Honam Mathematical J. **35** (2013), No. 1, pp. 029–035
http://dx.doi.org/10.5831/HMJ.2013.35.1.29

A NOTE ON THE BIVARIATE PARETO DISTRIBUTION

Bong Sik Cho[1] and Sun Young Jung[2]

Abstract. The Fisher information matrix plays a significant role in statistical inference in connection with estimation and properties of variance of estimators. Using Bivariate Lomax distribution, we can define "statistical model" and drive the Fisher information matrix of Bivariate Lomax distribution. In this paper, we correct the wrong of the paper [7].

1. Introduction

Information geometry is the differential geometric study of the manifold of probability measures or probability density functions. Recently, information geometric methods have been applied to many areas of the study of estimating functions and nuisance parameter, the dependency of Bayesian predictive distribution, the class of invariant priors for Bayesian inference, principal component analysis, independent component analysis and blind source separation.

Rao (1945) first noticed the importance of the differential-geometrical approach and introduced the Riemannian metric in a statistical manifold by using the Fisher information matrix and calculated the geodesic distance between two distributions for various statistical models. Since then many researchers have tried to obtain the properties of the Riemannian manifold of a statistical model. Efron (1975) elucidated the meaning of curvature for asymptotic statistical inference and pointed that the statistical curvature plays a fundamental role in the higher order asymptotic theory of statistical inference. Amari (1985) remarked

Received October 23, 2012. Accepted February 1, 2013.
2010 Mathematics Subject Classification. 62B10, 62H05, 94B27.
Key words and phrases. Fisher information, bivariate, pareto distribution.
[1]The frist author's research was supported by Wonkwang University in 2011.
[2]Corresponding author

that the two dimensional parameter space of the family of one dimensional normal distribution is a space of negative constant curvature and studied the -geometry of the families of the gamma, Gaussian, Mckey bivariate gamma and the Freund bivariate exponential. Recently, Adbel-All et al. (2003), Kass (1989), Kass and Vos (1997), Murray and Rice (1993) studied the probability density function from the view point of information geometry and use the geometric metrics to give a new description to the statistical distribution. Bivariate Pareto distributions are popular models in many applied areas. They are very versatile and a variety of uncertainties can be usefully modeled by them.

2. Bivariate Pareto distribution

The bivariate Lomax distribution with the joint survivor function has the form

$$F(x,y) = \frac{1}{(1+\theta x+\phi y)^a} \tag{2.1}$$

then

$$\frac{\partial F}{\partial x} = \frac{\partial}{\partial x}(1+\theta x+\phi y)^{-a} = -a\theta(1+\theta x+\phi y)^{-a-1}$$

$$\frac{\partial^2 F}{\partial x \partial y} = (-a)(-a-1)(\theta\phi)(1+\theta x+\phi y)^{-a-2}$$

and the bivariate Lomax distribution has the joint pdf of the form

$$f(x,y) = \frac{\partial^2 F}{\partial x \partial y} = \frac{a(a+1)\theta\phi}{(1+\theta x+\phi y)^{a+2}} \tag{2.2}$$

respectively, for $x>0,\ y>0,\ \theta>0,\ \phi>0$ and $a>0$.

For a given observation (x,y), the Fisher information matrix is defined by

$$(I_{jk}) = \left\{ E\left(\frac{\partial \log L(\theta)}{\partial \theta_j} \frac{\partial \log L(\theta)}{\partial \theta_k} \right) \right\}$$

for $j=1,2,\cdots,p$ and $k=1,2,\cdots,p$, where $L(\theta)=f(x,y)$ and $\theta=(\theta_1,\theta_2,\cdots,\theta_p)$ are the parameters of the pdf f. It has the meaning "information about the parameters θ contained in the observation (x,y)." The information matrix plays a significant role in statistical inference in connection with estimation, sufficiency and properties of variances of estimators. It is related to the covariance matrix of the estimate of θ.

From (2.2), the log-likelihood function is

$$\log L(a,\theta,\phi) = \log\{a(a+1)\theta\phi\} - (a+2)\log(1+\theta x+\phi y).$$

Taking the coordinates (a,θ,ϕ), the first-order derivatives are:

$$\begin{aligned}
\frac{\partial \log L}{\partial \theta} &= \frac{1}{\theta} - (a+2)\frac{x}{1+\theta x+\phi y}, \\
\frac{\partial \log L}{\partial \phi} &= \frac{1}{\phi} - (a+2)\frac{y}{1+\theta x+\phi y}, \quad \text{and} \\
\frac{\partial \log L}{\partial a} &= \frac{1}{a} + \frac{1}{a+1} - \log(1+\theta x+\phi y).
\end{aligned}$$

The second-order derivatives are:

$$\begin{aligned}
\frac{\partial^2 \log L}{\partial \theta^2} &= -\frac{1}{\theta^2} + (a+2)\frac{x^2}{(1+\theta x+\phi y)^2}, \\
\frac{\partial^2 \log L}{\partial \theta \partial \phi} &= (a+2)\frac{xy}{(1+\theta x+\phi y)^2}, \\
\frac{\partial^2 \log L}{\partial \theta \partial a} &= -\frac{x}{1+\theta x+\phi y}, \\
\frac{\partial^2 \log L}{\partial \phi^2} &= -\frac{1}{\phi^2} + (a+2)\frac{y^2}{(1+\theta x+\phi y)^2}, \\
\frac{\partial^2 \log L}{\partial \phi \partial a} &= -\frac{y}{1+\theta x+\phi y}, \quad \text{and} \\
\frac{\partial^2 \log L}{\partial a^2} &= -\frac{1}{a^2} - \frac{a}{(a+1)^2}.
\end{aligned}$$

Using the formula

$$E(X^m Y^n|a) = mn\int_0^\infty\int_0^\infty x^{m-1}y^{n-1}(1+\theta x+\phi y)^{-a}dydx \tag{2.3}$$

for $a > \max\{m, n-1\}, m < a, n < a+1$ (where denotes the shape parameter in (2.1)), we can express the element of the Fisher information matrix as

$$\begin{aligned}
E\left(-\frac{\partial^2 \log L}{\partial \theta^2}\right) &= \frac{1}{\theta^2} - \frac{a(a+1)(a+2)\theta\phi}{3}E(X^3Y|a+4), \\
E\left(-\frac{\partial^2 \log L}{\partial \theta \partial \phi}\right) &= -\frac{a(a+1)(a+2)\theta\phi}{4}E(X^2Y^2|a+4), \\
E\left(-\frac{\partial^2 \log L}{\partial \theta \partial a}\right) &= \frac{a(a+1)\theta\phi}{2}E(X^2Y|a+3),
\end{aligned}$$

$$E\left(-\frac{\partial^2 \log L}{\partial \phi^2}\right) = \frac{1}{\phi^2} - \frac{a(a+1)(a+2)\theta\phi}{3} E(XY^3|a+4),$$
$$E\left(-\frac{\partial^2 \log L}{\partial \phi \partial a}\right) = \frac{a(a+1)\theta\phi}{2} E(XY^2|a+3), \quad \text{and}$$
$$E\left(-\frac{\partial^2 \log L}{\partial a^2}\right) = \frac{1}{a^2} + \frac{1}{(a+1)^2}.$$

Theorem 2.1. If X and Y are jointly distributed according to (2.2) then

$$E(X^m Y^n|a) = \frac{mnB(n, a-n)B(m, a-m-n)}{\theta^m \phi^n}.$$

for $a > m + n \geq 1$, where B is a beta function.

Proof. Using the formula (2.3), we can express

$$\begin{aligned}
&E(X^m Y^n|a) \\
&= mn \int_0^\infty \int_0^\infty x^{m-1} y^{n-1} (1 + \theta x + \phi y)^{-a} dy dx \\
&= mn \int_0^\infty \int_0^\infty x^{m-1} y^{n-1} \left(\phi\left(y + \frac{1+\theta x}{\phi}\right)\right)^{-a} dy dx \\
&= mn \int_0^\infty x^{m-1} \phi^{-a} \int_0^\infty y^{n-1} (y+c)^{-a} dy dx \\
&= mn \int_0^\infty x^{m-1} \phi^{-a} \int_0^\infty c t^{n-1} (ct + c)^{-a} c \; dt dx \\
&= mn \int_0^\infty x^{m-1} \phi^{-a} c^{n-1-a+1} \int_0^\infty t^{n-1} (t+1)^{-a} dt dx \\
&= mn \int_0^\infty x^{m-1} \phi^{-a} c^{n-a} \int_0^1 \left(\frac{z}{1-z}\right)^{n-1} \left(\frac{z}{1-z} + 1\right)^{-a} \frac{1}{(1-z)^2} dz dx \\
&= mn \int_0^\infty x^{m-1} \phi^{-a} c^{n-a} \int_0^1 (z)^{n-1} (1-z)^{a-n-1} dz dx \\
&= mn \int_0^\infty x^{m-1} \phi^{-a} c^{n-a} B(n, a-n) dx \\
&= mn \int_0^\infty x^{m-1} \phi^{-a} \left(\frac{1+\theta x}{\phi}\right)^{n-a} B(n, a-n) dx
\end{aligned}$$

$$= mnB(n, a-n) \int_0^\infty x^{m-1} \phi^{-a} \left(\frac{\theta}{\phi} \left(x + \frac{1}{\theta} \right) \right)^{n-a} dx$$
$$= \frac{mnB(n, a-n)B(m, a-m-n)}{\theta^m \phi^n}.$$

□

By Theorem 2.1, the expectations above can be calculated as

$$E(X^3Y|a+4) = \frac{3B(1, a+3)B(3, a)}{\theta^3\phi} = \frac{6}{\theta^3\phi(a+4)(a+3)(a+2)(a+1)},$$

$$E(XY^3|a+4) = \frac{3B(3, a+1)B(1, a)}{\theta\phi^3} = \frac{6}{\theta\phi^3(a+4)(a+3)(a+2)(a+1)},$$
$$E(X^2Y^2|a+4) = \frac{4B(2, a+2)B(2, a)}{\theta^2\phi^2} = \frac{4}{\theta^2\phi^2(a+4)(a+3)(a+2)(a+1)},$$
$$E(X^2Y|a+3) = \frac{2B(1, a+2)B(2, a)}{\theta^2\phi} = \frac{2}{\theta^2\phi(a+3)(a+2)(a+1)},$$
and
$$E(XY^2|a+3) = \frac{2B(2, a+1)B(1, a)}{\theta\phi^2} = \frac{2}{\theta\phi^2(a+3)(a+2)(a+1)}.$$

Theorem 2.2. Let M be the family of bivariate Pareto distributions, then a, θ, ϕ is a local coordinate system, and M become a 3-manifold. Thus we can obtain the Fisher information matrix

$$E\left(-\frac{\partial^2 \log L}{\partial a^2}\right) = \frac{1}{a^2} + \frac{1}{(a+1)^2}$$
$$E\left(-\frac{\partial^2 \log L}{\partial a \partial \theta}\right) = \frac{a}{\theta(a+3)(a+2)}$$
$$E\left(-\frac{\partial^2 \log L}{\partial a \partial \phi}\right) = \frac{a}{\phi(a+3)(a+2)}$$
$$E\left(-\frac{\partial^2 \log L}{\partial \theta^2}\right) = \frac{1}{\theta^2} - \frac{2a}{\theta^2(a+4)(a+3)}$$
$$E\left(-\frac{\partial^2 \log L}{\partial \theta \partial \phi}\right) = -\frac{a}{\theta\phi(a+4)(a+3)}$$
$$E\left(-\frac{\partial^2 \log L}{\partial \phi^2}\right) = \frac{1}{\phi^2} - \frac{2a}{\phi^2(a+4)(a+3)}.$$

3. Conclusions

The Fisher information matrix(FIM) measures the curvature of the log-likelihood surface. Flat surfaces around the maximum do not inspire high confidence in estimated parameter values, while steep surfaces lead to sharp estimates. It is important to know the shape of a statistical model in the whole set of probability distributions. The information content is large if the FIM is large, because the likelihood is sharped peaked. we are sure that the maximum likelihood(ML) solution is a good estimate. If the curvature is small, then the likelihood probability distribution is very broad. So the ML estimate is not as good because the variance is very large.

References

[1] Abdel, N. H., Abd-Ellah, H. N., Moustata, H. M., Information geometry and statistical manifoeld, Chaos, Solitons and Fractal,(2003) 161-172

[2] Abdel, N. H., Mahmoud, M. A. W. and Abd-Ellah, H. N., Geometrical properties of Pareto distribution, Appl. Math. and Com. (2003) 321-339.

[3] Amari, S., Differential geometrical methods in statistics, Springer lecture notes in Statistics, (1985).

[4] Efron, B., Defining the curvature of a statistical problem, Annual. Statisitcs. vol 3. no. 6 (1975) 1109-1242.

[5] Kass, R. E. and Vos. P. W., Geometrical foundations of asymptotic inference, John Wiley and Sons, Inc., (1997).

[6] Kass, R. E. and Vos. P. W., The geometry of saympeoeic inference, August 1989 vol 4, no. 3, Statistical Science.

[7] Kotz, S. and Nadarajah, S. Information matrices for some bivariate Pareto distribution, Advances on Income Inequality and, 2008

[8] Murray, M. K. and Rice, J. W., Differential geometry and Statistics, Chapman and Hall, New York, (1993).

[9] Rao. C. R. Information and the accuracy attainable in the estimation of statistical parameters, Bull. Calcutta Math. Soc. 37, (1945) 81-91.

[10] Samuel L. Katz. and Saralees Nadarajah., Information matrices for some bivariate Pareto distributions.

[11] William, W. S. Chen, On computing Gaussian curvature of some well known distributions, Amer. Statist. Soc. : section on Bayesian statist. sci., (1999) 129-134.

Bong Sik Cho
Division of Mathematics and Information Statistics, Research Institute for Basic Sciences, Wonkwang University,

Iksan, Chonbuk 570-749, Korea.
E-mail: bscho@wonkwang.ac.kr

Sun Young Jung
17, Haengdang-dong, Seongdong-gu, Seoul 133-791, Korea.
E-mail: jsycbs@hanyang.ac.kr