



# What is the Fisher-Rao distance?

[WHAT IS?](#)

AUTHOR

Kisung You

PUBLISHED

February 5, 2025

## Introduction

How can we measure the difference between two probability distributions in a principled way? One approach is through the **Fisher-Rao distance**, a geometric measure of dissimilarity based on the **Fisher information matrix (FIM)**.

To understand this distance, we first examine the role of the FIM in statistical inference and its interpretation as a Riemannian metric. This metric structure naturally leads to the Fisher-Rao distance, which measures the shortest path—or geodesic—between distributions on the statistical manifold. However, directly computing this distance is often intractable.

To address this challenge, we introduce an alternative approach: the square-root transformation, which embeds probability densities into a Hilbert space. A key result follows from this transformation: the Fisher-Rao distance is exactly twice the geodesic distance between transformed densities on the unit sphere. This connection not only simplifies computations but also offers new insights into the geometry of probability distributions.

## Fisher Information Matrix

In a standard mathematical statistics course, one encounters the **Fisher information matrix (FIM)** for a probability density function  $p(x|\theta)$  with some parameter  $\theta \in \Theta \subset \mathbb{R}^d$ , defined as:

$$I(\theta) = \mathbb{E} \left[ \left( \frac{\partial}{\partial \theta} \log p(x|\theta) \right) \left( \frac{\partial}{\partial \theta} \log p(x|\theta) \right)^\top \right].$$

Alternatively, it can be expressed in terms of second derivatives:

$$I(\theta) = -\mathbb{E} \left[ \frac{\partial^2 \log p(x|\theta)}{\partial \theta \partial \theta^\top} \right].$$

The Fisher information matrix plays a central role in statistical inference. In **maximum likelihood estimation (MLE)**, it helps assess the asymptotic variance of the MLE:

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} \mathcal{N}(0, I^{-1}(\theta)).$$

The **Cramér-Rao bound** states that for any unbiased estimator  $\hat{\theta}$ :

$$\text{Cov}(\hat{\theta}) \succeq I^{-1}(\theta),$$

where  $A \succeq B$  indicates that  $A - B$  is positive semi-definite. In other words, the covariance matrix of any unbiased estimator is bounded below by the inverse of the FIM. In **Bayesian statistics**, the FIM is related to constructing priors<sup>1</sup>. The Jeffreys prior, an invariant and non-informative prior, is given by:

$$\pi(\theta) \propto \sqrt{\det(I(\theta))},$$

which is an objective, non-information prior that is invariant under reparametrization.

## Fisher-Rao Metric

---

A key discovery in **information geometry** (Amari et al. 2007) is that the FIM induces a **Riemannian metric** on the statistical manifold  $\mathcal{M} = \{p(x|\theta)\}$ , known as the **Fisher-Rao metric**. Given an infinitesimal displacement  $d\theta$  in the parameter space  $\Theta$ , the squared *length* of the displacement under the Fisher-Rao metric is given by:

$$ds^2 = d\theta^\top I(\theta) d\theta.$$

Thus, the Fisher-Rao metric captures the local geometry of the statistical manifold by taking the FIM as the local matrix form in a given coordinate system.

## Fisher-Rao Distance

---

Since the Fisher-Rao metric defines a Riemannian structure, one can define a distance between two points in the manifold. The **Fisher-Rao distance** between two distributions,  $p_1(x) = p(x|\theta_1)$  and  $p_2(x) = p(x|\theta_2)$ , is the geodesic length connecting  $\theta_1$  and  $\theta_2$ :

$$d_{FR}(p_1, p_2) := d_{FR}(\theta_1, \theta_2) = \inf_{\gamma} \int \sqrt{\dot{\gamma}(t)^\top I(\gamma(t)) \dot{\gamma}(t)} dt, \quad (1)$$

where the infimum is taken over all smooth curves  $\gamma : [0, 1] \rightarrow \Theta$  such that  $\gamma(0) = \theta_1$  and  $\gamma(1) = \theta_2$ . This formulation reveals that the Fisher-Rao distance is a Riemannian geodesic distance.

## Computing the Fisher-Rao Distance

---

While theoretically elegant, computing the Fisher-Rao distance as in [Equation 1](#) is challenging because it requires solving the geodesic equations:

$$\frac{d^2\theta^i}{dt^2} + \sum_{j,k} \Gamma_{jk}^i \frac{d\theta^j}{dt} \frac{d\theta^k}{dt} = 0,$$

where  $\Gamma_{jk}^i$  are Christoffel symbols derived from the Fisher information metric:

$$\Gamma_{jk}^i = \frac{1}{2} \sum_m I^{im} \left( \frac{\partial I_{mj}}{\partial \theta^k} + \frac{\partial I_{mk}}{\partial \theta^j} - \frac{\partial I_{jk}}{\partial \theta^m} \right).$$

Exact solutions are known to exist for only a handful of distributions ([Miyamoto et al. 2024](#)).

## Square-Root Transformation and Geodesic Distance

A practical alternative for computing the Fisher-Rao distance is the **square-root transformation**:

$$p(x) \mapsto \sqrt{p(x)}.$$

This procedure embeds probability densities into the Hilbert space  $L_2$  with the standard inner product. For  $f, g \in L_2$ , the inner product is given by

$$\langle f, g \rangle = \int f(x)g(x)dx.$$

It is straightforward to verify that the transformed functions lie on the infinite-dimensional unit sphere  $S^\infty$ . Specifically, let  $f \in L_2$  be the transformed function of some density  $\phi$ , i.e.,  $f(x) = \sqrt{\phi(x)}$ . Then, we have

$$\|f\|^2 = \langle f, f \rangle = \int f(x)^2 dx = \int \phi(x) = 1,$$

where the last equality follows from the fact that  $\phi(x)$  is a probability density function and thus integrates to 1.

On the unit sphere in  $L^2$ , the geodesic curve between two points is given by the great circle connecting them. Let  $\psi_1$  and  $\psi_2$  be two elements of  $S^\infty$ . Then, the geodesic distance between them is canonically known by  $\arccos(\langle \psi_1, \psi_2 \rangle)$ .

## Equivalence of Fisher-Rao Distance and Geodesic Distance of Square Root Densities

We are now ready to state the main result of this discussion. The Fisher-Rao distance between two distributions,  $p_1(x) = p(x|\theta_1)$  and  $p_2(x) = p(x|\theta_2)$ , is twice the geodesic distance between their square-root transformed densities,  $\psi_1(x) = \sqrt{p_1(x)}$  and  $\psi_2(x) = \sqrt{p_2(x)}$ , on  $S^\infty$ :

$$d_{FR}(p_1, p_2) = 2 \arccos \left( \int \sqrt{p_1(x)p_2(x)} dx \right) \quad (2)$$

To verify [Equation 2](#), consider an infinitesimal perturbation in  $p(x)$  by introducing a small variation:

$$p(x) \mapsto p(x) + \epsilon h(x), \quad (3)$$

where  $h(x)$  is an arbitrary function that integrates to 0 and  $\epsilon > 0$ . This constraint ensures that  $p(x) + \epsilon h(x)$  remains as a valid density. Our goal is to quantify how this small variation in  $p(x)$  induces a corresponding change in  $\psi(x)$ . In other words, we seek to measure the variation in the transformed function. Viewing the right-hand side of [Equation 3](#) as a function of  $\epsilon$ , we apply a first-order Taylor expansion to the square-root transformation:

$$\sqrt{p + \epsilon h} \approx \sqrt{p} + \frac{1}{2} \frac{\epsilon h}{\sqrt{p}} + O(\epsilon^2).$$

Thus, the infinitesimal perturbation in  $\psi(x)$  is:

$$\delta\psi(x) = \sqrt{p(x) + \epsilon h(x)} - \sqrt{p(x)} = \frac{1}{2} \frac{h(x)}{\sqrt{p(x)}} \epsilon + O(\epsilon^2). \quad (4)$$

From [Equation 4](#), we observe that  $\delta\psi(x)$  is linear in  $h(x)$ , demonstrating how small changes in  $p(x)$  translate to changes in  $\psi(x)$ <sup>2</sup>. Recall that the Fisher-Rao metric corresponds to infinitesimal displacements in probability space. Given the embedding of densities via the square-root transformation, it is natural to measure perturbations by computing the squared norm:

$$ds^2 = \|\delta\psi\|^2 = \int \left( \frac{1}{2} \frac{h(x)}{\sqrt{p(x)}} \right)^2 dx = \frac{\epsilon^2}{4} \int \frac{h(x)^2}{p(x)} dx. \quad (5)$$

Since  $h(x)$  represents an infinitesimal change in  $p(x)$ , we recognize that the integral in [Equation 5](#) corresponds precisely to the Fisher information metric evaluated for an infinitesimal displacement. Consequently, we obtain

$$ds^2 = \frac{1}{4} \int \frac{h(x)^2}{p(x)} dx.$$

which reveals that the Fisher-Rao metric is  $\frac{1}{4}$  times the standard metric on  $S^\infty$  induced by the square-root transformation.

By combining these observations with the fact that geodesic distances scale inversely with the metric factor, we conclude that the geodesic distance computed in the Fisher-Rao metric is twice the standard geodesic distance on the unit sphere in  $L^2$ . That is, for two densities  $p_1$  and  $p_2$  and their

transformations  $\psi_1 = \sqrt{p_1}$  and  $\psi_2 = \sqrt{p_2}$  in  $L_2$ , we have

$$d_{FR}(p_1, p_2) = 2 \times d_{S^\infty}(\psi_1, \psi_2) = 2 \arccos \left( \int \sqrt{p_1(x)p_2(x)} dx \right).$$

## Final Thoughts

1. In the derivation of the Fisher-Rao distance in terms of the scaled geodesic distance on  $S^\infty$ , no specific choice or constraint is imposed on  $\theta$ . This formulation naturally extends to nonparametric probability densities, as the geodesic distance in  $L^2$  depends solely on the embedding.
2. Despite the equivalence, some concerns remain regarding the computability of the Fisher-Rao distance. For instance, evaluating the integral can be challenging in many cases, particularly for high-dimensional distributions or arbitrary nonparametric densities, where numerical integration becomes intractable. Approximation via Monte Carlo methods, while useful, presents its own set of difficulties. Moreover, when densities are estimated, they may contain regions of extremely small values, leading to numerical precision issues.
3. I would like to express my appreciation to Prof. [Marco Radeschi](#) for his kindness and patience in enduring my endless questions in his differential geometry course - many of which, in hindsight, seem absurd.

## References

- Amari, Shun'ichi, Hiroshi Nagaoka, Shun'ichi Amari, and Shun'ichi Amari. 2007. *Methods of Information Geometry*. Translated by Daishi Harada. Nachdruck. Translations of Mathematical Monographs 191. Providence, Rhode Island: American Mathematical Society.
- Miyamoto, Henrique K., Fábio C. C. Meneghetti, Julianna Pinele, and Sueli I. R. Costa. 2024. "On Closed-Form Expressions for the Fisher-Rao Distance." *Information Geometry* 7 (2): 311–54. <https://doi.org/10.1007/s41884-024-00143-2>.

## Footnotes

1. There is an interesting connection to the objective Bayesian framework, which will appear in a later post. [↩](#)
2. From a geometric point of view,  $\delta\psi(x)$  belongs to the tangent space of  $S^\infty$  at the point  $\psi(x)$ . [↩](#)

## Citation

BibTeX citation:

```
@online{you2025,
  author = {You, Kisung},
  title = {What Is the {Fisher-Rao} Distance?},
  date = {2025-02-05},
  url = {https://kisungyou.com/Blog/blog_001_FisherRao.html},
  langid = {en}}
```

}

For attribution, please cite this work as:

You, Kisung. 2025. "What Is the Fisher-Rao Distance?" February 5, 2025.  
[https://kisungyou.com/Blog/blog\\_001\\_FisherRao.html](https://kisungyou.com/Blog/blog_001_FisherRao.html).

© Kisung You 2021-2025