2n+1)}\right) = \frac{2e}{\pi} \sum_{n=0}^{\infty} (-t^2)^n \int_0^1 x^{2n} (x+1)^{-(x+1)/2} (1-x)^{-(1-x)/2} \cos\frac{\pi x}{2} \, dx. \tag{5}$$

Setting

$$g(r) = \frac{-er}{4\pi^2} (1+r)^{-(1+r)/2} (1-r)^{-(1-r)/2} \left(-\pi \sin \frac{\pi r}{2} + \ln \frac{1-r}{1+r} \cos \frac{\pi r}{2} \right),$$

and using the spherical coordinate transformation, we have

$$\int_{C} (1 - it \cdot u)^{-1} f(u) du$$

$$= \int_{0}^{1} \int_{0}^{2\pi} \int_{0}^{\pi} (1 - it_{1}r \cos\theta \sin\phi - it_{2}r \sin\theta \sin\phi - it_{3}r \cos\phi)^{-1} \sin\phi g(r) d\phi d\theta dr$$

$$= \int_{0}^{1} \sum_{n=0}^{\infty} (ir)^{n} g(r) \int_{0}^{2\pi} \int_{0}^{\pi} (t_{1} \cos\theta \sin\phi + t_{2} \sin\theta \sin\phi + t_{3} \cos\phi)^{n} \sin\phi d\phi d\theta dr$$

$$= \int_{0}^{1} \sum_{n=0}^{\infty} (ir)^{n} g(r) \int_{0}^{2\pi} \int_{k=0}^{\pi} \sum_{k=0}^{n} {n \choose k} (t_{1} \cos\theta + t_{2} \sin\theta)^{k} t_{3}^{n-k} \sin^{k+1}\phi \cos^{n-k}\phi d\phi d\theta dr$$

$$= \int_{0}^{1} \sum_{n=0}^{\infty} \frac{4\pi (-t_{1}^{2} - t_{2}^{2} - t_{3}^{2})^{n} r^{2n}}{2n+1} g(r) dr \qquad (6)$$

$$= \frac{2e}{\pi} \sum_{n=0}^{\infty} (-t_{1}^{2} - t_{2}^{2} - t_{3}^{2})^{n} \int_{0}^{1} r^{2n} (1+r)^{-(1+r)/2} (1-r)^{-(1-r)/2} \cos\frac{\pi r}{2} dr \qquad (7)$$

$$= \exp\left(\sum_{n=1}^{\infty} \frac{(-t_{1}^{2} - t_{2}^{2} - t_{3}^{2})^{n}}{2n(2n+1)}\right). \qquad (8)$$

Identity (6) can be obtained by Eqs. (2) and (3). Identities (7) and (8) follow from Eq. (4) and Eq. (5), respectively. \Box

3 Distribution of a random functional of a Ferguson-Dirichlet process over the unit sphere

Ferguson [5] first defined the Ferguson-Dirichlet process. Let μ be a finite non-null measure on (Y,A), where Y is a Borel set in Euclidean space \mathbb{R}^n and A is the σ -field of Borel subsets of Y, and let U be a stochastic process indexed by elements of A. We say that U is a Ferguson-Dirichlet process with parameter μ , if for every finite measurable partition $\{B_1,\ldots,B_m\}$ of Y, the random vector $(U(B_1),\ldots,U(B_m))$ has a Dirichlet distribution with parameter $(\mu(B_1),\ldots,\mu(B_m))$, where $\mu(B_j)>0$ for all $j=1,\ldots,m$. A random vector $\mathbf{v}=(v_1,\ldots,v_m)'$ is said to have a Dirichlet distribution with parameter $\mathbf{b}=(b_1,\ldots,b_m)'$ where each $b_j>0$, if \mathbf{v} has the probability density function

$$f(\boldsymbol{\nu};\boldsymbol{b}) = \frac{\Gamma(b_1 + \dots + b_m)}{\prod_{j=1}^m \Gamma(b_j)} \prod_{j=1}^m v_j^{b_j - 1},$$

for all ν in the probability simplex $\{\nu \mid \text{each } \nu_i \geq 0, \nu_1 + \dots + \nu_m = 1\}.$

First, we give a trivariate c-characteristic function expression of any trivariate random functional of a Ferguson-Dirichlet process over a Borel set Y in Euclidean space in the next lemma.

Lemma 3. Let $\mathbf{w} = \int_Y \mathbf{h}(\mathbf{x}) dU(\mathbf{x})$ be a random functional where $\mathbf{h}(\mathbf{x}) = (h_1(\mathbf{x}), h_2(\mathbf{x}), h_3(\mathbf{x}))'$ is a bounded measurable function defined on a Borel set Y in Euclidean space \mathbb{R}^n , and U is a Ferguson-Dirichlet process with parameter μ on (Y,A). Then the trivariate c-characteristic function of \mathbf{w} can

be expressed as

$$g(t; \boldsymbol{w}, c) = \exp\left(-\int_{Y} \ln(1 - it \cdot \boldsymbol{h}(\boldsymbol{x})) d\mu(\boldsymbol{x})\right), \text{ where } c = \mu(Y).$$

Proof. For any $k \geq 2$, let $\{B_{k1}, B_{k2}, \ldots, B_{kk}\}$ be a partition of Y, $\boldsymbol{b}_{kj} \in B_{kj}$, $v_k = \max\{\text{volume}(B_{kj}) \mid 1 \leq j \leq k\}$, and $\lim_{k \to \infty} v_k = 0$. Then $(U(B_{k1}), \ldots, U(B_{kk}))$ follows a Dirichlet distribution with parameter $(\mu(B_{k1}), \ldots, \mu(B_{kk}))$. In addition, $\sum_{j=1}^k U(B_{kj}) = 1$ for all $k \geq 2$. Define $\boldsymbol{g}_k(\boldsymbol{x}) = \sum_{j=1}^k \boldsymbol{h}(\boldsymbol{b}_{kj}) \delta_{B_{kj}}(\boldsymbol{x})$ and $\boldsymbol{w}_k = \int_Y \boldsymbol{g}_k(\boldsymbol{x}) dU(\boldsymbol{x})$, where $\delta_{B_{kj}}(\boldsymbol{x})$ is 1, for $\boldsymbol{x} \in B_{kj}$; and is 0, otherwise. Then $\lim_{k \to \infty} \boldsymbol{g}_k(\boldsymbol{x}) = \boldsymbol{h}(\boldsymbol{x})$ for all $\boldsymbol{x} \in Y$, and $\boldsymbol{w}_k = \sum_{j=1}^k \boldsymbol{g}_k(\boldsymbol{b}_{kj}) U(B_{kj})$. The trivariate c-characteristic function of \boldsymbol{w}_k can be expressed as

$$g(t; \mathbf{w}_{k}, c) = E(1 - i\mathbf{t} \cdot \mathbf{w}_{k})^{-c}$$

$$= E\left(1 - i\sum_{j=1}^{k} [\mathbf{t} \cdot \mathbf{g}_{k}(\mathbf{b}_{kj})] U(B_{kj})\right)^{-c}$$

$$= E\left(\sum_{j=1}^{k} U(B_{kj})[1 - i\mathbf{t} \cdot \mathbf{g}_{k}(\mathbf{b}_{kj})]\right)^{-c}$$

$$= \mathcal{R}_{-c}(\mu(B_{k1}), \dots, \mu(B_{kk}); 1 - i\mathbf{t} \cdot \mathbf{g}_{k}(\mathbf{b}_{k1}), \dots, 1 - i\mathbf{t} \cdot \mathbf{g}_{k}(\mathbf{b}_{kk}))$$

$$= \prod_{j=1}^{k} (1 - i\mathbf{t} \cdot \mathbf{g}_{k}(\mathbf{b}_{kj}))^{-\mu(B_{kj})},$$

where \mathcal{R} is a Carlson's multiple hypergeometric function ([1]), and the last equality can be obtained by [1, formula 6.6.5]. Therefore, the limit of the trivariate c-characteristic function of w_k 's, as k approaches ∞ , is

$$\lim_{k \to \infty} g(t; \boldsymbol{w}_k, c) = \exp\left(\lim_{k \to \infty} \sum_{j=1}^k -\mu(B_{kj}) \ln(1 - i\boldsymbol{t} \cdot \boldsymbol{g}_k(\boldsymbol{b}_{kj}))\right)$$
$$= \exp\left(-\int_Y \ln(1 - i\boldsymbol{t} \cdot \boldsymbol{h}(\boldsymbol{x})) d\mu(\boldsymbol{x})\right).$$

In addition, by the Dominated Convergence Theorem, we have $\lim_{k\to\infty} w_k = w$. By [12, Theorem 2.4], we conclude that

$$g(t; w, c) = \exp\left(-\int_{Y} \ln(1 - it \cdot h(x)) d\mu(x)\right).$$

In the rest of this section, we study the random functional $u = \int_X x \, dU(x)$, where X is the unit sphere in \mathbb{R}^3 . We use Lemma 3 in the following theorem to first establish the trivariate c-characteristic function of u.

Theorem 4. Let $X = \{(x_1, x_2, x_3) \mid x_1^2 + x_2^2 + x_3^2 = 1\}$, and U be a Ferguson-Dirichlet process over X with uniform measure μ as its parameter, where $\mu(X) = c$. Then the trivariate c-characteristic

function of the random functional $\mathbf{u} = \int_X \mathbf{x} \, dU(\mathbf{x})$ can be expressed as

$$g(t; u, c) = \exp\left(\sum_{n=1}^{\infty} \frac{c}{2n(2n+1)} (-t_1^2 - t_2^2 - t_3^2)^n\right), \text{ where } t = (t_1, t_2, t_3)'.$$

Proof. First, we give the following two equations, which are about Appell's notations and can be shown easily.

$$\Gamma(a+n) = \Gamma(a)(a,n), \tag{9}$$

$$(a,2n) = 2^{2n} \left(\frac{a}{2}, n\right) \left(\frac{a+1}{2}, n\right). \tag{10}$$

By Lemma 3, we have

$$\begin{split} g(t;u,c) &= \exp\left(\frac{-c}{4\pi} \int_{X} \ln(1-it \cdot x) dx\right) \\ &= \exp\left(\frac{-c}{4\pi} \int_{0}^{\pi} \int_{0}^{2\pi} \ln(1-it_{1}\cos\theta_{1}-it_{2}\sin\theta_{1}\cos\theta_{2}-it_{3}\sin\theta_{1}\sin\theta_{2})\sin\theta_{1} d\theta_{2} d\theta_{1}\right) \\ &= \exp\left(\frac{c}{4\pi} \sum_{n=1}^{\infty} \frac{i^{n}}{n} \int_{0}^{\pi} \int_{0}^{2\pi} (t_{1}\cos\theta_{1}+t_{2}\sin\theta_{1}\cos\theta_{2}+t_{3}\sin\theta_{1}\sin\theta_{2})^{n}\sin\theta_{1} d\theta_{2} d\theta_{1}\right) \\ &= \exp\left(\frac{c}{4\pi} \sum_{n=1}^{\infty} \frac{i^{n}}{n} \sum_{k=0}^{n} \binom{n}{k} \int_{0}^{\pi} \int_{0}^{2\pi} (t_{1}\cos\theta_{1})^{k} \sin^{n-k+1}\theta_{1} (t_{2}\cos\theta_{2}+t_{3}\sin\theta_{2})^{n-k} d\theta_{2} d\theta_{1}\right) \\ &= \exp\left(\frac{c}{2} \sum_{n=1}^{\infty} \frac{(-1)^{n}}{2n} \sum_{k=0}^{n} \binom{2n}{2k} \frac{(1/2,n-k)(t_{2}^{2}+t_{3}^{2})^{n-k}}{(n-k)!} t_{1}^{2k} B(n-k+1,k+1/2)\right) \\ &= \exp\left(\sum_{n=1}^{\infty} \frac{c}{2n(2n+1)} (-t_{1}^{2}-t_{2}^{2}-t_{3}^{2})^{n}\right). \end{split}$$

The fifth identity can be obtained by Eqs. (2) and (3). The last identity follows from Eqs. (9) and (10). \Box

By [12, Lemma 2.2], Lemma 2, and Theorem 4, we can obtain the following corollary.

Corollary 5. The probability density function of $\mathbf{u} = \int_X \mathbf{x} \, dU(\mathbf{x})$, where U is a Ferguson-Dirichlet process over the unit sphere X with uniform probability measure as its parameter, is

$$f(u) = \frac{-e}{4\pi^2 r} (1+r)^{-(1+r)/2} (1-r)^{-(1-r)/2} \left(-\pi \sin \frac{\pi r}{2} + \ln \frac{1-r}{1+r} \cos \frac{\pi r}{2} \right),$$

where $r = \sqrt{u_1^2 + u_2^2 + u_3^2}$ and $u_1^2 + u_2^2 + u_3^2 < 1$.

4 Conclusions

In this paper, we obtain the trivariate *c*-characteristic function expression for a random functional of a Ferguson-Dirichlet process over any finite three-dimensional space. We also obtain the probability density function of the random functional of a Ferguson-Dirichlet process with uniform probability measure parameter over the unit sphere. This generalizes [11, Theorem 2].

Acknowledgements

The authors are grateful for the comments by two referees, which improved the presentation of this paper. This work was supported in part by the National Science Council, Taiwan.

References

- [1] Carlson, B.C. (1977). Special Functions of Applied Mathematics. Academic Press, New York. MR0590943
- [2] Cifarelli, D.M. and Regazzini, E. (1990). Distribution functions of means of a Dirichlet process. Ann. Statist. 18, 429–442. MR1041402. Correction (1994): Ann. Statist. 22, 1633–1634. MR1311994
- [3] Diaconis, P. and Kemperman, J. (1996). Some new tools for Dirichlet priors, in: J.M. Bernardo, J.O. Berger, A.P. Dawid, and A.F.M. Smith eds. *Bayesian Statistics* 5, pp. 97–106. Oxford University Press. MR1425401
- [4] Epifani, I., Guglielmi, A., and Melilli, E. (2006). A stochastic equation for the law of the random Dirichlet variance. *Statist. Probab. Lett.* **76**, 495–502. MR2266603
- [5] Ferguson, T.S. (1973). A Bayesian analysis of some nonparametric problems. *Ann. Statist.* **1**, 209–230. MR0350949
- [6] Gradshteyn, I.S. and Ryzhik, I.M. (2000). *Table of Integrals, Series, and Products*. Academic Press, New York, 6th ed. MR1773820
- [7] Gröbner, W. and Hofreiter, W. (1973). Integraltafel, Vol. 2. Springer-Verlag, New York, 5th ed.
- [8] Hannum, R.C., Hollander, M., and Langberg, N.A. (1981). Distributional results for random functionals of a Dirichlet process. *Ann. Probab.* **9**, 665–670. MR0630318
- [9] Hjort, N.L. and Ongaro, A. (2005). Exact inference for random Dirichlet means. *Stat. Inference Stoch. Process.* **8**, 227–254. MR2177313
- [10] Jiang, J. (1988). Starlike functions and linear functions of a Dirichlet distributed vector. SIAM J. Math. Anal. 19, 390–397. MR0930035
- [11] Jiang, T.J. (1991). Distribution of random functional of a Dirichlet process on the unit disk. *Statist. Probab. Lett.* **12**, 263–265. MR1130367
- [12] 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 Process. Appl.* 111, 77–95. MR2049570
- [13] Jiang, T.J. and Kuo, K.-L. (2006). On the random functional of the Ferguson-Dirichlet process. 2006 Proceeding of the Section on Bayesian Statistical Science of the American Statistical Association, pp. 52–59.
- [14] Lijoi, A. and Regazzini, E. (2004). Means of a Dirichlet process and multiple hypergeometric functions. *Ann. Probab.* **32**, 1469–1495. MR2060305

- [15] Lo, A.Y. (1984). On a class of Bayesian nonparametric estimates: I. density estimates. *Ann. Statist.* **12**, 351–357. MR0733519
- [16] 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. MR1936323
- [17] Yamato, H. (1984). Characteristic functions of means of distributions chosen from a Dirichlet process. *Ann. Probab.* **12**, 262–267. MR0723745