*Journal of the Korean Data & Information Science Society*
*2010,* **21**(3), 575–580

한국데이터정보과학회지

# A note on the geometric structure of the t-distribution[†]

Bong Sik Cho[1]· Sun Young Jung[2]

[1]Division of Mathematics and Information Statistics, Wonkwang University
[2]College of Engineering, Hanyang University

Received 24 March 2010, revised 5 May 2010, accepted 23 May 2010

**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 t-distribution using its Fisher's matrix is defined. The $\alpha$ -scalar curvatures to parameter space are calculated.

*Keywords*: Alpha-connection, alpha-scalar curvature, Fisher information matrix, Gaussian curvature.

## 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 (1982) introduced the $\alpha$-connection and $\alpha$-curvature. Then he pointed out important roles of the exponential and mixture curvatures and their duality in statistical inference. Amari (1985) remarked that the two dimensional parameter space of the family of one dimensional normal distribution is a space of negative constant curvature and studied the $\alpha$ -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 viewpoint of information geometry and use the geometric metrics to give a new description to the statistical distribution. Arwini and Dodson (2007) studied the $\alpha$ -geometry of the Weibull manifold. In this paper, we find the $\alpha$-connection and $\alpha$ -scalar curvature of the t-distribution.

[†] This paper was supported by Wonkwang University in 2008.
[1] Professor, Division of Mathematics and Information Statistics, Research institute for basic sciences, Wonkwang University, Iksan, Chonbuk 570-749, Korea.
[2] Corresponding author: Assistant professor, College of Engineering, Hanyang University, 17 Haengdang-dong Seongdong-gu Seoul 133-791, Korea. E-mail: jsycbs@hanyang.ac.kr

## 2. Geometry of parameter spaces

Let $S = \{p(x,\theta)|x:$ random variable, $\theta = (\theta_1, \cdots, \theta_n) \in R^n\}$ be a family of probability distributions. The family $S$ is regarded as an $n$-dimensional manifold having $\theta$ as a coordinate system. Rao (1945) has proved that when the Fisher information matrix $G(\theta) = (g_{ij}(\theta))$

$$g_{ij}(\theta) = -E\left[\partial_i \partial_j l\right]. \tag{2.1}$$

where $E$ denoted to the expectation with respect to $p(x,\theta)$ and $l(x,\theta) = \ln p(x,\theta)$, is non-degenerate, $S$ is a Riemannian manifold and $G(\theta)$ plays the role of a Riemannian metric tensor. In the following, we adapt the Einstein summation convention. The infinitesimal distance $dS$ between two nearby distributions $p(x,\theta)$ and $p(x,\theta+d\theta)$ is defined by

$$dS^2 = g_{ij}(\theta) d\theta_i d\theta_j.$$

The quantities

$$\Gamma^i_{jk} = \frac{1}{2} g^{il} \left( \frac{\partial g_{jl}}{\partial \theta_k} + \frac{\partial g_{kl}}{\partial \theta_j} - \frac{\partial g_{jk}}{\partial \theta_l} \right) \tag{2.2}$$

are called the Christoffel symbols and $(g^{il})$ is the contravariant metric tensor field of the covariant metric tensor field $(g_{il})$. Then we have

$$\Gamma^k_{ij} = g^{km} \Gamma_{ijm}, \quad \Gamma_{ijm} = \Gamma^k_{ij} g_{km}. \tag{2.3}$$

The $\alpha$-connection is defined by

$$\begin{aligned} \Gamma^{(\alpha)}_{ijk} &= E\left[ \left( \partial_i \partial_j l + \frac{1-\alpha}{2} \partial_i l \partial_j l \right) \partial_k l \right] \\ &= \Gamma^{(1)}_{ijk} + \frac{1-\alpha}{2} T_{ijk}, \end{aligned} \tag{2.4}$$

where $T_{ijk} = E\left[\partial_i l \partial_j l \partial_k l\right]$. From $\Gamma^{(\alpha)}_{ijk} = \Gamma^{(\beta)}_{ijk} + T_{ijk}(\beta - \alpha)/2$,

$$\Gamma^{(\alpha)}_{ijk} = \Gamma_{ijk} - \frac{\alpha}{2} T_{ijk}. \tag{2.5}$$

The Riemannian curvature tensor in $(S, g_{ij})$ is defined by

$$R^s_{ijk} = \Gamma^s_{ik,j} - \Gamma^s_{jk,i} + \Gamma^l_{ik} \Gamma^s_{jl} - \Gamma^l_{jk} \Gamma^s_{il}, \tag{2.6}$$

where comma denotes the partial derivative. Putting

$$R_{ijks} = R^l_{ijk} g_{ls}, \tag{2.7}$$

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

$$R_{ik} = R^{j}_{ijk} = R_{ijks} g^{sj}. \tag{2.8}$$

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

$$R = g^{ij} R_{ij}, \quad K = \frac{R_{1212}}{\det G}. \tag{2.9}$$

The $\alpha$-curvature tensors $R^{(\alpha)}_{ijkm}$ are defined by

$$R^{(\alpha)}_{ijkm} = \left(\Gamma^{s(\alpha)}_{ik,j} - \Gamma^{s(\alpha)}_{jk,i}\right) g_{sm} + \Gamma^{(\alpha)}_{jrm} \Gamma^{r(\alpha)}_{ik} - \Gamma^{(\alpha)}_{irm} \Gamma^{r(\alpha)}_{jk} \tag{2.10}$$

where $\Gamma^{k(\alpha)}_{ij} = \Gamma^{(\alpha)}_{ijm} g^{km}$. The $\alpha$ -scalar curvature $K^{\alpha}$ is defined by

$$K^{(\alpha)} = \frac{1}{n(n-1)} R^{(\alpha)}_{ijkm} g^{im} g^{jk}. \tag{2.11}$$

## 3. Geometric interpretation of student t-distribution

Let $\Omega$ be the location-scale manifold of density that has the form

$$\Omega = \left\{ f(x) = \frac{1}{v} \frac{\Gamma\left(\frac{r+1}{2}\right)}{\sqrt{\pi r}\,\Gamma\left(\frac{r}{2}\right)} \left(1 + \frac{1}{r}\left(\frac{x-u}{v}\right)^2\right)^{-\frac{r+1}{2}} | x \in R, (u,v) \in R \times R^{+} \right\} \tag{3.1}$$

where $u$ is location parameter and $v$ is scale parameter and $r$ is degrees of freedom. From Cho and Baek (2006), Fisher information matrix $(g_{ij})$ is given by

$$(g_{ij}) = \begin{pmatrix} \frac{r+1}{v^2(r+3)} & 0 \\ 0 & \frac{2r}{v^2(r+3)} \end{pmatrix}. \tag{3.2}$$

Let $z = \frac{x-u}{\sqrt{r}v}$. Then

$$1 = \int_{-\infty}^{\infty} f(x)dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{\pi}} \frac{\Gamma\left(\frac{r+1}{2}\right)}{\Gamma\left(\frac{r}{2}\right)} (1+z^2)^{-\frac{r+1}{2}} dz = \int_{-\infty}^{\infty} f(z)dz.$$

We can calculate

$$E\left(\frac{1}{Z^2+1}\right) = \frac{r}{r+1}, \quad E\left(\frac{1}{(Z^2+1)^2}\right) = \frac{r(r+2)}{(r+1)(r+3)}. \tag{3.3}$$

Moreover

$$E\left(\frac{1}{(Z^2+1)^3}\right) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{\pi}} \frac{\Gamma\left(\frac{r+1}{2}\right)}{\Gamma\left(\frac{r}{2}\right)} \left(1+z^2\right)^{-\frac{r+7}{2}} dz$$

$$= \frac{\Gamma\left(\frac{r+1}{2}\right)\Gamma\left(3+\frac{r}{2}\right)}{\Gamma\left(\frac{r}{2}\right)\Gamma\left(3+\frac{r+1}{2}\right)} = \frac{r(r+2)(r+4)}{(r+1)(r+3)(r+5)}. \tag{3.4}$$

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

$$R^{(\alpha)}_{1212} = -\frac{(r+1)\left\{(r+5)^2 - \alpha^2(r-1)^2\right\}}{v^4(r+3)(r+5)^2}. \tag{3.5}$$

**Proof**: From (2.2), (2.3) and (3.2)

$$\Gamma_{112} = \Gamma^k_{11} g_{k2} = \frac{r+1}{v^3(r+3)}, \quad \Gamma_{121} = \Gamma^k_{12} g_{k1} = -\frac{r+1}{v^3(r+3)},$$
$$\Gamma_{222} = \Gamma^k_{22} g_{k2} = -\frac{2r}{v^3(r+3)}, \quad \Gamma_{122} = \Gamma_{111} = 0. \tag{3.6}$$

By (2.4)

$$T_{112} = E[\partial_1 l \partial_1 l \partial_2 l] = \frac{4ab^2}{v^3}\left\{(2b-1)E\left[\frac{Z^4}{(1+Z^2)^3}\right] - E\left[\frac{Z^2}{(1+Z^2)^3}\right]\right\},$$
$$T_{222} = E[(\partial_2 l)^3] = \frac{1}{v^3}\left\{(2b-1)^3 - 6b(2b-1)^2 E\left[\frac{1}{1+Z^2}\right]\right. \tag{3.7}$$
$$\left. + 12b^2(2b-1)E\left[\frac{1}{(1+Z^2)^2}\right] - 8b^3 E\left[\frac{1}{(1+Z^2)^3}\right]\right\},$$
$$T_{111} = T_{122} = 0,$$

where $a = 1/r$ and $b = (r+1)/2$. From (3.3), (3.7) and partial traction decomposition

$$T_{112} = \frac{2(r+1)(r-1)}{v^3(r+3)(r+5)}, \quad T_{222} = \frac{8r(r-1)}{v^3(r+3)(r+5)}, \quad T_{111} = T_{122} = 0. \tag{3.8}$$

By (2.5), (3.5) and (3.8)

$$\begin{aligned}
\Gamma_{112}^{(\alpha)} &= \Gamma_{112} - \frac{\alpha}{2}T_{112} = \frac{(r+1)\{(r+5)-\alpha(r-1)\}}{v^3(r+3)(r+5)} \\
\Gamma_{222}^{(\alpha)} &= \Gamma_{222} - \frac{\alpha}{2}T_{222} = \frac{-2r\{(r+5)+2\alpha(r-1)\}}{v^3(r+3)(r+5)} \\
\Gamma_{121}^{(\alpha)} &= \Gamma_{121} - \frac{\alpha}{2}T_{121} = \frac{-(r+1)\{(r+5)+\alpha(r-1)\}}{v^3(r+3)(r+5)} \\
\Gamma_{111}^{(\alpha)} &= \Gamma_{122}^{(\alpha)} = 0.
\end{aligned} \tag{3.9}$$

Thus

$$\begin{aligned}
\Gamma_{11}^{2(\alpha)} &= \frac{(r+1)\{(r+5)-\alpha(r-1)\}}{2rv(r+5)} \\
\Gamma_{21}^{1(\alpha)} &= \frac{-\{(r+5)+\alpha(r-1)\}}{v(r+5)} \\
\Gamma_{21}^{2(\alpha)} &= \Gamma_{11}^{1(\alpha)} = 0.
\end{aligned} \tag{3.10}$$

By (2.10), (3.2), (3.9) and (3.10)

$$\begin{aligned}
R_{1212}^{(\alpha)} &= \left(\partial_2\Gamma_{11}^{s(\alpha)} - \partial_1\Gamma_{21}^{s(\alpha)}\right)g_{s2} + (\Gamma_{1r2}^{(\alpha)}\Gamma_{21}^{r(\alpha)} - \Gamma_{2r2}^{(\alpha)}\Gamma_{11}^{r(\alpha)}) \\
&= -\frac{(r+1)\left\{(r+5)^2 - \alpha^2(r-1)^2\right\}}{v^4(r+3)(r+5)^2}.
\end{aligned}$$

□

**Theorem 3.2.** The $\alpha$ -scalar curvature of the t-distribution is given by

$$K^{(\alpha)} = -\frac{(r+3)\left\{(r+5)^2 - \alpha^2(r-1)^2\right\}}{2r(r+5)^2}.$$

**Proof**: Since $K^{(\alpha)} = (1/2)(-2R_{1212}^{(\alpha)}g^{11}g^{22})$ from (2.11), we obtain the result. □

Thus if $\alpha = 0$, we have the following corollary.

**Corollary 3.3.** The scalar curvature $R$ and the Gaussian curvature $K$ of the t-distribution, are

$$R = -\frac{r+3}{r}, \quad K = \frac{1}{2}R.$$

**Example 3.4.** Metric for a single neuron (Wagenaar, 1998). A single N-input binary neuron with output defined as follows;

$$y(t) = sgn[\tanh \beta h(x) + \eta(t)]$$

with $h(x) = \sum_{i=1}^{N} J^i x_i + J^0$. In this, $J^i$ are connection weights, $J^0$ is the external field or bias, $x_i$ are the inputs, and $\eta(t)$ is a source of uniform random noise in $[-1, 1]$.

## 4. Conclusions

It should be very important to know the shape of a statistical model in the whole set of probability distributions. The Fisher information matrix (FIM) is the curvature of the likelihood at the mode. The information content is large if the FIM is large, because the likelihood is sharped peaked. we are very 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. A one-parameter family of affine connections are called the $\alpha$ -connections. The duality between the $\alpha$-connection and the $\alpha$-connection together with the metric play an essential role in this geometry. This kind of duality, having emerged from manifolds of probability distributions, is ubiquitous, appearing in a variety of problems which might have no explicit relation to probability theory. The notion of $\alpha$-curvature serves an important role in the asymptotic theory of statistical estimation, ancillary statistics, conditional inference and Bartlett adjustment in the likelihood ratio test.

## References

Adbel-All, N. H., Abd-Ellah, H. N. and Moustafa, H. M. (2003). Information geometry and statistical manifold. *Chaos, Solitons and Fractals*, **15**, 161-172.

Amari, S. (1982). Differential geometry of curved exponential families-curvatures and information loss. *Annals of Statistics*, **10**, 357-385.

Amari, S. (1985). Differential geometrical methods in statistics. *Springer Lecture Notes in Statistics*, **28**.

Arwini, K. A. and Dodson, C. T. J. (2007). *Alpha-geometry of the Weibull manifold*, Second Basic Sciences Conference, Al-Fatah University, Tripoli, Libya 4-8.

Cho, B. S. and Baek, H. Y. (2006). Geometric properties of t-distribution. *Honam Mathematical Journal*, **28**, 433-438.

Efron, B. (1975). Defining the curvature of a statistical problem. *Annals of Statistics*, **3**, 1109-1242.

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

Kass, R. E. (1989). The geometry of asymptotic inference. *Statistical Science*, **4**, 188- 219.

Rao, C. R. (1945). Information and the accuracy attainable in estimation of statistical parameters. *Bulletin of the Calcutta Mathematical Society*, **37**, 81-91.

Wagenaar, D. A. (1998). *Information geometry for neural networks*, Term paper for reading course with A. C. C. Coolen, King's College, London.