On information geometry of the inverse Gaussian distribution

Oğuzhan Bahadır, Umut Çako

Abstract. The inverse Gaussian distribution is a common distribution in many different fields, such as life tests, psychology, demography, linguistics, environment and finance. In this paper, we give the geometric interpretation of the inverse Gaussian distribution from the viewpoint of information geometry. We obtain the Fisher information matrix, Riemannian connections and Gaussian curvature of the inverse Gaussian distribution. Then we investigate the KL-divergence, J-divergence, geodesic distance and the relations between of them.

M.S.C. 2010: 53C22, 60D05, 62E99.
Key words: Divergence; geodesic distance; information geometry; inverse Gaussian distribution; Riemannian connections.

1 Introduction

The inverse Gaussian distribution has a history dating back to 1915 when Schrödinger and Smoluchowski presented independent derivations of the density of the first passage time distribution of Brownian motion with positive drift [10]. Tweedie has shown the inverse relationship between the cumulant generating function of the first passage time distribution and that of the normal distribution, Tweedie proposed the name inverse Gaussian (IG) for the first passage time distribution [11]. Wald has derived the limiting form of IG [12]. Therefore, it is also called as Wald's distribution, particularly in the Russian literature [6].

The IG is used to model non-negative, positively skewed data and has a wide variety of applications in business, survival analysis, finance, medicine, and even in labor dispute resolution. The tail of the distribution decreases slowly compared to the normal distribution. Therefore, it is suitable for modeling phenomena where there is a greater likelihood of getting extremely large values compared to the normal distribution.

In the present paper, we consider the geometric structure of the IG. Firstly, we give the Fisher information matrix, the Riemannian connections, the Gaussian curvature under the coordinate system (μ, λ). We can see that the IG is an exponential

BSG Proceedings, vol. 28 (2021), pp. 12-19.

family distribution. Secondly, we give the KL-divergence and J-divergence in this distribution. Furthermore we give the geodesic distance and the relations between of them.

2 Preliminaries

Definition 2.1. A continuous random variable $X \subset \mathbb{R}^n$ is a random variable taking a continuous range of values; its probability distribution has a probability density function (pdf), $p(x)$, which is a non-negative function, integrable with respect to the Lebesgue measure on X, i.e. a function $p : X \to \mathbb{R}$ satisfying [4]

$$p(x) \geq 0, \;\; (\forall x \in X) \;\; and \;\; \int p(x)dx = 1.$$

Consider a family S of probability distributions on X. Suppose each element of S, a probability distribution, may be parameterized using n real-valued variables $(\theta^1, \theta^2, \cdots, \theta^n)$ so that

$$S = \left\{p_\theta = p(x;\theta) \mid \theta = (\theta^1, \theta^2, \cdots, \theta^n) \in \Theta\right\}$$

where Θ is a subset of $\mathbb{R}^n$ and the mapping $\theta \to p_\theta$ is injective. We call such S an n-dimensional statistical model on X [2].

Definition 2.2. Consider $n+1$ real-valued smooth functions $C(x), F_i(x)$ on X such that $C(x), F_1(x), \cdots, F_n(x)$ are linearly independent. Then define the normalization function

$$\psi(\theta) = ln\left(\int e^{C(x)+\theta^i F_i(x)}dx\right)$$

and consider the exponential family of probability densities

$$p_\theta(x) = h(x)e^{\theta^i F_i(x)-\psi(\theta)} \tag{2.1}$$

with $h(x) = e^{C(x)}$, $x \in X$. The statistical model $S = \{p(x;\theta)\}$, is called an exponential family and θ^i are its natural parameters [4].

Definition 2.3. Let $S = \{p_\theta \mid \theta \in \Theta\}$ be an n-dimensional statistical model. Given a point θ, the Fisher information matrix of S at θ is the $n \times n$ matrix $[g_{ij}(\theta)]$, where the $(i,j)^{th}$ element $g_{ij}(\theta)$ is defined by the equation below

$$g_{ij}(\theta) = E_\theta\left[\partial_i l_\theta \partial_j l_\theta\right] = \int \partial_i l(x;\theta)\partial_j l(x;\theta)p(x;\theta)dx$$

where $\partial_i = \frac{\partial}{\partial\theta^i}$, $l_\theta = l(x;\theta) = lnp(x;\theta)$, and E_θ denotes the expectation with respect to the distribution p_θ. We note that it is possible to write g_{ij} as [2]

$$g_{ij}(\theta) = -E\left[\partial_i\partial_j lnp(x;\theta)\right]. \tag{2.2}$$

Definition 2.4. The metric tensor $g_{ij}(\theta)$ brings about a natural affine connection, whose coefficients are given by

$$\Gamma_{ij}^{k} = \frac{1}{2}\sum_{r} g^{kr}\left(\partial_j g_{ri} + \partial_i g_{rj} - \partial_r g_{ji}\right) \tag{2.3}$$

which is called Christoffel symbol (g^{kr} is the inverse of the Fisher information matrix g_{kr}) [3].

Definition 2.5. The Riemann curvature tensor is defined by [8]

$$R_{ijk}^{l} = \partial_j \Gamma_{ki}^{l} - \partial_k \Gamma_{ji}^{l} + \sum_{m} \Gamma_{jm}^{l}\Gamma_{ki}^{m} - \Gamma_{km}^{l}\Gamma_{ji}^{m}. \tag{2.4}$$

The Riemann tensor can be written using its covariant components as the following:

$$R_{ijkl} = \sum_{m} g_{im} R_{jkl}^{m}. \tag{2.5}$$

We can obtain a tensor of rank two as

$$R_{ij} = \sum_{k} g^{kl} R_{kilj}. \tag{2.6}$$

This tensor is called the Ricci tensor. The scalar curvature R is defined by

$$R = \sum_{ij} g^{ij} R_{ij}. \tag{2.7}$$

If $n = 2$, then

$$K = \frac{1}{2} R \tag{2.8}$$

is called the Gaussian curvature [5].

Definition 2.6. When S is a Riemannian manifold, we can define a Riemannian geodesic. The distance along a geodesic, between two points p and q is defined by the integral

$$S(p,q) = \left| \int_{p}^{q} \sqrt{\sum_{ij} g_{ij}(\theta) d\theta^i d\theta^j} \right|. \tag{2.9}$$

Definition 2.7. The divergence function is given by

$$D_{KL}(p \parallel q) = \int p(x) ln \frac{p(x)}{q(x)} dx \tag{2.10}$$

which is known as the Kullback-Leibler divergence [1]. It is well known that $D_{KL}(p \parallel q)$ is non-negative, additive but not symmetric. To obtain a symmetric measure, we can define

$$J(p,q) = D_{KL}(p \parallel q) + D_{KL}(q \parallel p) \tag{2.11}$$

which is called the J-divergence [9].

3 Geometric structure of IG statistical model

Let $X = (0, \infty)$ and $\Theta = (0, \infty) \times (0, \infty)$. The probability distribution function of an IG with two-parameter $\theta = (\theta^1, \theta^2) = (\mu, \lambda) \in \Theta$ is given by the formula

$$p(x; \theta) = p(x; \mu, \lambda) = \left(\frac{\lambda}{2\pi x^3}\right)^{1/2} \cdot exp\left\{-\frac{\lambda(x-\mu)^2}{2\mu^2 x}\right\} \tag{3.1}$$

where $x \in X$. The parameter μ is the mean of the distribution and λ is a scale parameter [7]. Using the well-known properties of a pdf

$$\int p(x)dx = 1, \quad \int xp(x)dx = E[x], \quad \int \frac{1}{x}p(x)dx = E\left[\frac{1}{x}\right] \tag{3.2}$$

and by a straightforward calculation, we get

$$E[x] = \mu, \quad E\left[\frac{1}{x}\right] = \frac{1}{\mu} + \frac{1}{\lambda}. \tag{3.3}$$

Proposition 3.1. *The IG is an exponential family distribution.*

Proof. The log-likelihood function of IG is

$$lnp(x; \theta) = lnp(x; \mu, \lambda) = \frac{1}{2}ln\lambda - \frac{1}{2}ln(2\pi x^3) - \frac{\lambda x}{2\mu^2} + \frac{\lambda}{\mu} - \frac{\lambda}{2x}. \tag{3.4}$$

From (3.4), the IG pdf (3.1) can be rewritten as

$$p(x; \theta) = e^{lnp(x;\theta)} = \frac{1}{\sqrt{2\pi x^3}} \cdot exp\left\{-\frac{\lambda}{2\mu^2}x - \frac{\lambda}{2x} + \frac{\lambda}{\mu} + \frac{1}{2}ln\lambda\right\}.$$

Set

$$h(x) = \frac{1}{\sqrt{2\pi x^3}}$$

$$\theta^1 = -\frac{\lambda}{2\mu^2}, \quad \theta^2 = -\frac{\lambda}{2}$$

$$F_1(x) = x, \quad F_2(x) = \frac{1}{x}$$

then the potential function $\psi(\theta)$ can be written as

$$\psi(\theta) = -\frac{\lambda}{\mu} - \frac{1}{2}ln\lambda = -2\sqrt{\theta^1 \cdot \theta^2} - \frac{1}{2}ln(-2\theta^2)$$

so from (2.1), it is an exponential family distribution. □

Proposition 3.2. *The Fisher information matrix for the IG distribution is given by*

$$g_{ij}(\theta) = \begin{pmatrix} \frac{\lambda}{\mu^3} & 0 \\ 0 & \frac{1}{2\lambda^2} \end{pmatrix}. \tag{3.5}$$

Proof. Using (3.4), we have

$$\begin{aligned} \partial_\mu\partial_\mu lnp(x;\mu,\lambda) &= -\frac{3\lambda x}{\mu^4}+\frac{2\lambda}{\mu^3} \\ \partial_\lambda\partial_\lambda lnp(x;\mu,\lambda) &= -\frac{1}{2\lambda^2} \\ \partial_\lambda\partial_\mu lnp(x;\mu,\lambda) = \partial_\mu\partial_\lambda lnp(x;\mu,\lambda) &= \frac{x}{\mu^3}-\frac{1}{\mu^2}. \end{aligned} \tag{3.6}$$

Using (2.2), (3.3) and (3.6) we get the Fisher information matrix with respect to the coordinate system (μ, λ). □

From (3.5), we can get the inverse of the Fisher information matrix

$$g^{ij}(\theta) = \begin{pmatrix} \frac{\mu^3}{\lambda} & 0 \\ 0 & 2\lambda^2 \end{pmatrix}. \tag{3.7}$$

3.1 The Riemannian and scalar curvatures

From (2.3), (3.5) and (3.7) we can see that the Christoffel symbols as follows:

$$\begin{gathered} \Gamma^1_{11} = -\frac{3}{2\mu},\ \Gamma^1_{21} = \Gamma^1_{12} = \frac{1}{2\lambda},\ \Gamma^1_{22} = \Gamma^2_{21} = \Gamma^2_{12} = 0, \\ \Gamma^2_{11} = -\frac{\lambda^2}{\mu^3},\ \Gamma^2_{22} = -\frac{1}{\lambda}. \end{gathered} \tag{3.8}$$

Using (2.4) and (3.8), we get the nonzero components of the Riemannian curvature tensor

$$R^1_{212} = -R^1_{221} = -\frac{1}{4\lambda^2},\ R^2_{121} = -R^2_{112} = -\frac{\lambda}{2\mu^3}. \tag{3.9}$$

From (2.5) and (3.9), we have

$$R_{1212} = R_{2121} = -R_{1221} = -R_{2112} = -\frac{1}{4\lambda\mu^3}, \tag{3.10}$$

and the other components are equal zero. Using (2.6), we find that Ricci curvature tensor is given by

$$R_{12} = R_{21} = 0,\ R_{11} = -\frac{\lambda}{2\mu^3},\ R_{22} = -\frac{1}{4\lambda^2}. \tag{3.11}$$

Using (2.7) and (2.8), we have the following theorem.

Theorem 3.3. *The scalar curvature and the Gaussian curvature of IG are given by*

$$R = -1 \ \textit{and}\ K = -\frac{1}{2}. \tag{3.12}$$

4 The relation between the divergence and the geodesic distance

4.1 The geodesic distance

Let $p(x;\mu_p,\lambda_p)$ and $q(x;\mu_q,\lambda_q)$ be a IG pdf. From (2.9) and (3.5), we have the geodesic distance between two points,

$$S = S(p,q) = \left| \int_p^q \sqrt{\frac{\lambda}{\mu^3}(d\mu)^2 + \frac{1}{2\lambda^2}(d\lambda)^2} \right|. \tag{4.1}$$

We consider the following two cases:

(i). If μ is fixed, from (4.1), the geodesic distance is

$$S = \frac{1}{\sqrt{2}} \left| ln \frac{\lambda_q}{\lambda_p} \right|. \tag{4.2}$$

(ii). If λ is fixed, from (4.1), we obtain that in this case the geodesic distance is

$$S = 2 \left| \sqrt{\frac{\lambda}{\mu_p}} - \sqrt{\frac{\lambda}{\mu_q}} \right|. \tag{4.3}$$

4.2 The KL-divergence

Theorem 4.1. *The KL-divergence of IG pdf $p(x;\theta_p)$ and $q(x;\theta_q)$ satisfies the following equation:*

$$D_{KL}(p \parallel q) = \frac{1}{2} ln \frac{\lambda_p}{\lambda_q} - \frac{1}{2} + \frac{\lambda_q \mu_p}{2{\mu_q}^2} - \frac{\lambda_q}{\mu_q} + \frac{\lambda_q}{2\mu_p} + \frac{\lambda_q}{2\lambda_p}. \tag{4.4}$$

Proof. From (3.1), $p(x;\theta_p)$ and $q(x;\theta_q)$ are defined by the following equations:

$$\begin{aligned} p(x;\theta_p) &= \left(\frac{\lambda_p}{2\pi x^3}\right)^{1/2} \cdot exp\left\{-\frac{\lambda_p(x-\mu_p)^2}{2{\mu_p}^2 x}\right\}, \\ q(x;\theta_q) &= \left(\frac{\lambda_q}{2\pi x^3}\right)^{1/2} \cdot exp\left\{-\frac{\lambda_q(x-\mu_q)^2}{2{\mu_q}^2 x}\right\}, \end{aligned} \tag{4.5}$$

where $\theta_p = (\mu_p, \lambda_p)$, $\theta_q = (\mu_q, \lambda_q)$. Using (2.10), (3.2), (3.3) and (4.5)

$$\begin{aligned} D_{KL}(p \parallel q) &= \int_0^\infty p(x) \left(\frac{1}{2} ln \frac{\lambda_p}{\lambda_q} - \frac{\lambda_p(x-\mu_p)^2}{2{\mu_p}^2 x} + \frac{\lambda_q(x-\mu_q)^2}{2{\mu_q}^2 x} \right) dx \\ &= \frac{1}{2} ln \frac{\lambda_p}{\lambda_q} \int_0^\infty p(x)\, dx \\ &\quad - \frac{\lambda_p}{2{\mu_p}^2} \left(\int_0^\infty x p(x)\, dx - 2\mu_p \int_0^\infty p(x)\, dx + {\mu_p}^2 \int_0^\infty \frac{1}{x} p(x)\, dx \right) \\ &\quad + \frac{\lambda_q}{2{\mu_q}^2} \left(\int_0^\infty x p(x)\, dx - 2\mu_q \int_0^\infty p(x)\, dx + {\mu_q}^2 \int_0^\infty \frac{1}{x} p(x)\, dx \right). \end{aligned}$$

Thus, using (3.2) and (3.3), we arrive at the proof of the theorem. □

Similarly, we have

$$D_{KL}(q \parallel p) = \frac{1}{2} ln \frac{\lambda_q}{\lambda_p} - \frac{1}{2} + \frac{\lambda_p \mu_q}{2{\mu_p}^2} - \frac{\lambda_p}{\mu_p} + \frac{\lambda_p}{2\mu_q} + \frac{\lambda_p}{2\lambda_q}. \tag{4.6}$$

Special cases:

(i). If $\mu_p = \mu_q = \mu$, we can see that (4.4) reduced to

$$D_{KL}(p \parallel q) = \frac{1}{2} ln \frac{\lambda_p}{\lambda_q} - \frac{1}{2} + \frac{\lambda_q}{2\lambda_p}. \tag{4.7}$$

(ii). If $\lambda_p = \lambda_q = \lambda$, (4.4) takes the form

$$D_{KL}(p \parallel q) = \frac{\lambda(\mu_p - \mu_q)^2}{2{\mu_q}^2 \mu_p}. \tag{4.8}$$

Then from (4.2), (4.3), (4.7) and (4.8) we have the following theorem.

Theorem 4.2. *Let $p(x; \mu_p, \lambda_p)$ and $q(x; \mu_q, \lambda_q)$ be a IG pdf. Then the KL-divergence and the geodesic distance are connected in the following ways*

(i). when μ fixed,

$$D_{KL}(p \parallel q) = \frac{1}{2} \left(e^{\sqrt{2}S} - \sqrt{2}S - 1 \right),$$

(ii). when λ fixed,

$$D_{KL}(p \parallel q) = \frac{\left(\sqrt{\mu_p} + \sqrt{\mu_q}\right)^2}{8\mu_q} S^2.$$

4.3 The J-divergence

Theorem 4.3. *The J-divergence of IG pdf $p(x; \theta_p)$ and $q(x; \theta_q)$ satisfies the following equation:*

$$J(p, q) = -1 + \frac{\lambda_q \mu_p}{2{\mu_q}^2} - \frac{\lambda_q}{\mu_q} + \frac{\lambda_q}{2\mu_p} + \frac{\lambda_q}{2\lambda_p} + \frac{\lambda_p \mu_q}{2{\mu_p}^2} - \frac{\lambda_p}{\mu_p} + \frac{\lambda_p}{2\mu_q} + \frac{\lambda_p}{2\lambda_q}. \tag{4.9}$$

Proof. Using (2.11), (4.4) and (4.6), we arrive at the proof of the theorem. □

Special cases:

(i). If $\mu_p = \mu_q = \mu$, we can see that (4.9) reduced to

$$J(p, q) = -1 + \frac{\lambda_q}{2\lambda_p} + \frac{\lambda_p}{2\lambda_q}. \tag{4.10}$$

(ii). If $\lambda_p = \lambda_q = \lambda$, (4.9) takes the form

$$J(p, q) = \frac{\lambda}{2} (\mu_p + \mu_q) \left(\frac{\mu_p - \mu_q}{\mu_p \mu_q} \right)^2. \tag{4.11}$$

Then from (4.2), (4.3), (4.10) and (4.11) we have the following theorem.

Theorem 4.4. *Let $p(x;\mu_p,\lambda_p)$ and $q(x;\mu_q,\lambda_q)$ be a IG pdf. Then the J-divergence and the geodesic distance are connected in the following ways*

(i). when μ fixed,

$$J(p,q) = cosh(\sqrt{2}S) - 1,$$

(ii). when λ fixed,

$$J(p,q) = \frac{(\mu_p + \mu_q)\left(\sqrt{\mu_p} + \sqrt{\mu_q}\right)^2}{8\mu_p\mu_q} S^2.$$

References

[1] S. Amari, A. Cichocki, *Information geometry of divergence functions*, Bulletin of the Polish Academy of Sciences:Technical Sciences. 58, 1 (2010), 183-195.

[2] S. Amari, H. Nagaoka, *Methods of Information Geometry*, Oxford University Press, New York 2000.

[3] W.M. Boothby, *An Introduction to Differentiable Manifolds and Riemannian Geometry*, Elsevier, Singapore 2007.

[4] O. Calin, C. Udrişte, *Geometric Modeling in Probability and Statistics*, Springer, New York 2010.

[5] B. Chen, *Geometry of Submanifolds*, Marcel Dekker, New York 1973.

[6] R.S. Chhikara, J.L. Folks, *The inverse Gaussian distribution and its statistical application: a review*, J.R. Statist. Soc. B40, 3 (1978), 263-289.

[7] R.S. Chhikara, J.L. Folks, *The Inverse Gaussian Distribution Theory, Methodology and Applications*, Marcel Dekker, New York 1989.

[8] A. Gray, E. Abbena, S. Salamon, *Modern Differential Geometry of Curves and Surfaces with Mathematica*, Chapman & Hall/CRC, Boca Raton 2006.

[9] J. Lin, *Divergence measures based on the Shannon entropy*, IEEE Transactions on Information Theory. 37, 1 (1991), 145-151.

[10] V. Seshadri, *The Inverse Gaussian Distribution Statistical Theory and Applications*, Springer, New York 1999.

[11] M.C.K. Tweedie, *Inverse statistical variates*, Nature. 155, 3937 (1945), 453.

[12] A. Wald, *Sequential Analysis*, Wiley, New York 1947.

Authors' addresses:

Oğuzhan Bahadır
Department of Mathematics, Faculty of Arts and Sciences,
K.S.U. Kahramanmaras, Turkey.
E-mail: oguzbaha@gmail.com , obahadir@ksu.edu.tr

Umut Çako
Goksun Science High School, T.C. Ministry Of Education,
Goksun, 46600, Kahramanmaras, Turkey.
E-mail: umutcako@hotmail.com