

THE GEOMETRY OF THE DIRICHLET MANIFOLD

FENGWEI ZHONG, HUAFEI SUN, AND ZHENNING ZHANG

ABSTRACT. In the present paper, we investigate the geometric structures of the Dirichlet manifold composed of the Dirichlet distribution. We show that the Dirichlet distribution is an exponential family distribution. We consider its dual structures and give its geometric metrics, and obtain the geometric structures of the lower dimension cases of the Dirichlet manifold. In particular, the Beta distribution is a 2-dimensional Dirichlet distribution. Also, we construct an affine immersion of the Dirichlet manifold. At last, we give the e -flat hierarchical structures and the orthogonal foliations of the Dirichlet manifold. All these work will enrich the theoretical work of the Dirichlet distribution and will be great help for its further applications.

1. Introduction

Scientific researches and human activities generate lots of data, sometimes they are incomplete, redundant or erroneous, and probability methods are fairly useful in exploiting the patterns present in these data. Dirichlet distribution is the multivariate generalization of the Beta distribution, and it has many applications in engineering, also in contrast to other distributions, it can offer more considerable flexibility and ease of use, so it is a good choice for data modeling.

In this paper, we will give an intensive study to the n -dimensional Dirichlet distribution from the viewpoint of information geometry. Using the theory of information geometry, we can solve some problems with statistical characteristics. Since the set of the probability density function with parameters can be regarded as a manifold, it is quite meaningful to give its geometric structures.

Let $X = (x_1, x_2, \dots, x_{n-1})$ be an $(n-1)$ -variate positive random vector satisfying $x_1 + \dots + x_{n-1} \leq 1$. The probability density function of the Dirichlet distribution with the parameter vector $\nu = (\nu_0, \dots, \nu_{n-1})$ can be written as

$$(1.1) \quad f(X; \nu) = \frac{\Gamma(\nu_0 + \dots + \nu_{n-1})}{\Gamma(\nu_0) \cdots \Gamma(\nu_{n-1})} x_0^{\nu_0-1} \cdots x_{n-1}^{\nu_{n-1}-1},$$

Received October 26, 2006.

2000 *Mathematics Subject Classification.* 53B20, 62B10.

Key words and phrases. Dirichlet manifold, α -geometric structures, foliation, immersion.
This subject is supported partially by the Foundation of China ministry of education.

where $x_0 := 1 - x_1 - \cdots - x_{n-1}$, and the parameters $\nu_i > 0$, $i = 0, 1, \dots, n-1$. The Dirichlet distribution is a multivariate continuous distribution, and it has close relation with gamma distribution. Let $\{Y_i\}_{i=1}^{n-1} \sim \text{Gam}(\nu_i, 1)$ be independent. For $i = 1, \dots, n-1$. Let

$$x_i := \frac{Y_i}{\sum_{i=1}^{n-1} Y_i}$$

$X := (x_1, \dots, x_{n-1})'$, then $X \sim \text{Dirichlet}(\nu)$.

Since the parameters ν_0, \dots, ν_{n-1} of the Dirichlet distribution play the role of the coordinate system, it is obvious that the Dirichlet manifold is n -dimensional. In particular, when $n = 2$, the Dirichlet distribution becomes the Beta distribution, and we have shown all the geometric metrics in [6].

The present paper is organized as follows: Firstly, we show that the Dirichlet distribution is an exponential family distribution and give the dual geometric structures of the Dirichlet manifold. We investigate the α -geometric structure of the Dirichlet manifold. We obtain the Fisher information matrix, α -connections, and α -curvatures. Secondly, we consider the cases of lower dimension. Thirdly, we obtain an affine immersion of the Dirichlet manifold. At last, we give the e -flat hierarchical structures and the orthogonal foliations of the Dirichlet manifold.

2. The geometric structures of the Dirichlet manifold

Definition 2.1. The set

$$(2.1) \quad S = \left\{ f(X; \nu) \mid f(X; \nu) = \frac{\Gamma(\nu_0 + \cdots + \nu_{n-1})}{\Gamma(\nu_0) \cdots \Gamma(\nu_{n-1})} x_0^{\nu_0-1} \cdots x_{n-1}^{\nu_{n-1}-1}, \right. \\ \left. (\nu_0, \dots, \nu_{n-1}) \in \underbrace{R^+ \times \cdots \times R^+}_n \right\}$$

is called the Dirichlet manifold, where $(\nu_0, \dots, \nu_{n-1})$ plays the role of the coordinate system.

Proposition 2.1. *The Dirichlet distribution is an exponential family distribution.*

Proof. The Dirichlet probability density function (1.1) can be rewritten as

$$(2.2) \quad f(X) = \exp \left\{ \nu_0 \log x_0 + \nu_1 \log x_1 + \cdots + \nu_{n-1} \log x_{n-1} \right. \\ \left. + (-\log x_0 - \log x_1 - \cdots - \log x_{n-1}) \right. \\ \left. - \left(\log \Gamma(\nu_0) + \log \Gamma(\nu_1) + \cdots + \log \Gamma(\nu_{n-1}) \right. \right. \\ \left. \left. - \log \Gamma(\nu_0 + \nu_1 + \cdots + \nu_{n-1}) \right) \right\},$$

then, let

$$\begin{aligned}
 & y_1 = \log x_0, \theta_1 = \nu_0, \\
 & y_2 = \log x_1, \theta_2 = \nu_1, \\
 (2.3) \quad & \vdots \\
 & y_n = \log x_{n-1}, \theta_n = \nu_{n-1}, \\
 & M(y) = -(y_1 + y_2 + \dots + y_n),
 \end{aligned}$$

so the potential function $\psi(\theta)$ can be presented as

$$(2.4) \quad \psi(\theta) = \log \Gamma(\theta_1) + \log \Gamma(\theta_2) + \dots + \log \Gamma(\theta_n) - \log \Gamma(\theta_1 + \theta_2 + \dots + \theta_n).$$

Therefore, the Dirichlet probability density function can be denoted by

$$(2.5) \quad f(y) = \exp \left\{ \sum_{i=1}^n \theta_i y_i - \psi(\theta) + M(y) \right\},$$

where $\theta = (\theta_1, \dots, \theta_n) = (\nu_0, \dots, \nu_{n-1})$ is call the natural coordinate system of the Dirichlet manifold, and $M(y)$ is a function merely depending on y .

From (2.5), we know that the Dirichlet distribution is an exponential family distribution. Obviously, the Dirichlet manifold we consider here is n -dimensional. □

Remark. The Dirichlet manifold is ± 1 -flat.

Proposition 2.2. *The Dirichlet manifold has the following dual structures.*

Under the natural coordinate system $\theta = (\theta_1, \dots, \theta_n) = (\nu_0, \dots, \nu_{n-1})$, and the potential function $\psi(\theta)$,

(i) $\eta = (\eta^1, \dots, \eta^n) = (\Psi(\theta_1) - \Psi_1(\theta_1 + \dots + \theta_n), \dots, \Psi(\theta_n) - \Psi_n(\theta_1 + \dots + \theta_n))$, *is the dual coordinate system, and is also called expectation coordinate system, where $\Psi(\theta_i) = \Gamma'(\theta_i)/\Gamma(\theta_i)$ is the digamma function, and $\Psi_i(\theta_1 + \dots + \theta_n) = \Gamma'_i(\theta_1 + \dots + \theta_n)/\Gamma(\theta_1 + \dots + \theta_n), i = 1, \dots, n$.*

(ii) $\phi(\eta) = \sum_{i=1}^n \theta_i \left(\Psi(\theta_i) - \Psi_i(\theta_1 + \dots + \theta_n) \right) - \log \Gamma(\theta_1) - \log \Gamma(\theta_2) - \dots - \log \Gamma(\theta_n) + \log \Gamma(\theta_1 + \theta_2 + \dots + \theta_n)$, *is the dual potential function with respect to the expectation coordinate system.*

Proof. Since the expectation coordinate system can be obtained by

$$\eta^i = \frac{\partial \psi(\theta)}{\partial \theta_i},$$

we get the expectation coordinate system of the Dirichlet manifold by

$$(2.6) \quad (\eta^1, \dots, \eta^n) = \left(\Psi(\theta_1) - \Psi_1(\theta_1 + \dots + \theta_n), \dots, \Psi(\theta_n) - \Psi_n(\theta_1 + \dots + \theta_n) \right).$$

The dual potential function with respect to the expectation coordinate system is obtained by

$$\begin{aligned}
 & \phi(\eta) \\
 &= \sum \theta_i \eta^i - \psi(\theta) \\
 (2.7) \quad &= \sum_{i=1}^n \theta_i \left(\Psi(\theta_i) - \Psi_i(\theta_1 + \dots + \theta_n) \right) - \log \Gamma(\theta_1) \\
 &\quad - \log \Gamma(\theta_2) - \dots - \log \Gamma(\theta_n) + \log \Gamma(\theta_1 + \theta_2 + \dots + \theta_n).
 \end{aligned}$$

This finishes the proof of Proposition 2.2. □

Under the natural coordinate system, we use the potential function with respect to the natural coordinate system to give some simple formulae for the geometric metrics ([1], [2])

$$\begin{aligned}
 & g_{ij}(\theta) = \partial_i \partial_j \psi(\theta), \\
 & T_{ijk}(\theta) = \partial_i \partial_j \partial_k \psi(\theta) = \partial_k(g_{ij}), \\
 (2.8) \quad & \Gamma_{ijk}^{(\alpha)}(\theta) = \frac{1-\alpha}{2} T_{ijk}(\theta), \\
 & R_{ijkl}^{(\alpha)} = \frac{1-\alpha^2}{4} (T_{kmi} T_{jln} - T_{kmj} T_{iln}) g^{mn}.
 \end{aligned}$$

Since $\partial_i \psi(\theta) = \Psi(\theta_i) - \Psi_i(\theta_1 + \dots + \theta_n), i = 1, \dots, n$, combining (2.4) with (2.8), by a calculation, we get the following geometric metrics.

i) when $i = j$, we get

$$(2.9) \quad g_{ii}(\theta) = \partial_i \partial_i \psi(\theta) = \Psi'(\theta_i) - \Psi_{ii}(\theta_1 + \dots + \theta_n), \quad i = 1, \dots, n,$$

ii) when $i \neq j$, we get

$$(2.10) \quad g_{ij}(\theta) = \partial_i \partial_j \psi(\theta) = -\Psi_{ij}(\theta_1 + \dots + \theta_n), \quad i, j = 1, \dots, n,$$

where $\Psi(\theta_i) = \Gamma'(\theta_i)/\Gamma(\theta_i)$ is the digamma function, $\Psi_{ij}(\theta_1 + \dots + \theta_n) = \partial^2 \log \Gamma(\theta_1 + \dots + \theta_n) / \partial \theta_i \partial \theta_j, i, j = 1, \dots, n$, and we denote it by Ψ_{ij} for brevity.

So the Fisher information matrix of the Dirichlet manifold can be written as

$$(2.11) \quad (g_{ij}) = \begin{pmatrix} \Psi'(\theta_1) - \Psi_{11} & -\Psi_{12} & \dots & -\Psi_{1n} \\ -\Psi_{21} & \Psi'(\theta_2) - \Psi_{22} & \dots & -\Psi_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ -\Psi_{n1} & -\Psi_{n2} & \dots & \Psi'(\theta_n) - \Psi_{nn} \end{pmatrix},$$

which is an n -dimensional positive definite matrix.

The cubic tensor T_{ijk} is symmetric and is given by

$$(2.12) \quad T_{iii} = \Psi''(\theta_i) - \Psi_{iii}(\theta_1 + \dots + \theta_n) \quad \text{when } i = j = k,$$

and for other cases

$$(2.13) \quad T_{ijk} = -\Psi_{ijk}(\theta_1 + \dots + \theta_n),$$

where $\Psi_{ijk}(\theta_1 + \dots + \theta_n) = \partial^3 \log \Gamma(\theta_1 + \dots + \theta_n) / \partial \theta_i \partial \theta_j \partial \theta_k$, $i, j, k = 1, \dots, n$, and we denote it by Ψ_{ijk} for brevity.

Then we obtain the α -connections of the Dirichlet manifold as follows:

When $i = j = k$,

$$(2.14) \quad \Gamma_{iii}^{(\alpha)} = \frac{1 - \alpha}{2} \left(\Psi''(\theta_i) - \Psi_{iii}(\theta_1 + \dots + \theta_n) \right), \quad i = 1, \dots, n,$$

and for other cases

$$(2.15) \quad \Gamma_{ijk}^{(\alpha)} = -\frac{1 - \alpha}{2} \Psi_{ijk}(\theta_1 + \dots + \theta_n), \quad i, j, k = 1, \dots, n.$$

In the following, we will consider lower dimensional Dirichlet manifold.

3. The geometric structures of the three dimensional Dirichlet manifold

Remark. When $n = 2$, the Dirichlet distribution changes into

$$(3.1) \quad f(x; \nu) = \frac{\Gamma(\nu_0 + \nu_1)}{\Gamma(\nu_0)\Gamma(\nu_1)} (1 - x)^{\nu_0 - 1} x^{\nu_1 - 1},$$

which is the Beta distribution. The geometric structures of the Beta manifold have been shown in [6].

In this section, we mainly investigate the geometric structures of the 3-dimensional Dirichlet manifold, and the corresponding Dirichlet distribution can be rewritten as

$$(3.2) \quad f(X; \nu) = \frac{\Gamma(\nu_0 + \nu_1 + \nu_2)}{\Gamma(\nu_0)\Gamma(\nu_1)\Gamma(\nu_2)} (1 - x_1 - x_2)^{\nu_0 - 1} x_1^{\nu_1 - 1} x_2^{\nu_2 - 1},$$

where $X = (x_1, x_2)$.

It corresponds to a simplex embedding to a 3-dimensional space (see Figure 1). For $0 < \nu < 1$, multiple modes appear in the corners.

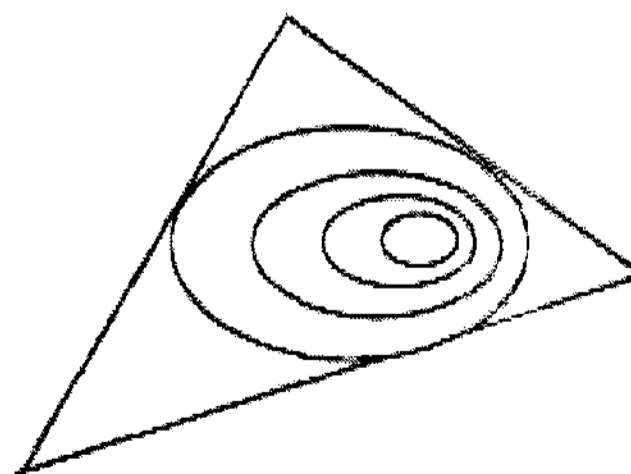


Figure 1. Geometric description of the Dirichlet distribution when $n=3$.

Definition 3.1. The set

$$(3.3) \quad S_3 = \left\{ f(X; \nu) \mid f(X; \nu) = \frac{\Gamma(\nu_0 + \nu_1 + \nu_2)}{\Gamma(\nu_0)\Gamma(\nu_1)\Gamma(\nu_2)} (1 - x_1 - x_2)^{\nu_0-1} x_1^{\nu_1-1} x_2^{\nu_2-1}, \right. \\ \left. (\nu_0, \nu_1, \nu_2) \in R^+ \times R^+ \times R^+ \right\}$$

forms a special 3-dimensional Dirichlet manifold, where (ν_0, ν_1, ν_2) plays the role of the natural coordinate system of S_3 .

From (2.11), we see that the Fisher information matrix of S_3 satisfies

$$(3.4) \quad (g_{ij}) = \begin{pmatrix} \Psi'(\theta_1) - \Psi_{11} & -\Psi_{12} & -\Psi_{13} \\ -\Psi_{12} & \Psi'(\theta_2) - \Psi_{22} & -\Psi_{23} \\ -\Psi_{13} & -\Psi_{23} & \Psi'(\theta_3) - \Psi_{33} \end{pmatrix},$$

and by a complicated calculation, the inverse of the Fisher information matrix (g_{ij}) , (g^{ij}) is given by

$$(3.5) \quad (g^{ij}) = \frac{1}{A} \begin{pmatrix} B(\theta_2, \theta_3) & C(\theta_3, \theta_1, \theta_2) & C(\theta_2, \theta_1, \theta_3) \\ C(\theta_3, \theta_1, \theta_2) & B(\theta_1, \theta_3) & C(\theta_1, \theta_2, \theta_3) \\ C(\theta_2, \theta_1, \theta_3) & C(\theta_1, \theta_2, \theta_3) & B(\theta_1, \theta_2) \end{pmatrix},$$

where

$$A = (\Psi'(\theta_1) - \Psi_{11})(\Psi'(\theta_2) - \Psi_{22})(\Psi'(\theta_3) - \Psi_{33}) - 2\Psi_{12}\Psi_{13}\Psi_{23} \\ + (\Psi_{13})^2(\Psi'(\theta_2) - \Psi_{22}) + (\Psi_{23})^2(\Psi'(\theta_1) - \Psi_{11}) \\ + (\Psi_{12})^2(\Psi'(\theta_3) - \Psi_{33}),$$

$$B(x, y) = \Psi'(x)\Psi'(y) - \Psi'(x)\Psi_{yy} - \Psi'(y)\Psi_{xx} + \Psi_{xx}\Psi_{yy} - (\Psi_{xy})^2,$$

$$C(x, y, z) = \Psi'(x)\Psi_{yz} - \Psi_{xx}\Psi_{yz} + \Psi_{xy}\Psi_{xz}.$$

From (2.12), (2.13), (2.14) and (2.15), we get the cubic tensor and α -connections of S_3 :

$$(3.6) \quad T_{111} = \Psi''(\theta_1) - \Psi_{111}, \quad T_{222} = \Psi''(\theta_2) - \Psi_{222}, \quad T_{333} = \Psi''(\theta_3) - \Psi_{333}, \\ T_{112} = T_{121} = T_{211} = -\Psi_{112}, \quad T_{113} = T_{131} = T_{311} = -\Psi_{113}, \\ T_{122} = T_{212} = T_{221} = -\Psi_{122}, \quad T_{133} = T_{313} = T_{331} = -\Psi_{133}, \\ T_{223} = T_{232} = T_{322} = -\Psi_{223}, \quad T_{233} = T_{323} = T_{322} = -\Psi_{233}, \\ T_{123} = T_{231} = T_{312} = T_{213} = T_{321} = T_{132} = -\Psi_{123},$$

and

$$\begin{aligned} \Gamma_{111}^{(\alpha)} &= \frac{1-\alpha}{2}(\Psi''(\theta_1) - \Psi_{111}), \quad \Gamma_{222}^{(\alpha)} = \frac{1-\alpha}{2}(\Psi''(\theta_2) - \Psi_{222}), \\ \Gamma_{333}^{(\alpha)} &= \frac{1-\alpha}{2}(\Psi''(\theta_3) - \Psi_{333}), \\ \Gamma_{112}^{(\alpha)} &= \Gamma_{121}^{(\alpha)} = \Gamma_{211}^{(\alpha)} = -\frac{1-\alpha}{2}\Psi_{112}, \quad \Gamma_{113}^{(\alpha)} = \Gamma_{131}^{(\alpha)} = \Gamma_{311}^{(\alpha)} = -\frac{1-\alpha}{2}\Psi_{113}, \\ \Gamma_{122}^{(\alpha)} &= \Gamma_{212}^{(\alpha)} = \Gamma_{221}^{(\alpha)} = -\frac{1-\alpha}{2}\Psi_{122}, \quad \Gamma_{133}^{(\alpha)} = \Gamma_{313}^{(\alpha)} = \Gamma_{331}^{(\alpha)} = -\frac{1-\alpha}{2}\Psi_{133}, \\ \Gamma_{223}^{(\alpha)} &= \Gamma_{232}^{(\alpha)} = \Gamma_{322}^{(\alpha)} = -\frac{1-\alpha}{2}\Psi_{223}, \quad \Gamma_{233}^{(\alpha)} = \Gamma_{323}^{(\alpha)} = \Gamma_{322}^{(\alpha)} = -\frac{1-\alpha}{2}\Psi_{233}, \\ \Gamma_{123}^{(\alpha)} &= \Gamma_{231}^{(\alpha)} = \Gamma_{312}^{(\alpha)} = \Gamma_{213}^{(\alpha)} = \Gamma_{321}^{(\alpha)} = \Gamma_{132}^{(\alpha)} = -\frac{1-\alpha}{2}\Psi_{123}. \end{aligned}$$

Combing (2.8), (3.5) with (3.6), we get the nonzero components of the curvature tensor

$$\begin{aligned} R_{1212}^{(\alpha)} &= \frac{1-\alpha^2}{4A} \left[(-\Psi''(\theta_1)\Psi_{122} + \Psi_{111}\Psi_{122} - (\Psi_{112})^2)B(\theta_2, \theta_3) + (\Psi''(\theta_1)\Psi''(\theta_2) \right. \\ &\quad - \Psi''(\theta_1)\Psi_{222} - \Psi''(\theta_2)\Psi_{111} + \Psi_{111}\Psi_{222} - \Psi_{112}\Psi_{122})C(\theta_3, \theta_1, \theta_2) \\ &\quad + (-\Psi''(\theta_1)\Psi_{223} + \Psi_{111}\Psi_{223} + \Psi_{113}\Psi_{122} - 2\Psi_{112}\Psi_{123})C(\theta_2, \theta_1, \theta_3) \\ &\quad + (-\Psi''(\theta_2)\Psi_{112} + \Psi_{222}\Psi_{112} - (\Psi_{122})^2)B(\theta_1, \theta_3) + (\Psi_{112}\Psi_{223} - \Psi''(\theta_2)\Psi_{113} \\ &\quad \left. + \Psi_{113}\Psi_{222} - 2\Psi_{123}\Psi_{122})C(\theta_1, \theta_2, \theta_3) + (\Psi_{113}\Psi_{223} - (\Psi_{123})^2)B(\theta_1, \theta_2) \right], \\ R_{1313}^{(\alpha)} &= \frac{1-\alpha^2}{4A} \left[(-\Psi''(\theta_1)\Psi_{133} + \Psi_{111}\Psi_{133} - (\Psi_{113})^2)B(\theta_2, \theta_3) \right. \\ &\quad + (-\Psi''(\theta_1)\Psi_{233} + \Psi_{111}\Psi_{233} + \Psi_{112}\Psi_{133} - 2\Psi_{113}\Psi_{123})C(\theta_3, \theta_1, \theta_2) \\ &\quad + (\Psi''(\theta_1)\Psi''(\theta_3) - \Psi''(\theta_1)\Psi_{333} - \Psi''(\theta_3)\Psi_{111} + \Psi_{111}\Psi_{333} - \Psi_{113}\Psi_{133}) \\ &\quad \cdot C(\theta_2, \theta_1, \theta_3) + (\Psi_{112}\Psi_{233} - (\Psi_{123})^2)B(\theta_1, \theta_3) \\ &\quad + (\Psi_{113}\Psi_{233} - \Psi''(\theta_3)\Psi_{112} + \Psi_{112}\Psi_{333} - 2(\Psi_{123})^2)C(\theta_1, \theta_2, \theta_3) \\ &\quad \left. + (-\Psi''(\theta_3)\Psi_{113} + \Psi_{333}\Psi_{113} - (\Psi_{133})^2)B(\theta_1, \theta_2) \right], \\ R_{2323}^{(\alpha)} &= \frac{1-\alpha^2}{4A} \left[(\Psi_{122}\Psi_{133} - (\Psi_{123})^2)B(\theta_2, \theta_3) \right. \\ &\quad + (-\Psi''(\theta_2)\Psi_{133} + \Psi_{222}\Psi_{133} + \Psi_{122}\Psi_{233} - 2\Psi_{223}\Psi_{123})C(\theta_3, \theta_1, \theta_2) \\ &\quad + (\Psi_{233}\Psi_{133} - \Psi''(\theta_3)\Psi_{122} + \Psi_{122}\Psi_{333} - 2\Psi_{123}\Psi_{233})C(\theta_2, \theta_1, \theta_3) \\ &\quad + (-\Psi''(\theta_2)\Psi_{233} + \Psi_{222}\Psi_{233} - (\Psi_{223})^2)B(\theta_1, \theta_3) \\ &\quad + (\Psi''(\theta_2)\Psi''(\theta_3) - \Psi''(\theta_2)\Psi_{333} - \Psi''(\theta_3)\Psi_{222} + \Psi_{222}\Psi_{333} - \Psi_{223}\Psi_{233}) \\ &\quad \cdot C(\theta_1, \theta_2, \theta_3) + (-\Psi''(\theta_3)\Psi_{223} + \Psi_{333}\Psi_{223} - (\Psi_{233})^2)B(\theta_1, \theta_2) \left. \right], \\ R_{2113}^{(\alpha)} &= \frac{1-\alpha^2}{4A} \left[(\Psi_{112}\Psi_{113} + \Psi''(\theta_1)\Psi_{123} - \Psi_{111}\Psi_{123})B(\theta_2, \theta_3) + (\Psi_{122}\Psi_{113} + \Psi''(\theta_1)\Psi_{223} \right. \\ &\quad - \Psi_{111}\Psi_{223})C(\theta_3, \theta_1, \theta_2) + (\Psi_{112}\Psi_{133} + \Psi''(\theta_1)\Psi_{233} - \Psi_{111}\Psi_{233})C(\theta_2, \theta_1, \theta_3) \\ &\quad + (\Psi_{122}\Psi_{123} - \Psi_{112}\Psi_{223})B(\theta_1, \theta_3) \\ &\quad + (\Psi_{122}\Psi_{133} + (\Psi_{123})^2 - \Psi_{112}\Psi_{233} - \Psi_{113}\Psi_{233})C(\theta_1, \theta_2, \theta_3) \\ &\quad \left. + (\Psi_{133}\Psi_{123} - \Psi_{113}\Psi_{233})B(\theta_1, \theta_2) \right], \end{aligned}$$

$$\begin{aligned}
R_{1223}^{(\alpha)} &= \frac{1-\alpha^2}{4A} \left[(\Psi_{112}\Psi_{123} - \Psi_{122}\Psi_{113})B(\theta_2, \theta_3) + (\Psi_{112}\Psi_{223} + \Psi''(\theta_2)\Psi_{113} - \Psi_{222}\Psi_{113}) \right. \\
&\quad \times C(\theta_3, \theta_1, \theta_2) + (\Psi_{112}\Psi_{233} + (\Psi_{123})^2 - \Psi_{122}\Psi_{133} - \Psi_{113}\Psi_{223})C(\theta_2, \theta_1, \theta_3) \\
&\quad + (\Psi_{122}\Psi_{223} + \Psi''(\theta_2)\Psi_{123} - \Psi_{222}\Psi_{123})B(\theta_1, \theta_3) \\
&\quad + (\Psi_{122}\Psi_{233} + \Psi''(\theta_2)\Psi_{133} - \Psi_{222}\Psi_{133})C(\theta_1, \theta_2, \theta_3) \\
&\quad \left. + (\Psi_{123}\Psi_{233} - \Psi_{223}\Psi_{133})B(\theta_1, \theta_2) \right], \\
R_{1332}^{(\alpha)} &= \frac{1-\alpha^2}{4A} \left[(\Psi_{113}\Psi_{123} - \Psi_{133}\Psi_{112})B(\theta_2, \theta_3) + (\Psi_{113}\Psi_{223} + (\Psi_{123})^2 - \Psi_{133}\Psi_{122} \right. \\
&\quad - \Psi_{233}\Psi_{112})C(\theta_3, \theta_1, \theta_2) + (\Psi_{113}\Psi_{233} + \Psi''(\theta_3)\Psi_{112} - \Psi_{112}\Psi_{333})C(\theta_2, \theta_1, \theta_3) \\
&\quad + (\Psi_{123}\Psi_{223} - \Psi_{233}\Psi_{122})B(\theta_1, \theta_3) + (\Psi_{133}\Psi_{223} + \Psi''(\theta_3)\Psi_{122} - \Psi_{333}\Psi_{122}) \\
&\quad \left. \times C(\theta_1, \theta_2, \theta_3) + (\Psi_{133}\Psi_{233} + \Psi''(\theta_3)\Psi_{123} - \Psi_{333}\Psi_{123})B(\theta_1, \theta_2) \right].
\end{aligned}$$

So the α -sectional curvatures of the Dirichlet manifold are given by

$$\begin{aligned}
K_{1212}^{(\alpha)} &= \frac{1-\alpha^2}{4AB(\theta_1, \theta_2)} \left[(-\Psi''(\theta_1)\Psi_{122} + \Psi_{111}\Psi_{122} - (\Psi_{112})^2)B(\theta_2, \theta_3) \right. \\
&\quad + (\Psi''(\theta_1)\Psi''(\theta_2) - \Psi''(\theta_1)\Psi_{222} - \Psi''(\theta_2)\Psi_{111} + \Psi_{111}\Psi_{222} - \Psi_{112}\Psi_{122})C(\theta_3, \theta_1, \theta_2) \\
&\quad + (-\Psi''(\theta_1)\Psi_{223} + \Psi_{111}\Psi_{223} + \Psi_{113}\Psi_{122} - 2\Psi_{112}\Psi_{123})C(\theta_2, \theta_1, \theta_3) \\
&\quad + (-\Psi''(\theta_2)\Psi_{112} + \Psi_{222}\Psi_{112} - (\Psi_{122})^2)B(\theta_1, \theta_3) \\
&\quad + (\Psi_{112}\Psi_{223} - \Psi''(\theta_2)\Psi_{113} + \Psi_{113}\Psi_{222} - 2\Psi_{123}\Psi_{122})C(\theta_1, \theta_2, \theta_3) \\
&\quad \left. + (\Psi_{113}\Psi_{223} - (\Psi_{123})^2)B(\theta_1, \theta_2) \right], \\
K_{1313}^{(\alpha)} &= \frac{1-\alpha^2}{4AB(\theta_1, \theta_3)} \left[(-\Psi''(\theta_1)\Psi_{133} + \Psi_{111}\Psi_{133} - (\Psi_{113})^2)B(\theta_2, \theta_3) \right. \\
&\quad + (-\Psi''(\theta_1)\Psi_{233} + \Psi_{111}\Psi_{233} + \Psi_{112}\Psi_{133} - 2\Psi_{113}\Psi_{123})C(\theta_3, \theta_1, \theta_2) \\
&\quad + (\Psi''(\theta_1)\Psi''(\theta_3) - \Psi''(\theta_1)\Psi_{333} - \Psi''(\theta_3)\Psi_{111} + \Psi_{111}\Psi_{333} - \Psi_{113}\Psi_{133})C(\theta_2, \theta_1, \theta_3) \\
&\quad + (\Psi_{112}\Psi_{233} - (\Psi_{123})^2)B(\theta_1, \theta_3) \\
&\quad + (\Psi_{113}\Psi_{233} - \Psi''(\theta_3)\Psi_{112} + \Psi_{112}\Psi_{333} - 2(\Psi_{123})^2)C(\theta_1, \theta_2, \theta_3) \\
&\quad \left. + (-\Psi''(\theta_3)\Psi_{113} + \Psi_{333}\Psi_{113} - (\Psi_{133})^2)B(\theta_1, \theta_2) \right], \\
K_{2323}^{(\alpha)} &= \frac{1-\alpha^2}{4AB(\theta_2, \theta_3)} \left[(\Psi_{122}\Psi_{133} - (\Psi_{123})^2)B(\theta_2, \theta_3) \right. \\
&\quad + (-\Psi''(\theta_2)\Psi_{133} + \Psi_{222}\Psi_{133} + \Psi_{122}\Psi_{233} - 2\Psi_{223}\Psi_{123})C(\theta_3, \theta_1, \theta_2) \\
&\quad + (\Psi_{233}\Psi_{133} - \Psi''(\theta_3)\Psi_{122} + \Psi_{122}\Psi_{333} - 2\Psi_{123}\Psi_{233})C(\theta_2, \theta_1, \theta_3) \\
&\quad + (-\Psi''(\theta_2)\Psi_{233} + \Psi_{222}\Psi_{233} - (\Psi_{223})^2)B(\theta_1, \theta_3) \\
&\quad + (\Psi''(\theta_2)\Psi''(\theta_3) - \Psi''(\theta_2)\Psi_{333} - \Psi''(\theta_3)\Psi_{222} + \Psi_{222}\Psi_{333} - \Psi_{223}\Psi_{233})C(\theta_1, \theta_2, \theta_3) \\
&\quad \left. + (-\Psi''(\theta_3)\Psi_{223} + \Psi_{333}\Psi_{223} - (\Psi_{233})^2)B(\theta_1, \theta_2) \right].
\end{aligned}$$

Using the formulae $R_{ij} = R_{ikjl}g^{kl}$, and $R = R_{ikjl}g^{kl}g^{ij}$, we can obtain the Ricci curvatures and the mean curvature of S_3 . We omit them here.

4. Affine immersion

Let M be an m -dimensional manifold, f be an immersion from M to \mathbb{R}^{m+1} , and ξ be a vector field along f . For arbitrary $x \in \mathbb{R}^{m+1}$, we identify $T_x \mathbb{R}^{m+1} \cong \mathbb{R}^{m+1}$. The pair $\{f, \xi\}$ is said to be an affine immersion from M to \mathbb{R}^{m+1} if, for each point $P \in M$, the following formula holds

$$T_{f(P)} \mathbb{R}^{m+1} = f_*(T_P M) \oplus \text{span } \xi_P,$$

where ξ is called a transversal vector field.

Denoting by D the standard flat affine connection of \mathbb{R}^{m+1} , we have the following decompositions

$$\begin{aligned} D_X f_* Y &= f_*(\nabla_X Y) + h(X, Y)\xi, \\ D_X \xi &= -f_*(Sh(X)) + \tau(X)\xi. \end{aligned}$$

The induced objects ∇ , h , Sh and τ are the induced connection, the affine fundamental form, the affine shape operator and the transversal connection form, respectively.

Therefore, we have the following

Theorem 4.1. *Since (S, h, ∇, ∇^*) is the n -dimensional Dirichlet manifold, it is dually flat space with a global coordinate system. θ is an affine coordinate system of ∇ , and ψ is a θ -potential function. Then the Dirichlet manifold (S, h, ∇) can be immersed in \mathbb{R}^{n+1} by the following way*

$$f : S \rightarrow \mathbb{R}^{n+1} \left(\begin{array}{c} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_n \end{array} \right) \rightarrow \left(\begin{array}{c} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_n \\ \log \Gamma(\theta_1) + \log \Gamma(\theta_2) + \dots + \log \Gamma(\theta_n) - \log \Gamma(\theta_1 + \theta_2 + \dots + \theta_n) \end{array} \right),$$

which is called a graph immersion from S into \mathbb{R}^{n+1} , and the transversal vector $\xi = (\underbrace{0, \dots, 0}_n, 1)'$.

5. The hierarchical structures and the orthogonal foliations of the Dirichlet manifold

Denoting by

$$S_k = \left\{ f(X; \nu) \mid f(X; \nu) = \frac{\Gamma(\nu_0 + \dots + \nu_{k-1})}{\Gamma(\nu_0) \dots \Gamma(\nu_{k-1})} x_0^{\nu_0-1} \dots x_{k-1}^{\nu_{k-1}-1}, \right. \\ \left. (\nu_0, \dots, \nu_{k-1}) \in \underbrace{R^+ \times \dots \times R^+}_k \right\}$$

a k -dimensional Dirichlet manifold, where $(\nu_0, \dots, \nu_{k-1})$ plays the role of the natural coordinate system, and $k \leq n$, then we have the following theorem.

Theorem 5.1. S_k is an e -flat submanifold of the dual flat manifold $S(S = S_n)$.

Proof. Let $(\nu_0, \dots, \nu_{k-1})$ be a coordinate system of S_k , and from Proposition 2.2, we see that $\theta = (\theta_1, \dots, \theta_n) = (\nu_0, \dots, \nu_{k-1}, \nu_k, \dots, \nu_{n-1})$ is an e -affine coordinate system of S satisfying $k \leq n$.

From the statistical manifold S_k to the manifold S , we define the following map

$$\begin{aligned} \sigma : S_k &\rightarrow S \\ (\nu_0, \dots, \nu_{k-1}) &\mapsto (\nu_0, \dots, \nu_{k-1}, \underbrace{0, \dots, 0}_{n-k}), \end{aligned}$$

then σ is a continuous smooth map from S_k to S . So we can see that S_k is an immersed submanifold of S .

Since S is ± 1 -flat, and

$$\begin{aligned} \nu_0 &= \nu_0 + 0\nu_1 + \dots + 0\nu_{n-1}, \\ \nu_1 &= \nu_1 + 0\nu_1 + 0\nu_2 + \dots + 0\nu_{n-1}, \\ &\vdots \\ \nu_{k-1} &= \nu_{k-1} + 0\nu_1 + \dots + 0\nu_{k-2} + 0\nu_k + \dots + 0\nu_{n-1}, \end{aligned}$$

that is, S_k can be written as a linear subspace in the e -affine coordinate system θ of S .

From the conclusion in [1], we obtain that S_k is an e -flat submanifold of S .

This completes the proof of the Theorem 5.1. \square

Remark. From the result of the Theorem 5.1, we can see that S_k is an e -flat submanifold of S_{k+1} .

Definition 5.1. A nested series of e -flat submanifolds

$$S_1 \subset S_2 \subset \dots \subset S_{n-1} \subset S_n = S$$

is called an e -flat hierarchical structure, or, shortly, the e -structure, where every S_k is an e -flat submanifold of S_{k+1} .

More details about e -hierarchical structure see [2].

Now let us introduce a new coordinate system of the Dirichlet manifold, which is called mixed coordinate system. In this case, the corporate in the n -dimensional Dirichlet manifold can be represented as

$$\xi_k = (\eta_{k-}, \theta_{k+}) = (\eta_1, \dots, \eta_k, \theta_{k+1}, \dots, \theta_n),$$

where $\eta_{k-} = (\eta_1, \dots, \eta_k)$, and $\theta_{k+} = (\theta_{k+1}, \dots, \theta_n)$.

We define two subsets $E_k(a_{k+})$, and $M_k(b_{k-})$ of the Dirichlet manifold as follows

$$\begin{aligned} E_k(a_{k+}) &= \{p(x; \theta) | \theta_{k+} = a_{k+}\}, \\ M_k(b_{k-}) &= \{p(x; \eta) | \eta_{k-} = b_{k-}\}, \end{aligned}$$

where $a_{k+} = (a_{k+1}, \dots, a_n)$, and $b_{k-} = (b_1, \dots, b_k)$, and $a_{k+1}, \dots, a_n, b_1, \dots, b_k$ are all constants.

Theorem 5.2. *A nested series*

$$E_1(0) \subset E_2(0) \subset \cdots \subset E_{n-1}(0) \subset E_n(0) = S$$

*is an e-flat hierarchical structure of the Dirichlet manifold, where $a_{k+} = 0$;
A nested series*

$$M_{n-1}(0) \subset M_{n-2}(0) \subset \cdots \subset M_1(0) \subset M_0(0) = S$$

is an m-flat hierarchical structure of the Dirichlet manifold, where $b_{k-} = 0$.

Proof. From the definition of $E_k(a_{k+})$, we know that $E_k(a_{k+})$ is a submanifold of $E_{k+1}(a_{(k+1)+})$ with the fixed parameter θ_{k+1} . Here, we take $a_{k+} = 0$, and it is clear that $E_n(0) = S$. Since S is in the form of (2.1), it belongs to an exponential distribution family, and it is ± 1 -flat. For $E_{n-1}(0)$ can be written as a linear submanifold in the e -affine coordinates θ of $E_n(0)$, $E_{n-1}(0)$ is an e -flat submanifold of $E_n(0)$, and $E_{n-1}(0)$ is automatically dually flat. Similarly, we find that $E_k(0)$ is an e -flat submanifold of $E_{k+1}(0)$, and $E_k(0)$ is automatically dually flat. So we see that the nested series

$$E_1(0) \subset E_2(0) \subset \cdots \subset E_{n-1}(0) \subset E_n(0) = S$$

is an e -flat hierarchical structure of the Dirichlet manifold.

The proof of the m -flat hierarchical structure of the Dirichlet manifold is similar with the above, we omit it here.

This finishes the proof of the Theorem 5.2. □

Proposition 5.1. *The submanifolds M_k and E_k of the Dirichlet manifold are orthogonal at any point.*

Proof. Let e_I be the tangent vector along the natural coordinate curve θ_I of the E_k . The tangent space at a point in E_k , is a linear space spanned by these tangent vectors, namely, $\{e_1, e_2, \dots, e_k\}$ forms a basis of the tangent space of E_k . Let e^J be the tangent vector along the expectional coordinate curve η of the M_k . The tangent space at a point in M_k , is a linear space spanned by these tangent vectors, namely, $\{e^{k+1}, \dots, e^n\}$ forms a basis of the tangent space of M_k .

So the inner product of the tangent vectors e^J and e_I at any point $\xi_k = (\eta_{k-}, \theta_{K+}) = (\eta_1, \dots, \eta_k, \theta_{k+1}, \dots, \theta_n)$ in the n -dimensional Dirichlet manifold is

$$e_I \cdot e^J = 0,$$

and hence E_k and M_k are orthogonal at ξ_k .

This completes the proof of the Proposition 5.1. □

Thus, E_k and M_k give the orthogonal foliations of the Dirichlet manifold.

From Definition 5.1, we know that S_k is an e -flat submanifold of S , and is dual flat automatically. Then any point p in S can be projected to S_k through the dual geodesic. The projection p^* is the unique point to make the distance

from p to S_k smallest in the sense of the Kullback divergence([1]). p^* can be denoted by

$$p^* = \prod_{S_k} p.$$

In information theory, the information between any fixed point p_0 in S_k and p can be decomposed into two parts under the help of the Pythagoras theorem ([2]), that is,

$$D[p, p_0] = D[p, p^*] + D[p^*, p_0].$$

Here, $D[p, p^*]$ is regarded as the information amount representing effects of p higher than k , whereas $D[p^*, p_0]$ is regarded as the information amount representing effects of p not higher than k .

Acknowledgement. The authors would like to thank Professor Hiroshi Matsuzoe for his valuable suggestions. The second author would like to thank Professor Anmin Li for his pure-hearted encouragement.

References

- [1] S. Amari, *Differential-Geometrical Methods in Statistics*, Lecture Notes in Statistics, 28. Springer-Verlag, New York, 1985.
- [2] ———, *Information geometry on hierarchy of probability distributions*, IEEE Trans. Inform. Theory **47** (2001), no. 5, 1701–1711.
- [3] S. Amari and H. Nagaoka, *Methods of Information Geometry*, Translated from the 1993 Japanese original by Daishi Harada. Translations of Mathematical Monographs, 191. American Mathematical Society, Providence, RI; Oxford University Press, Oxford, 2000.
- [4] C. T. J Dodson and H. Matsuzoe, *An affine embedding of the gamma manifold*, Appl. Sci. **5** (2003), no. 1, 7–12.
- [5] A. M. Li, U. Simon, and G. S. Zhao, *Global Affine Differential Geometry of Hypersurfaces*, de Gruyter Expositions in Mathematics, 11. Walter de Gruyter & Co., Berlin, 1993.
- [6] F. Zhong, H. Sun, and Z. Zhang, *Information geometry of the Beta distribution*, preprint.

FENGWEI ZHONG
DEPARTMENT OF MATHEMATICS
BEIJING INSTITUTE OF TECHNOLOGY
BEIJING 100081, CHINA
E-mail address: killowind@sina.com

HUAFEI SUN
DEPARTMENT OF MATHEMATICS
BEIJING INSTITUTE OF TECHNOLOGY
BEIJING 100081, CHINA
E-mail address: sunhuafei@263.net

ZHENNING ZHANG
DEPARTMENT OF MATHEMATICS
BEIJING INSTITUTE OF TECHNOLOGY
BEIJING 100081, CHINA
E-mail address: ningning0327@163.com