On Finding Geodesic Equation of Student T Distribution

William W. S. Chen[1]

[1] Department of Statisticse, The George Washington University, USA

Correspondence: William W. S. Chen , Department of Statistics, The George Washington University, Washington D. C. 20013, USA. E-mail: williamwschen@gmail.com

Abstract

Student t distribution has been widely applied in the course of statistics. In this paper, we focus on finding a geodesic equation of the two parameter student t distributions. To find this equation, we applied both the well-known Darboux Theorem and a triply of partial differential equations taken from Struik D. J. (Struik, D. J., 1961) or Grey A (Grey A., 1993), As expected, the two different approaches reach the same type of results. The solution proposed in this paper could be used as a general solution of the geodesic equation for the student t distribution.

Mathematical Subject Classification 62E99

Keywords: Darboux theorem, geodesic equation, small sample, size, student t distribution, triply partial differential equation.

1. Introduction

The student t distribution was first discovered by W.S. Gosset. Since the Irish brewery for which Gosset was working did not want the other breweries to know the statistical method they were using, Gosset published under the pseudonym of a student. Most statistical textbooks describe the t distribution in the following way: If $X_1, X_2, \ldots X_n$ are

independent, identically distributed, random variables, each having the same normal distribution with the expected value u and standard deviation v, then $\sqrt{n}(\overline{X} - u)/v$ has a unit normal distribution. This statistic can be used in the construction of tests and confidence intervals relating to the value of u, provided that v is known. If v is not known, it is reasonable to replace it by the sample estimator " s ", given the statistic $T = \sqrt{n}(\overline{X} - u)/s$. This process has been used for some time without allowing for differences between the distribution of $\sqrt{n}(\overline{X} - u)/v$ and $\sqrt{n}(\overline{X} - u)/s$. Statisticians realized that the two distributions are not identical, but the determination of the actual distribution had difficulties. Gosset obtained the distribution of $T' = T/\sqrt{n-1}$ and gave a short table of it's cumulative distribution function. We can show that T' is distributed as a ratio of a unit normal variable, z, and Chi, $\chi_{(n-1)}$, where the two variables are mutually independent. The divisor $\sqrt{n-1}$ was introduced by Fisher(1925a) who defined t with v degree of freedom as the distribution of $t_v = z(\frac{\chi_v}{v})^2$. This quantity is usually called student t and the corresponding distribution is called the student t distribution. In this paper, we used two different algorithms to find the geodesic equation of the student t distribution.

2. List the Fundamental Tensor

The probability density function for the student t distribution is given by:

$$f(x) = \frac{1}{v} \frac{\Gamma(\frac{r+1}{2})}{\sqrt{\pi r}\,\Gamma(\frac{r}{2})} [1 + \frac{1}{r}(\frac{x-u}{v})^2]^{-\frac{r+1}{2}} \qquad x \in R, \quad (u,v) \in R \times R_+ \} .$$

where u is a location parameter, v is a scale parameter and r is defined as the degree of freedom.

Define

$$a = \frac{1}{r}, \qquad b = \frac{r+1}{2}, \qquad c_r = \frac{\Gamma(\frac{r+1}{2})}{\sqrt{\pi r}\,\Gamma(\frac{r}{2})}.$$

Then,

$$\ln f(x) = \ln c_r - b\ln(1 + a(\frac{x-u}{v})^2) - \ln v. \tag{2.1}$$

From above equation (2.1), we derive the first and second partial derivatives:

$$\frac{\partial \ln f}{\partial u} = \frac{2ab(x-u)}{v^2 + a(x-u)^2},$$
$$\frac{\partial^2 \ln f}{\partial u^2} = \frac{2ab(a(x-u)^2 - v^2)}{(a(x-u)^2 + v^2)^2}. \tag{2.2}$$

$$\frac{\partial \ln f}{\partial v} = \frac{2ab(x-u)^2 v^{-3}}{1 + a(x-u)^2 v^{-2}} - v^{-1},$$
$$\frac{\partial^2 \ln f}{\partial v^2} = v^{-2}\left[1 + \frac{-6ab(x-u)^2 v^{-2}(1 + a(x-u)^2 v^{-2}) + 4a^2 b(x-u)^4 v^{-4}}{(1 + a(x-u)^2 v^{-2})^2}\right] \tag{2.3}$$

$$\frac{\partial^2 \ln f}{\partial v \partial u} = \frac{-4abv(x-u)}{(a(x-u)^2 + v^2)^2} \tag{2.4}$$

Then we take the expected values of (2.2),(2.3)and (2.4) to derive the metric tensor components for the student t distribution:

$$E = -E(\frac{\partial^2 \ln f}{\partial u^2}) = -2ab(\frac{-r}{v^2(r+3)}) = \frac{r+1}{v^2(r+3)}, \tag{2.5}$$

$$F = -E(\frac{\partial^2 \ln f}{\partial v \partial u}) = 0, \tag{2.6}$$

$$G = -E(\frac{\partial^2 \ln f}{\partial v^2}) = \frac{-1}{v^2}(1 - 3\frac{r+1}{r+3}). = \frac{1}{v^2}(\frac{2r}{r+3}) \tag{2.7}$$

More detailed proof for equation (2.5), (2.6) and (2.7) can be found in Chen W.W.S.[3]. Using the above results we can further derive their derivatives and six well known Christoffel Symbols as follows:

$$E_u = 0; \quad E_v = \frac{-2(r+1)}{v^3(r+3)}; \quad G_u = 0; \quad G_v = \frac{-4r}{v^3(r+3)}$$

$$EG - F^2 = EG = \frac{2r(r+1)}{v^4(r+3)^2}; \quad F = 0; \quad F_u = 0; \quad F_v = 0 \tag{2.8}$$

$$\Gamma^1_{11} = \frac{E_u}{2E} = 0; \qquad \Gamma^2_{12} = \frac{G_u}{2G} = 0;$$

$$\Gamma^2_{11} = \frac{-E_v}{2G} = \frac{2(r+1)}{v^3(r+3)} \frac{v^2(r+3)}{4r} = \frac{r+1}{2rv}$$

(2.9)

$$\Gamma^1_{22} = \frac{-G_u}{2E} = 0; \qquad \Gamma^1_{12} = \frac{E_v}{2E} = \frac{-2(r+1)}{v^3(r+3)} \frac{v^2(r+3)}{2(r+1)} = \frac{-1}{v}$$

$$\Gamma^2_{22} = \frac{G_v}{2G} = \frac{-4r}{v^3(r+3)} \frac{v^2(r+3)}{4r} = \frac{-1}{v}$$

(2.10)

3. The Geodesic Equation

To find the geodesic equation of the student t distribution, we solve a triply of partial differential equations, given in the appendix I. We seek its solution in the following section.

$$\frac{d^2u}{ds^2} - \frac{2}{v} \frac{du\,dv}{ds\,ds} = 0$$

(3.1)

$$\frac{d^2v}{ds^2} + \frac{r+1}{2rv}\left(\frac{du}{ds}\right)^2 - \frac{1}{v}\left(\frac{dv}{ds}\right)^2 = 0$$

(3.2)

$$ds^2 = \frac{r+1}{v^2(r+3)}du^2 + \frac{2r}{v^2(r+3)}dv^2$$

(3.3)

We only need two out of above three equations to find the student t model geodesic equation. We will choose the first (3.1) and the third (3.3) equations. To simplify the notation, we let

$$p = \frac{du}{ds} \qquad \text{then} \quad \frac{dp}{ds} - \frac{2}{v}\left(p\frac{dv}{ds}\right) = 0$$

(3.4)

$$\frac{\frac{dp}{ds}}{p} - \frac{2}{v}\frac{dv}{ds} = 0$$

(3.5)

Integrate (3.5) on both sides with respect to p, to get

$$\ln p - 2\ln v = C_1 \qquad \text{or} \qquad \ln pv^{-2} = C_1$$

$$pv^{-2} = e^{C_1} = A_t \qquad \frac{du}{ds} = A_t v^2$$

Where C_1 is an arbitrary constant and A_t is a temperary constant. We will define its value later. Finally, we derive:

$$ds^2 = \frac{du^2}{A_t^2 v^4}$$

(3.6)

Substitute equation (3.6) into equation (3.3)

$$\frac{du^2}{A_t^2 v^4} = \frac{r+1}{v^2(r+3)}du^2 + \frac{2r}{v^2(r+3)}dv^2$$

$$[(r+3) - A_t^2 v^2 (r+1)]du^2 = 2rA_t^2 v^2 dv^2$$

$$du^2 = \frac{2rA_t^2 v^2 dv^2}{(r+3) - A_t^2 v^2 (r+1)} \tag{3.7}$$

Then take the square root of equation (3.7), to get

$$du = \frac{\pm\sqrt{2r} A_t v dv}{\sqrt{(r+3) - A_t^2 v^2 (r+1)}}$$

Integrate the equation on both sides to derive the geodesic equation of the student t distribution as follows:

$$\pm u \pm \int \frac{\sqrt{2r} A_t v dv}{\sqrt{(r+3) - A_t^2 v^2 (r+1)}} = B \tag{3.8}$$

Where A_t and B are an arbitrary constant.

Alternatively, we can find the geodesic equation of the student t distribution by solving one partial differential equation. This idea originated from the French mathematician Darboux and is now known as Darboux's theory.

$$\nabla Z = 1; \quad \text{i.e.} \quad \frac{E Z_v^2 - 2F Z_u Z_v + G Z_u^2}{EG - F^2} = 1$$

where

$$E = -E(\frac{\partial^2 \ln f}{\partial u^2}) = \frac{r+1}{v^2(r+3)}, \quad F = -E(\frac{\partial^2 \ln f}{\partial v \partial u}) = 0,$$

$$G = -E(\frac{\partial^2 \ln f}{\partial v^2}) = \frac{-1}{v^2}(1 - 3\frac{r+1}{r+3}). = \frac{1}{v^2}(\frac{2r}{r+3})$$

$$\nabla Z = 1; \text{ equivalent to :} \quad \frac{r+1}{v^2(r+3)} Z_v^2 + \frac{2r}{v^2(r+3)} Z_u^2 = \frac{2r(r+1)}{v^4(r+3)^2}$$

$$or \quad (r+1)Z_v^2 + 2rZ_u^2 = \frac{2r(r+1)}{v^2(r+3)} \tag{3.9}$$

From (3.9), we derive

$$Z_u^2 = \frac{r+1}{v^2(r+3)} - \frac{r+1}{2r} Z_v^2 = A^2 \tag{3.10}$$

Then (3.10) separated into two parts as follows;

Part 1:

$$Z_u^2 = A^2, \quad Z_u = \pm A \qquad or \quad Z = \pm Au \tag{3.11}$$

Part 2:

$$\frac{r+1}{v^2(r+3)} - \frac{r+1}{2r} Z_v^2 = A^2$$

$$Z_v = \pm[\frac{2r}{r+1}(\frac{r+1}{v^2(r+3)} - A^2)]^{\frac{1}{2}}$$

$$Z = \pm \int [\frac{2r}{r+1}(\frac{r+1}{v^2(r+3)} - A^2)]^{\frac{1}{2}} dv \qquad (3.12)$$

We put (3.11) and (3.12) together to find one general solution for equation (3.9):

$$Z = \pm Au \pm \int [\frac{2r}{r+1}(\frac{r+1}{v^2(r+3)} - A^2)]^{\frac{1}{2}} dv \qquad (3.13)$$

Now, to find the Geodesic equation of the student t distribution we only need to differentiate equation (3.13) by the constant A.

$$i.e. \; \frac{\partial Z}{\partial A} = B.$$

$$\pm u \pm \int \frac{\sqrt{2r}Avdv}{\sqrt{\frac{(r+1)^2}{r+3} - A^2v^2(r+1)}} = B \qquad (3.14)$$

We found that equations (3.8) and (3.14) are of the same type. The difference is only by a constant. The difference may be adjusted by using the constant A_t.

4. Concluding Remarks

Rao(1945) presented a "geodesic distance" (or "Rao distance"), which has outstanding theoretical properties.

However,it was based on a demanding differential geometrical approach. This "geodesic distance" concept, a generalization of the well-known Mahalanobis distance, had to wait until more interest in differential geometry was raised by Efron. This paper uses a simple econometric problem to demonstrate the reason the student t geodesic equation is useful. Let A be a stock represented by its yield $y \sim N(\mu, \sigma_0^2)$, with the unknown expected yield μ and the known risk σ_0^2. Assume we

want to test $H_0 \; \mu = \mu_0 \; versus \; H_a \; \mu \neq \mu_0$ where μ_0 is some specified value with a sample size of one. The optimal test in this situation has a critical region $H = \langle \bar{x} : / |\bar{x} - \mu_0| > t_{1-\alpha/2} \; \sigma_0 \rangle$. The test seeks to answer the question: Is the distance between the two normal populations $N(\mu_0, \sigma_0^2)$ and $N(\bar{x}, \sigma_0^2)$, big enough to reject H_0 ? The answer depends on σ and on the distributional assumption. If we let σ tend to infinitely large,then the distance between $N(\mu_0, \sigma^2)$ and $N(\bar{x}, \sigma^2)$ should converge to zero. For σ to tend to zero,then the distance will become infinitely large. For this reason, the family of t distribution should not be identified with a flat plane but with a curved surface. This is why the geodesic equation should be used instead of the t distribution.

5. References

Atkinson, C., & Mitchell, A. F. S. (1981). Rao's distance measure. *Sankhya A, 43*, 345-65.

Chen W. W. S. (2014). A note on finding geodesic equation Of two parameters gamma distribution. *Applied Mathematics, 5*, 3511-3517. https://doi.org/10.4236/am.2014.521328

Chen W. W. S. (2016). On finding geodesic equation of two parameters *Inverse Gaussian distribution. British Journal of Mathematics & Computer Science, 15*(1), 1-8, Article no. BJMCS. 24577. https://doi.org/10.9734/bjmcs/2016/24577

Chen W. W. S. (2016). On the comparison of statistical Curvature with Gaussian curvature. I*nternational Journal of Mathematics and Statistics Studies, 4*(6), 13-23, December.

Grey A. (1993). *Modern differential geometry of curves and surfaces.* CRC Press, Inc. Boca Raton.

Jensen, U. (1995). *A review of the derivation and Calculation of Rao distances with an application to portfolio theory*. 413-462, Blackwell, Cambridge.

Kass, R. E., & Vos, P. W. (1997). *Geometrical foundation of asymptotic inference. John Wiley & Son, Inc. New York*. https://doi.org/10.1002/9781118165980

Oller, J. M. (1986). Information metric extreme value and logistic probability distributions. *Sankya A, 49*, 17-23.

Struik, D. J. (1961). *Lectures on classical differential geometry.* Second Edition. Dover Publications, Inc.

Appendix I

We list the six well known Christoffel Symbols as follows. For detail derivation see Struik or Grey.

$$\Gamma_{11}^1 = \frac{GE_u - 2FF_u + FE_v}{2(EG - F^2)}, \qquad \Gamma_{12}^2 = \frac{EG_u - FE_v}{2(EG - F^2)}$$

$$\Gamma_{11}^2 = \frac{2EF_u - EE_v - FE_u}{2(EG - F^{2)})}, \qquad \Gamma_{22}^1 = \frac{2GF_v - GG_u - FG_v}{2(EG - F^2)}$$

$$\Gamma_{12}^1 = \frac{GE_v - FG_u}{2(EG - F^2)}, \qquad \Gamma_{22}^2 = \frac{EG_v - 2FF_v + FG_u}{2(EG - F^2)}$$

In general, the solution of the geodesic equation depends upon a pair of partial differential equations as below.

$$\frac{d^2u}{ds^2} + \Gamma_{11}^1 \left(\frac{du}{ds}\right)^2 + 2\Gamma_{12}^1 \left(\frac{du}{ds}\frac{dv}{ds}\right) + \Gamma_{22}^1 \left(\frac{dv}{ds}\right)^2 = 0$$

$$\frac{d^2v}{ds^2} + \Gamma_{11}^2 \left(\frac{du}{ds}\right)^2 + 2\Gamma_{12}^2 \left(\frac{du}{ds}\frac{dv}{ds}\right) + \Gamma_{22}^2 \left(\frac{dv}{ds}\right)^2 = 0$$

Copyrights