# The geometric structure of the Pareto distribution *

Linyu Peng, Huafei Sun and Lin Jiu
Beijing Institute of Technology

**Abstract**

In this paper, we characterize the Pareto manifold from the viewpoint of information geometry and give the Ricci curvatures, the Gaussian curvature, the Kullback divergence, the J-divergence and the geodesic equations. Also, some examples on the application of the Pareto distribution are provided.

**Resumen**

En este artículo, caracterizamos la variedad de Pareto desde el punto de vista de la geometría de la información y calculamos las curvaturas de Ricci, la curvatura Gaussiana, la divergencia de Kullback, la J-divergencia y las ecuaciones de las geodésicas. También damos algunas aplicaciones de la distribución de Pareto.

**Key words.** The Pareto distribution, Ricci curvature, Gaussian curvature, divergence

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

## 1 Introduction

It is well known that information geometry has been widely applied into various fields, such as statistical inference, system control and neural network. Recently, scholars studied the statistical manifolds from the viewpoint of information geometry, and using the geometric metrics gave a new description of the statistical distribution. Here, the parameters of the probability density function play an important role in statistical manifold and can be regarded as the coordinate system. In [1], Amari studied the exponential distribution families and Dodson ([4]) and his colleagues investigated some special exponential distributions such as the bivariate normal distribution, the Gamma distribution,

*The subject is partially supported by the Foundation of China Education Ministry.

the Mckay distribution and the Frund distribution and gave their geometric structures. It is clear that the exponential distribution shows a "lovely" form with the potential function, by using which the calculation becomes more convenient. The Pareto model is very famous, but the Pareto distribution can not be written as a form with the potential function, which means that the calculation becomes difficult. In this paper, we study the geometric structure of the Pareto distribution.

## 2 Preliminaries.

**Definition 2.1.** For a density function $f(x;\theta)$, where $\theta = (\theta^1, \theta^2, \ldots, \theta^n)$, the function $l(x;\theta)$ is defined by

$$l(x;\theta) = \ln f(x;\theta). \tag{2.1}$$

**Definition 2.2.** We call $M = \{l(x;\theta) | (\theta^1, \theta^2, \ldots, \theta^n) \in R^n\}$ an $n$-dimensional distribution manifold, where $(\theta^1, \theta^2, \ldots, \theta^n)$ plays the role of the coordinate system.

**Definition 2.3.** The Fisher information matrix $(g_{ij})$ is defined by

$$(g_{ij}) = (E_\theta[\partial_i l \quad \partial_j l]), \quad i,j = 1,2,\ldots,n, \tag{2.2}$$

where

$$\partial_i l = \frac{\partial}{\partial \theta^i} l(x;\theta), \quad i = 1,2,\ldots,n.$$

Then we get the inverse matrix of $(g_{ij})$

$$(g^{ij}) = (g_{ij})^{-1}, \quad i,j = 1,2,\ldots,n. \tag{2.3}$$

**Definition 2.4.** The Riemannian connection $\Gamma_{ijk}$ are defined by

$$\Gamma_{ijk} = \frac{1}{2}(\partial_i g_{jk} + \partial_j g_{ki} - \partial_k g_{ij}), \quad i,j,k = 1,2,\ldots,n \tag{2.4}$$

and the $\alpha$-connection are defined by

$$\Gamma^{(\alpha)}_{ijk} = \Gamma_{ijk} - \frac{\alpha}{2} T_{ijk}, \quad i,j,k = 1,2,\ldots,n, \tag{2.5}$$

where

$$T_{ijk} = E[\partial_i l \quad \partial_j l \quad \partial_k l], \quad i,j,k = 1,2,\ldots,n. \tag{2.6}$$

**Definition 2.5.** Under the $\theta$ coordinate system, the $\alpha$-curvature tensors $R^{(\alpha)}_{ijkl}$ are defined by

$$R^{(\alpha)}_{ijkl} = \left(\partial_j \Gamma^{s(\alpha)}_{ik} - \partial_i \Gamma^{s(\alpha)}_{jk}\right) g_{sl} + \left(\Gamma^{(\alpha)}_{jtl} \Gamma^{t(\alpha)}_{ik} - \Gamma^{(\alpha)}_{itl} \Gamma^{t(\alpha)}_{jk}\right), \quad i,j,k,l,s,t = 1,2,\ldots,n, \tag{2.7}$$

where

$$\Gamma_{ij}^{k(\alpha)} = \Gamma_{ijs}^{(\alpha)} g^{sk}, \quad i,j,k,s = 1,2,\ldots,n. \tag{2.8}$$

**Definition 2.6.** The $\alpha$-Ricci curvatures $R_{ik}^{(\alpha)}$ are given by

$$R_{ik}^{(\alpha)} = R_{ijkl}^{(\alpha)} g^{jl}, \quad i,j,k,l = 1,2,\ldots,n. \tag{2.9}$$

**Definition 2.7.** The $\alpha$-sectional curvatures $K_{ijij}^{(\alpha)}$ are defined by

$$K_{ijij}^{(\alpha)} = \frac{R_{ijij}^{(\alpha)}}{g_{ii}g_{jj} - (g_{ij})^2}, \quad i,j = 1,2,\ldots,n. \tag{2.10}$$

Specially, when $n = 2$, the $\alpha$-sectional curvature $K_{1212}^{(\alpha)} = K^{(\alpha)}$ is called the $\alpha$-Gaussian curvature and

$$K^{(\alpha)} = \frac{R_{1212}^{(\alpha)}}{\det(g_{ij})}. \tag{2.11}$$

**Definition 2.8.** Assume $p(x;\theta_p)$ and $q(x;\theta_q)$ are two points on the manifold $M$, the Kullback divergence $K(p,q)$ is defined by

$$\begin{aligned} K(p,q) &= E_{\theta_p}[\ln \frac{p(x;\theta_p)}{q(x;\theta_q)}] \\ &= \int p(x;\theta_p) \ln \frac{p(x;\theta_p)}{q(x;\theta_q)} dx \end{aligned} \tag{2.12}$$

and the J-divergence is defined by

$$J(p,q) = \int (p(x;\theta_p) - q(x;\theta_q)) \ln \frac{p(x;\theta_p)}{q(x;\theta_q)} dx. \tag{2.13}$$

When the two points $p(x;\theta_p)$ and $q(x;\theta_q)$ are close enough, from the Taylor's formula, one can see that

$$K(\theta, \theta + d\theta) = \frac{1}{2} ds^2$$

and

$$J(\theta, \theta + d\theta) = ds^2.$$

**Definition 2.9.** The geodesic equations of manifold $M$ with coordinate $\theta = (\theta^1, \theta^2, \ldots, \theta^n)$ are defined by

$$\frac{d^2\theta^k}{dt^2} + \Gamma_{ij}^k \frac{d\theta^i}{dt} \frac{d\theta^j}{dt} = 0. \tag{2.14}$$

## 3 The Pareto manifold.

The set

$$\left\{p(x;\theta)|p(x;\theta) = \lambda\mu^{\lambda}x^{-\lambda-1}, \theta = (\theta^1,\theta^2) = (\lambda,\mu), x > \mu, \lambda > 0, \mu > 0\right\}$$

is called the Pareto manifold, where

$$p(x;\theta) = \lambda\mu^{\lambda}x^{-\lambda-1}, x > \mu, \lambda > 0, \mu > 0$$

is the probability density function of the Pareto distribution.

**Proposition 3.1.** The nonzero component of the $\alpha$-curvature tensor is given by

$$R^{(\alpha)}_{1212} = \frac{(\lambda^3-\lambda)\alpha^2 + (-2\lambda^3 + 2\lambda^2 - 3\lambda + 2)\alpha + 3\lambda - 2}{4\lambda\mu^2(1-\lambda)^2}.$$

**Proof.** Defining

$$\begin{aligned} l(x;\theta) &= \ln p(x,\theta) \\ &= \ln\lambda + \lambda\ln\mu - (\lambda+1)\ln x, \end{aligned}$$

then we see that

$$\partial_1 l = \frac{1}{\lambda} + \ln\mu - \ln x,\ \partial_2 l = \frac{\lambda}{\mu}$$

and

$$\partial_1\partial_1 l = -\frac{1}{\lambda^2},\ \partial_1\partial_2 l = \partial_2\partial_1 l = \frac{1}{\mu},\ \partial_2\partial_2 l = -\frac{\lambda}{\mu^2}.$$

From (2.2), we get the Fisher information matrix

$$(g_{ij}) = \begin{pmatrix} \frac{1}{\lambda^2} & -\frac{1}{\mu} \\ -\frac{1}{\mu} & \frac{\lambda}{\mu^2} \end{pmatrix}$$

and

$$\det(g_{ij}) = \frac{1-\lambda}{\mu^2\lambda}.$$

Thus the square of the arc length is given by

$$(ds)^2 = \frac{1}{\lambda^2}(d\lambda)^2 - \frac{2}{\mu}d\lambda d\mu + \frac{1}{\mu^2}(d\mu)^2.$$

The inverse matrix of $(g_{ij})$ is given by

$$(g^{ij}) = \begin{pmatrix} \frac{\lambda^2}{1-\lambda} & \frac{\lambda\mu}{1-\lambda} \\ \frac{\lambda\mu}{1-\lambda} & \frac{\mu^2}{\lambda(1-\lambda)} \end{pmatrix}.$$

From (2.6), we can get

$$T_{111} = -\frac{2}{\lambda^3}, \quad T_{121} = T_{211} = T_{112} = \frac{1}{\lambda\mu}, \tag{3.1}$$

$$T_{221} = T_{212} = T_{122} = 0, \quad T_{222} = \frac{\lambda^3}{\mu^3}. \tag{3.2}$$

And from (2.4), we can get

$$\Gamma_{111} = -\frac{1}{\lambda^3}, \quad \Gamma_{112} = \Gamma_{121} = \Gamma_{211} = 0, \tag{3.3}$$

$$\Gamma_{122} = \Gamma_{212} = \Gamma_{221} = \frac{1}{2\mu^2}, \quad \Gamma_{222} = -\frac{\lambda}{\mu^3}. \tag{3.4}$$

Then from (2.5), (3.1), (3.2), (3.3) and (3.4), we can get

$$\Gamma^{(\alpha)}_{111} = \frac{\alpha-1}{\lambda^3}, \quad \Gamma^{(\alpha)}_{112} = \Gamma^{(\alpha)}_{121} = \Gamma^{(\alpha)}_{211} = -\frac{\alpha}{2\lambda\mu}, \tag{3.5}$$

$$\Gamma^{(\alpha)}_{212} = \Gamma^{(\alpha)}_{122} = \Gamma^{(\alpha)}_{221} = \frac{1}{2\mu^2}, \quad \Gamma^{(\alpha)}_{222} = -\frac{2\lambda+\lambda^3\alpha}{2\mu^3}. \tag{3.6}$$

From (2.8), (3.5) and (3.6), we get

$$\Gamma^{1(\alpha)}_{11} = \frac{2\alpha-\alpha\lambda-2}{2\lambda(1-\lambda)}, \quad \Gamma^{2(\alpha)}_{11} = \frac{\mu(\alpha-2)}{2\lambda^2(1-\lambda)}, \tag{3.7}$$

$$\Gamma^{1(\alpha)}_{12} = \frac{\lambda-\alpha\lambda}{2\mu(1-\lambda)}, \quad \Gamma^{2(\alpha)}_{12} = \frac{1-\alpha\lambda}{2\lambda(1-\lambda)}, \quad \Gamma^{1(\alpha)}_{21} = \frac{\lambda-\alpha\lambda}{2\mu(1-\lambda)} \tag{3.8}$$

and

$$\Gamma^{2(\alpha)}_{21} = \frac{1-\alpha\lambda}{2\lambda(1-\lambda)}, \quad \Gamma^{1(\alpha)}_{22} = -\frac{\lambda^2+\alpha\lambda^4}{2\mu^2(1-\lambda)}, \quad \Gamma^{2(\alpha)}_{22} = \frac{\lambda-2+\alpha\lambda^2}{2\mu(1-\lambda)}. \tag{3.9}$$

From (2.7), (3.5), (3.6), (3.7), (3.8) and (3.9), we can get

$$R^{(\alpha)}_{1212} = \frac{(\lambda^3-\lambda)\alpha^2 + (-2\lambda^3+2\lambda^2-3\lambda+2)\alpha + 3\lambda - 2}{4\lambda\mu^2(1-\lambda)^2}. \tag{3.10}$$

This completes the proof of Proposition 3.1.

**Theorem 3.1.** The $\alpha$-Gaussian curvature and the $\alpha$-Ricci curvatures of the Pareto manifold are given by

$$K^{(\alpha)} = \frac{(\lambda^3-\lambda)\alpha^2 + (-2\lambda^3+2\lambda^2-3\lambda+2)\alpha + 3\lambda - 2}{4(1-\lambda)^3}$$

and

$$R_{11}^{(\alpha)} = \frac{(\lambda^3-\lambda)\alpha^2 + (-2\lambda^3+2\lambda^2-3\lambda+2)\alpha + 3\lambda - 2}{4\lambda^2(1-\lambda)^3},$$
$$R_{12}^{(\alpha)} = R_{21}^{(\alpha)} = -\frac{(\lambda^3-\lambda)\alpha^2 + (-2\lambda^3+2\lambda^2-3\lambda+2)\alpha + 3\lambda - 2}{4\mu(1-\lambda)^3},$$
$$R_{22}^{(\alpha)} = \frac{(\lambda^3-\lambda)\alpha^2 + (-2\lambda^3+2\lambda^2-3\lambda+2)\alpha + 3\lambda - 2}{4\mu^2(1-\lambda)^3}.$$

**Proof.** From the definition 2.6 and 2.7, by a direct calculation, we can obtain Theorem 3.1, immediately.

From Theorem 3.1 we can get the following

**Corollary 3.1.** When $\alpha = 0$, the Gaussian curvature and the Ricci curvatures satisfy

$$K^{(0)} = \frac{3\lambda - 2}{4(1-\lambda)^3}$$

and

$$R_{11}^{(0)} = \frac{3\lambda-2}{4\lambda^2(1-\lambda)^3}, \quad R_{12}^{(0)} = R_{21}^{(0)} = -\frac{3\lambda-2}{4\mu(1-\lambda)^3}, \quad R_{22}^{(0)} = \frac{3\lambda-2}{4\mu^2(1-\lambda)^3}.$$

**Theorem 3.2.** On the Pareto manifold the Kullback divergence is given by

$$K(p,q) = \ln\frac{\lambda_p}{\lambda_q} + \lambda_q \ln\frac{\mu_p}{\mu_q} + \frac{\lambda_q - \lambda_p}{\lambda_p}$$

and the J-divergence is given by

$$J(p,q) = (\lambda_q - \lambda_p)(\ln\mu_p - \ln\mu_q) + \frac{(\lambda_p-\lambda_q)^2}{\lambda_p\lambda_q}.$$

**Proof.** From (2.12), we can get

$$\begin{aligned} K(p,q) &= E_{\theta_p}[\ln\frac{p(x;\theta_p)}{q(x;\theta_q)}] \\ &= \int p(x;\theta_p)\ln\frac{p(x;\theta_p)}{q(x;\theta_q)}dx \\ &= \ln\frac{\lambda_p}{\lambda_q} + \lambda_q\ln\frac{\mu_p}{\mu_q} + \frac{\lambda_q-\lambda_p}{\lambda_p}. \end{aligned}$$

Then from (2.13), we can get

$$\begin{aligned} J(p,q) &= \int (p(x;\theta_p) - q(x;\theta_q))\ln\frac{p(x;\theta_p)}{q(x;\theta_q)}dx \\ &= K(p,q) + K(q,p) \\ &= (\lambda_q-\lambda_p)(\ln\mu_p - \ln\mu_q) + \frac{(\lambda_p-\lambda_q)^2}{\lambda_p\lambda_q}. \end{aligned}$$

**Corollary 3.2.** When $\mu_p = \mu_q$ , then

$$K(p,q) = \ln \frac{\lambda_p}{\lambda_q} + \frac{\lambda_q - \lambda_p}{\lambda_p}, \quad J(p,q) = \frac{(\lambda_p - \lambda_q)^2}{\lambda_p \lambda_q}.$$

When $\lambda_p = \lambda_q = \lambda$ , then

$$K(p,q) = \lambda \ln \frac{\mu_p}{\mu_q}, \quad J(p,q) = 0.$$

**Theorem 3.3.** The geodesic equations are given by

$$\frac{d^2\lambda}{dt^2} - \frac{1}{\lambda(1-\lambda)}\left(\frac{d\lambda}{dt}\right)^2 + \frac{\lambda}{\mu(1-\lambda)}\frac{d\lambda}{dt}\frac{d\mu}{dt} - \frac{\lambda^2}{2\mu^2(1-\lambda)}\left(\frac{d\mu}{dt}\right)^2 = 0, \quad (3.11)$$

$$\frac{d^2\mu}{dt^2} - \frac{\mu}{\lambda^2(1-\lambda)}\left(\frac{d\lambda}{dt}\right)^2 + \frac{1}{\lambda(1-\lambda)}\frac{d\lambda}{dt}\frac{d\mu}{dt} + \frac{\lambda-2}{2\mu(1-\lambda)}\left(\frac{d\mu}{dt}\right)^2 = 0. \quad (3.12)$$

**Proof.** By the definition 2.9 and using $\Gamma_{ij}^k$ which we have calculated above, we can get the geodesic equations immediately.

In particular, for fixed $\lambda$, from (3.12) we get the solution with respect to $\mu$ that $\mu^{\frac{\lambda}{2(\lambda-1)}} = c_1 t + c_2$ or $\mu = constant$. Similarly, for fixed $\mu$, from (3.11) we get the solution with respect to $\lambda$ that $\lambda - \ln \lambda = c_3 t + c_4$ or $\lambda = constant$.

## 4 Applications.

The Pareto distribution can been used in various fields. J. Shi and H. Zhu ([5]) modeled the TELNET originator traffic by using the Pareto distribution and shown that packet interarrivals of the TELNET originator can be well modeled by Pareto distribution in large time scales. In [6], three modified Edf-type tests, the Kolmogorov-Smirnov (K-S), Anderson-Darling (A-D), and Cramer-von Mises (C-vM), were developed for the Pareto distribution with unknown parameters of location and scale and known shape parameter. In [7], L. Ouyang and S. Wu presented the prediction intervals on future ordered-observation in a sample of size from a Pareto distribution with known shape parameter. Then a useful method was defined for obtaining a bound on life-test duration for sample from a population having Pareto distributed lifetimes.

## 5 Figures of the Gaussian curvature $K^{(\alpha)}$.

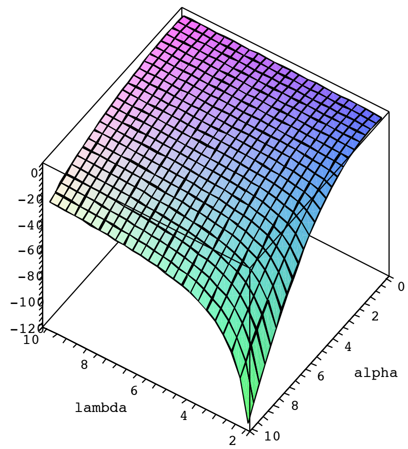


Fig 1. The $\alpha$-Gaussian curvature $K^{(\alpha)}$

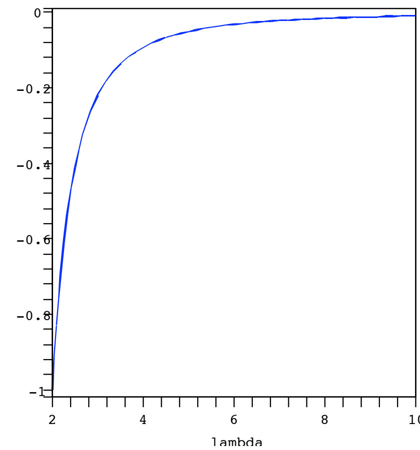


Fig 2. The Gaussian curvature $K^{(0)}$

**Acknowledgement.** The authors would like to thank the referees for the valuable suggestions. Without their help, we couldn't get the final version.

## References

[1] S. Amari, Differential geometrical methods in statistics, Springer lecture Notes in Statistics, 28 (1985).

[2] S. Amari, Methods of information geometry, TMM 191, Oxford University Press, Oxford, 2000.

[3] S. Amari, Differential geometry of a parametric family of invertible linear systems Riemannian metric, dual affine connections and divergence, Mathematical systems theory, 20 (1987b), 53-82.

[4] C. T. J. Dodson, Spatial statistical and information geometry for parametric statistical models of galaxy clustering, Springer Netherlands, 10 (1999).

[5] J. Shi and H. Zhu, The application of the Pareto distribution in modeling TELNET orignator traffic, Journal of computer research and development, 6 (2000), 746-751.

[6] James E. Porter III, Joseph W. Coleman and Albert H. Moore, Modified KS, AD, and C-VM tests for the Pareto distribution with unknown location & scale parameters, IEEE transactions on reliability, 1 (1992).

[7] L. Ouyang and S. Wu, Prediction intervals for an ordered observation from a Pareto distribution, IEEE transactions on reliability, 2 (1994).

Linyu Peng
Department of Mathematics
Beijing Institute of Technology
Beijing, 100081 China
e-mail: bityuquansuperman@126.com

Huafei Sun
Department of Mathematics
Beijing Institute of Technology
Beijing, 100081 China
e-mail: sunhuafei@263.net

Lin Jiu
Department of Mathematics
Beijing Institute of Technology
Beijing, 100081 China
e-mail: xinjilonely@gmail.com