Honam Mathematical J. **29** (2007), No. 1, pp. 75–81

A NOTE ON THE GEOMETRICAL PROPERTIES OF THE NORMAL DISTRIBUTION

Bong Sik Cho

Abstract. The Fisher information matrix plays a significant role in statistical inference in connection with estimation and properties of variance of estimators. In this paper, the parameter space of the normal distribution using its Fisher's matrix is defined. The Riemannian curvature and J-divergence to parameter space are calculated.

1. Introduction

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) defined the statistical curvature of statistical model and pointed out that the statistical curvature plays a fundamental role in the higher order asymptotic theory of statistical inference. Amari(1980) remarked that two dimensional parameter space of the family of one dimensional normal distribution is a space of negative constant curvature. In this paper, we find the Fisher information matrix and J-divergence of the normal distribution.

Received Feb. 7, 2007. Accepted Mar. 28, 2007.
2000 Mathematics Subject Classification: 62B10, 62H05, 94B27.
Key words and phrases: Fisher information, J-divergence.
This paper was supported by Wonkwang University in 2006.

2. Geometry of parameter spaces

Let $S = \{p(x,\theta)\,|\,x$: random variable, $\theta = (\theta^1, \cdots, \theta^n) \in \mathbb{R}^n\}$ be a family of probability distributions. The family S is regarded as an n-dimensional manifold having θ as a coordinate system. Rao[8] has proved that when the Fisher information matrix $G(\theta) = (g_{ij}(\theta))$

$$\begin{aligned} g_{ij}(\theta) &= \int \frac{\partial \ln p(x,\theta)}{\partial \theta^i} \frac{\partial \ln p(x,\theta)}{\partial \theta^j} p(x,\theta) d\theta \\ &= -E\left[\frac{\partial^2 \ln p}{\partial \theta^i \partial \theta^j}\right] \end{aligned} \tag{2.1}$$

where E denoted to the expectation with respect to $p(x,\theta)$, is nondegenerate, S is a Riemannian manifold, and $G(\theta)$ plays the role of a Riemannian metric tensor. The infinitesimal distance dS between two nearby distributions $p(x,\theta)$ and $p(x,\theta + d\theta)$ is defined by

$$dS^2 = \sum_{i,j=1}^{n} g_{ij}(\theta) d\theta^i d\theta^j.$$

The quantities

$$\Gamma^i_{jk} = \frac{1}{2} \sum_l g^{il} \left(\frac{\partial g_{jl}}{\partial \theta^k} + \frac{\partial g_{kl}}{\partial \theta^j} - \frac{\partial g_{jk}}{\partial \theta^l} \right) \tag{2.2}$$

are called the *Christoffel symbols* and (g^{il}) is the contravariant metric tensor field of the covariant metric tensor field (g_{il}). The Riemannian curvature tensor in (S, g_{ij}) is defined by

$$R^s_{ijk} = \Gamma^s_{ik,j} - \Gamma^s_{jk,i} + \Gamma^l_{ik}\Gamma^s_{jl} - \Gamma^l_{jk}\Gamma^s_{il} \quad \text{sum on } l$$

where comma denotes the partial derivative. Putting

$$R_{ijks} = \sum_l R^l_{ijk} g_{ls}, \tag{2.3}$$

we can write as:

$$\begin{aligned} &R_{ijks} + R_{jkis} + R_{kijs} = 0, \\ &R_{ijks} = -R_{jiks} = -R_{ijsk} = R_{ksij}. \end{aligned}$$

Thus the Ricci tensor is given as

$$R_{ik} = \sum_j R^j_{ijk} = \sum_{s,j} R_{ijks} g^{sj}.$$

The scalar curvature R and the Gaussian curvature K are defined by

$$R = \sum_{i,j} g^{ij} R_{ij}, \quad K = \frac{R_{1212}}{\det G}. \tag{2.4}$$

The Kullback-Leibler distance between two probability density function $p(x, \theta_p)$ and $q(x, \theta_q)$ is given by

$$I(p, q) = \int p(x, \theta_p) \ln \frac{p(x, \theta_p)}{q(x, \theta_q)} dx. \tag{2.5}$$

The J-divergence between two extremely closed probability density function $p = f(x, \theta)$ and $q = f(\theta, \theta + d\theta)$ is given by

$$\begin{aligned} J(p,q) &= \int (p-q) \ln \frac{p}{q} dx \\ &= \left[\int \frac{\partial \ln f}{\partial \theta^i} \frac{\partial \ln f}{\partial \theta^j} f dx \right] d\theta^i d\theta^j \\ &= \sum_{i,j} g_{ij}(\theta) d\theta^i d\theta^j \\ &= dS^2. \end{aligned} \tag{2.6}$$

properties of the normal distribution

3. Geometrical properties of the normal distribution

Theorem 3.1. *Let Ω be a location scale surface of density that has the following general form*

$$\Omega = \{ f(x) = \frac{1}{\sqrt{2\pi v^2}} \, e^{-\frac{(x-u)^2}{2v^2}} \mid (u, v) \in \mathbb{R} \times \mathbb{R}_+ \} \tag{3.1}$$

where u is the location parameter and v is the scale parameter. Then the Gaussian curvature of normal distribution is $K = -\frac{1}{2}$.

Proof.

$$\ln f = -\frac{1}{2}\ln(2\pi v^2) - \frac{(x-u)^2}{2v^2},$$
$$\frac{\partial}{\partial u}\ln f = \frac{x-u}{v^2},$$
$$\frac{\partial}{\partial v}\ln f = -\frac{1}{v} + \frac{(x-u)^2}{v^3}.$$

Then $E[x] = \int_{-\infty}^{\infty} xf(x)dx = u$, $V[x] = \int_{-\infty}^{\infty}(x-u)^2 f(x)dx = v^2$. Thus, by (2.1)

$$\begin{aligned} g_{11} &= -E\left[\frac{\partial^2}{\partial u^2}\ln f\right] = E\left[\frac{1}{v^2}\right] = \frac{1}{v^2}, \\ g_{12} &= g_{21} = -E\left[\frac{\partial^2}{\partial v \partial u}\ln f\right] = E\left[\frac{2(x-u)}{v^3}\right] \\ &= \frac{2}{v^3}\left(E[x] - E[u]\right) = 0, \\ g_{22} &= -E\left[\frac{\partial^2}{\partial v^2}\ln f\right] = E\left[-\frac{1}{v^2} + \frac{3(x-u)^2}{v^4}\right] \\ &= -\frac{1}{v^2} + \frac{3}{v^4}E[(x-u)^2] = -\frac{1}{v^2} + \frac{3v^2}{v^4} = \frac{2}{v^2}. \end{aligned}$$

From (2.2), (2.3) and (2.4)

$$\Gamma_{11}^1 = \Gamma_{12}^2 = \Gamma_{22}^1 = 0,\ \Gamma_{12}^1 = -\frac{1}{v},\ \Gamma_{11}^2 = \frac{1}{2v},\ \Gamma_{22}^2 = -\frac{1}{v},$$
$$R_{121}^2 = \frac{\partial}{\partial v}\Gamma_{11}^2 - \Gamma_{21}^1\Gamma_{11}^2 + \Gamma_{11}^2\Gamma_{22}^2 = -\frac{1}{2v^2},$$
$$R_{1212} = R_{121}^2 g_{22} = -\frac{1}{v^4},$$
$$K = \frac{R_{1212}}{\det G} = -\frac{1}{2}.$$

□

Therefore the normal family can be identified with the upper-half plane

$$H = \{(x,\, y) \mid y > 0\}$$

with the metric

$$\begin{aligned} g\left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}\right) &= 0, \\ g\left(\frac{\partial}{\partial x}, \frac{\partial}{\partial x}\right) &= \frac{1}{y^2}, \\ \text{and} \quad g\left(\frac{\partial}{\partial y}, \frac{\partial}{\partial y}\right) &= \frac{2}{y^2}. \end{aligned}$$

The geometry of this upper-half plane is hyperbolic geometry, which was the first non-Euclidean geometry to be discovered. It is striking that the simplest non-trivial family of distributions should give rise to the simplest non-Euclidean geometry.

Theorem 3.2. *The Kullback-Leibler distance of the normal probability density functions $p(x, \theta_p)$ and $q(x, \theta_q)$ satisfies the following equation.*

$$I(p, q) = \ln\frac{v_q}{v_p} + \frac{{v_p}^2}{2{v_q}^2} - \frac{1}{2} + \frac{(u_p - u_q)^2}{2{v_q}^2} \tag{3.2}$$

Proof. Let

$$\begin{aligned} p(x) &= \frac{1}{\sqrt{2\pi {v_p}^2}}\, e^{-\frac{(x-u_p)^2}{2{v_p}^2}}, \\ q(x) &= \frac{1}{\sqrt{2\pi {v_q}^2}}\, e^{-\frac{(x-u_q)^2}{2{v_q}^2}}. \end{aligned}$$

Then, by (2.5)

$$\begin{aligned} I(p,q) &= \int p(x) \ln \left(\frac{\frac{1}{v_p\sqrt{2\pi}}\, e^{-\frac{(x-u_p)^2}{2{v_p}^2}}}{\frac{1}{v_q\sqrt{2\pi}}\, e^{-\frac{(x-u_q)^2}{2{v_q}^2}}} \right) dx \\ &= \int p(x) \left(\ln\frac{v_q}{v_p} + \frac{(x-u_q)^2}{2{v_q}^2} - \frac{(x-u_p)^2}{2{v_p}^2} \right) dx \\ &= \int p(x) \ln\frac{v_q}{v_p} dx + \int p(x) \left(\frac{(x-u_q)^2}{2{v_q}^2} - \frac{(x-u_p)^2}{2{v_p}^2} \right) dx. \end{aligned}$$

From $(x-u_q)=(x-u_p)+(u_p-u_q)$,

$$\begin{aligned}
&\int p(x)\left(\frac{(x-u_q)^2}{2{v_q}^2}-\frac{(x-u_p)^2}{2{v_p}^2}\right)dx \\
= \quad & \int p(x)\left(\frac{(x-u_p)^2}{2{v_q}^2}+\frac{(u_p-u_q)^2}{2{v_q}^2}+\frac{2(x-u_p)(u_p-u_q)}{2{v_q}^2}-\frac{(x-u_p)^2}{2{v_p}^2}\right)dx \\
= \quad & \int p(x)\left(\frac{{v_p}^2}{{v_p}^2}\frac{(x-u_p)^2}{2{v_q}^2}-\frac{(x-u_p)^2}{2{v_p}^2}\right)dx \\
& +\int p(x)\left(\frac{(u_p-u_q)^2}{2{v_q}^2}+\frac{(x-u_p)(u_p-u_q)}{{v_q}^2}\right)dx \\
= \quad & \frac{1}{2{v_p}^2}\left(\frac{{v_p}^2}{{v_q}^2}-1\right)\int p(x)(x-u_p)^2dx+\frac{(u_p-u_q)^2}{2{v_q}^2} \\
& +\frac{(u_p-u_q)}{{v_q}^2}\int p(x)(x-u_p)dx \\
= \quad & \frac{1}{2{v_p}^2}\left(\frac{{v_p}^2}{{v_q}^2}-1\right)E((x-u_p)^2)+\frac{(u_p-u_q)^2}{2{v_q}^2}+\frac{(u_p-u_q)}{{v_q}^2}E(x-u_p).
\end{aligned}$$

Since $E(x-u_p)=0$ and $E((x-u_p)^2)={v_p}^2$,

$$I(p,q)=\ln\frac{v_q}{v_p}+\frac{{v_p}^2}{2{v_q}^2}-\frac{1}{2}+\frac{(u_p-u_q)^2}{2{v_q}^2}. \qquad \square$$

Theorem 3.3. *The J-divergence of the normal probability density functions $p(x,\theta_p)$ and $q(x,\theta_q)$ satisfies the following equation.*

$$J(p,q)=\frac{({u_p}^2-{v_q}^2)^2}{2{v_p}^2{v_q}^2}+\frac{(u_p-u_q)^2({v_p}^2+{v_q}^2)}{2{v_p}^2{v_q}^2} \tag{3.3}$$

Proof. From (2.6),

$$J(p,q)=I(p,q)+I(q,p). \tag{3.4}$$

From (3.2),

$$I(p,q)=\ln\frac{v_q}{v_p}+\frac{{v_p}^2}{2{v_q}^2}-\frac{1}{2}+\frac{(u_p-u_q)^2}{2{v_q}^2}. \tag{3.5}$$

Similarly, we have

$$I(q,p)=\ln\frac{v_p}{v_q}+\frac{{v_q}^2}{2{v_p}^2}-\frac{1}{2}+\frac{(u_q-u_p)^2}{2{v_p}^2}. \tag{3.6}$$

Using (3.4), (3.5) and (3.6), we get the relation (3.3). $\square$

If $u_p = u_q$, (3.3) reduced to

$$J(p,q) = \frac{({v_p}^2 - {v_q}^2)^2}{2{v_p}^2{v_q}^2}.$$

If $v_p = v_q = v$, (3.3) takes the form

$$J(p,q) = \frac{(u_p - u_q)^2}{v^2}.$$

References

[1] Abdel, N. H., Mahmoud, M. A. W., Moustafa, H. M. *Information geometry and statistical manifold,* Chaos, Solitons and Fractals **15** (2003) 161-172.

[2] Abdel, N. H., Mahmoud, M. A. W., Abd-Ellah, H. N. *Geometrical properties of Pareto distribution*, Appl. Math. and Computation **145** (2003) 321-339.

[3] Amari, S. *Differential geometrical methods in statistics*, Lecture notes in statistics, **28**, Springer Verlag, Heidelberg (1986).

[4] Amari, S. and Nagaoka, H. *Methods of information geometry,* **191**, A. M. S. (1993).

[5] Chen, W. W. S. *On computing Gaussian curvature of some well known distributions*, A. S. A. : section on Bayesian statistical science (1999), 129-134.

[6] Costa, S. I. R. and Santos, S. A. and Strapasson, J. E. *Fisher information matrix and hyperbolic geometry,* Information Theory Workshop, 2005 IEEE (2005), 28-30.

[7] Edwards, R. D. *International distributions of the age at death and mortality convergence*, Population association of America 2004 Annual Meeting Program, Boston, Massachusetts, April **1-3** (2004).

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

Bong Sik Cho
Division of Mathematics and Information Statistics
Research institute for basic sciences
Wonkwang University
Iksan, Chonbuk, Korea 570-749
E-mail: bscho@wonkwang.ac.kr