Journal of Positive School Psychology
2022, Vol. 6, No. 4, 6326 – 6330 http://journalppw.com

Using Hellinger Method To Evaluate The Distance For Continuous Distributions And Discrete Distributions

[1]Nada Mohammed Abbas, [2]Audie Sabrie

[1]*Department of Mathematics/Education College for Pure Sciences/ University of Babylon, Iraq, pure.nada.moh@uobabylon.edu.iq*
[2]*Department of Mathematics/ College of Basic Education University of Babylon, Iraq, and udaisabri51@gmail.com*

Abstract

Hillinger presented a method to calculating the distance between distributions of parameter by utilizing the concept of contrast function. In this work Hillinger method is used to measure the distance for continuous distributions and discrete distributions respectively, (i,e) between theire probability density functions of two gamma distributions , two exponential distributions as a special case of gamma distribution , also for two Poisson distributions as discrete distributions.

Keywords: Hellinger Method, continuous distributions and discrete distributions.

1. INTRODUCTION

T Several methods are presented by different researchers to measure the distance between two distributions . Assuming that all the information for erect such distances is included in the probability density function (p.d.f) of a random vector . Rao (1945) depend on Fisher information as a base to calculate the distance [3] , H.Jeffery [4] proposed a method for finding the distance between two distributions by constructing contrast functions the same as the symmetric Kullback leibher relative entropy , and it is studied by Atkinson and Michell (1981) admit us to define a distance on the parametric space have the feature of being invariant under any admissible transformation of the parameters . Rao distance between two points of is defined as their geodesic distance. Carles and Joseph (1985) and Gruber (2003) they using Rao's method to measure the distance between continuous distribution , Amari (1987) studied the connection of geometry with statistical inferences . For more detail see [7] , S.Eguchi (1992) presented a differential geometry as an approach to statistical inference and geometry of minimum contrast [6] .

In this paper Hillinger method is used to measure the distance between two continuous distributions and two discrete distributions respectively depending on the contrast functions since the contrast functions are distance like quantities which the a symmetric vicinity for two probability density functions on a statistical manifold and statistical model [6],[9] and it gives all the data of the statistical manifold if it exists[10] ,[1] .

The Riemannian metric induced by the contrast function by its second order derivative [6] on statistical model is the Fisher Riemann information matrix .

2. MATHEMATICAL BACKGROUND

Definition 2.1 :- [4]

A random variable is a function which can take on any value from the sample space and having range of some set of real numbers is known as the random variable of the experiment . Random variables can be classified as :

I. Discrete Random Variable .

II. Continuous Random Variable.

Definition 2.2:- [4]

These are the random variable ,which can take on only finite number of values in a finite observation interval .So we can say , that discrete random variable has distinct values that can be counted .

Definition 2.3:- [11]

A random variable that takes on an infinite number of values is known as a continuous random variable.

Definition 2.4:- [11]

Let be a continuous random variable ,the probability density function (p.d.f) of is denoted by or and it satisfies the following properties :

i. $f(x) \geq 0, \forall x, -\infty < x < \infty$

ii. $\int_{-\infty}^{\infty} f(x)dx = 1$

iii. $p(a \leq X \leq b) = \int_a^b f(x)dx$

Definition2. 5:- [8]

Let X be a random variable from a distribution with pdf $f(x;\theta)$ of a continuous type such that the parameter θ does not appear in endpoints of the interval in which $f(x;\theta) > 0$ and that we can interchange integration and differentiation with respect to θ .the fisher information , $I(\theta)$ in a single observation X about θ is given by

$$I(\theta) = \int_{-\infty}^{\infty}\left[\frac{\partial \ln f(x;\theta)}{\partial\theta}\right]^2 f(x;\theta)dx = E\left[\left(\frac{\partial \ln f(x;\theta)}{\partial\theta}\right)^2\right]$$

That is , $I(\theta)$ is the expected value of the square of the random variable $\frac{\partial \ln f(x;\theta)}{\partial\theta}$

Definition 2.6:- [5]

$$D_{KL}(P\|Q) = E_{X \sim P}\left(\log\frac{P(X)}{Q(X)}\right) = -E_{X \sim P}\left(\log Q(X)\right) - H(X \sim P)$$

.

The term $-E_{X \sim P}(\log Q(X))$ is called the cross entropy with the following properties :-

a. $D_{KL}(P\|Q) \geq 0, \forall P, Q$

b. $D_{KL}(P\|Q) = 0$ if and only if P and Q are equal almost surely

c. KL divergence is asymmetric (i.e) $D_{KL}(P\|Q) \neq D_{KL}(Q\|P)$.

Definition 2.7:- [6]

Topological space (X, Θ_X) is called second countable if it has a countable basis

Definition 2.8:-[7]

A topological space M is called n- dimensional topological manifold if M is locally euclidean at any point , Hausdorff and second countable .

Definition 2.9:- [1]

A second countable Hausdorff space M together with an n-dimensional maximal atlas is called smooth manifold of dimension dim(M)= n .

Definition 2.10:- [9]

Let W and Z be smooth manifolds . A continuous map $f : W \to Z$ is called smooth if for all charts (φ, U) of W , (ψ, V) of Z $\Psi \circ f \circ \varphi^{-1} : \varphi(U \cap f^{-1}(V)) \to \Psi(V)$ is smooth

Definition 2.11:- [6]

The set of all tangent vectors to M at a point p is the tangent space $(T_p M)$.

Definition 2.12:- [2]

Let M be smooth manifold . A riemannian metric g on M is a smooth family of inner products on the tangent spaces of M . Namely , g associates to each $p \in M$ a positive definite symmetric bilinear on $T_p M$,

$g_p : T_pM \times T_pM \to R$ and the smoothness condition on g refers to the fact the function $p \in M \to g_p(X_p, Y_p) \in R$ is smooth .

Definition 2.13 :- [2]

In the mathematical field of differential geometry , a metric tensor is a type of function which takes as input a pair of tangent vectors and at a point of a surface (or higher dimensional differentiable manifold) and produces a real number scalar in a way that generalizes many of the familiar properties of the dot product of vectors in Euclidean space .

In the same way as a dot product , metric tensors are used to define the length of and angle between tangent vectors . Through integration , the metric tensor allows one to define and compute the length of curves on the manifold .

Definition 2.14:- [5]

Fisher information metric is a metric tensor for a statistical differential manifold such that the coefficients of the expected Fisher information matrix as equal to the coefficients of the first fundamental form (Riemannian metric)on the space of probabilities . Fisher information metric can be used to cclculate the informational difference between measurement and it takes the following form

$$g_{ij} = \int \frac{\partial^2 \log f(x,\theta)}{\partial\theta_i \partial\theta_j} f(x,\theta) dx$$

Definition 2.15:-

Let E be an open set in R^n , $\xi_1, \xi_2 \in \mathrm{E}$ and let S be a smooth manifold . A contrast function on S is a smooth mapping $D_S(\Box\|\Box) : S \times S \to R$, such that any parametrization $\Phi : E \to S$ makes $D(\xi_1\|\xi_2) = D_S(\Phi(\xi_1)\|\Phi(\xi_2))$ a contrast function on E , where $D_S(\bullet\|\bullet)$ represented the separation notation between two points . contrast function $D_S(\bullet\|\bullet)$ has the following properties :-

a) Positive : $D(\xi_1, \xi_2) \geq 0, \forall \xi_1, \xi_2 \in S$

b) Non- degenerate : $D(\xi_1, \xi_2) = 0 \Leftrightarrow \xi_1 = \xi_2$

c) The first variation along the diagonal $\{\xi_1 = \xi_2\}$ vanishes :

$$\partial_{\xi_1^i} D(\xi_1\|\xi_2)_{|\xi_1=\xi_2} = \partial_{\xi_2^i} D(\xi_1\|\xi_2)_{|\xi_1=\xi_2} = 0$$

d)The Hessian along the diagonal $\xi^0 = \xi$,

$$g_{ij}(\xi_1) = \partial_{\xi_2^i} \partial_{\xi_2^j} D(\xi_1\|\xi_2)_{|\xi_2=\xi_1}$$

is strictly positive definite and smooth with respect to ξ_1.

Note 2.16:- The operator $D_f(\bullet\|\bullet)$ is a contrast function on the statistical model $S = \{p_\xi\}$ since it satisfies the properties in definition 2.15, for more detail see [5],[10] .

Definition 2.17:- [10]

Let $f(u) = 4(1-\sqrt{u})$ be contrast function and let $q(x), p(x)$ any (p.d.f) , then the Hellinger distance which is denoted by $H(p,q)$ is

$$H(p,q) = \sqrt{4 - 4\int_0^\infty \sqrt{p(x)q(x)}dx} .$$

Now we presented the following examples :

i) For continuous distributions :-

Example 2.18:-

Let two Gamma distribution

$$g(x) = \frac{1}{\Gamma\alpha\beta^\alpha} e^{-\frac{t}{\beta}} t^{\alpha-1}$$ and

$$z(x) \frac{1}{\Gamma a b^a} e^{-\frac{t}{b}} t^{a-1}, \forall t \geq 0, \alpha, \beta, a, b > 0$$ if we assume that $\alpha = a = 1$ then the Hillinger distance given by $H(g,z) = \sqrt{4 - \frac{8\sqrt{\alpha\beta}}{\alpha+\beta}}$.

As a special case of the Gamma distribution play an important role in statistics is the exponential distribution which possesses the memoryless property for $\alpha = a = 1$

Example 2.19:-

Let two exponential distributions with

$p(x)=\frac{1}{\alpha}e^{-\frac{1}{\alpha}x}$ and

$q(x)=\frac{1}{\beta}e^{-\frac{1}{\beta}x}$ $,x\geq 0,\alpha,\beta>0$

Then the Hellinger distance $H^2(p,q)$ is given by

$$H^2(p,q)=4-4\int_0^\infty\sqrt{p(x)q(x)}dx$$

$$=4-4\int^\infty\sqrt{(\frac{1}{\alpha}e^{-\frac{1}{\alpha}x})(\frac{1}{\beta}e^{-\frac{1}{\beta}x})}dx$$

$$=4-4\sqrt{\frac{1}{\alpha\beta}\frac{2\alpha\beta}{\beta+\alpha}}$$

Hence the Hellinger is $H(p,q)=2\sqrt{1-\frac{2\sqrt{\alpha\beta}}{\alpha+\beta}}$.

ii) For discrete distributions:-

Example 2.20:-

Consider two Poisson distributions where the p.d.f ' s are $r(t)=\frac{10^t}{t!}e^{-10}$ and $s(t)=\frac{8^t e^{-8}}{t!}$,

then $H(r,s)=2\left(1-\sum_{t\geq 0}\sqrt{r(t)s(t)}\right)^{\frac{1}{2}}$

$$=2\left(1-\sum_{t\geq 0}\sqrt{(\frac{10^t e^{-10}}{t!})(\frac{8^t e^{-8}}{t!}}\right)^{\frac{1}{2}}$$

Now , we calculate as following:

$$\sum_{t\geq 0}\sqrt{(\frac{10^t e^{-10}}{t!})(\frac{8^t e^{-8}}{t!})}=\sum_{t\geq 0}\frac{(\sqrt{180})^t e^{-9}}{t!}$$

$$=e^{-9}e^{\sqrt{180}}\sum_{t\geq 0}\frac{(\sqrt{180})^t}{t!}e^{-\sqrt{180}}$$

$$=e^{\sqrt{180}-9}$$

So,

$$H(r,s)=2\left(1-\sum_{t\geq 0}\sqrt{r(t)s(t)}\right)^{\frac{1}{2}}=2\sqrt{1-e^{\sqrt{180}-9}}$$

.

3. Conclusion

Many geometrical procedures from different researches are presented to measure the dissimilarity between two probability density functions (p.d.f). In current study Hillingers approach is used to evalute the distance between two continuous distributions and two discrete distributions respectively as an application . Hellingers distance depend on the concept of contrast function on space of density functions , since a contrast functional on statistical manifolds are a natural extensions of Kullback- Leibler relative entropy from statistical models. [9]

Reference

[1] O'neill, Barrett. Elementary differential geometry. Elsevier, 2006.

[2] Lipschutz, Martin M. Schaum's outline of differential geometry. McGraw Hill Professional, 1969.

[3] Atkinson, Colin, and Ann FS Mitchell. "Rao's distance measure." Sankhyā: The Indian Journal of Statistics, Series A (1981): 345-365.

[4] Gruber.M.H.J,(2003)"Some Interesting connections between statistics and Differential Geometry"

[5] S. -I. Amari, O. E. Barndorff-Nielsen, R. E. Kass, S. L. Lauritzen and C. R. Rao,Differential Geometry in Statistical Inferences, IMS Lecture Notes-Monograph Series, 10, Institute of Mathematical Statistics, Hayward, 1987.

[6] S. Eguchi, Geometry of minimum contrast, Hiroshima Math. J. 22 (1992), 631-647.

[7] Do Carmo, Manfredo P. Differential forms and applications. Springer Science & Business Media, 1998.

[8] Costa, Sueli IR, Sandra A. Santos, and Joao E. Strapasson. "Fisher information distance: A geometrical reading." Discrete Applied Mathematics 197 (2015): 59-69

[9] Burns, K., & Gidea, M. (2005). Differential Geometry and Topology: With a view to dynamical systems. Chapman and Hall/CRC.

[10] Matumoto, Takao. "Any statistical manifold has a contrast function—On the -functions taking the minimum at the diagonal of the product manifold." Hiroshima Math. J 23.2 (1993): 327-332

[11] Freund, J. E., & Walpole, R. E. (1980). Mathematical statistics 3rd edn.