Information Geometry

Pinak Joshi

Supervised by Diarmuid Crowley and Thorsten Hertl (Department of Mathematics and Statistics, University of Melbourne)

What is Information Geometry?

Information geometry equips a set of probability distributions with a **Riemannian metric** (specifically the Fisher-Rao metric). This induces a rich geometry on the set, which yields many quantities **invariant under parametrisation**, thus revealing many intrinsic features of our set of probability distributions – a few of which we explore here.

How do we measure distances on curved surfaces?

What does it mean to measure the distance between two points on a curved surface—for example, between the North Pole and Melbourne?

Gauss, under a similar question, extended classical (Euclidean) geometry to curved surfaces by measuring lengths and angles locally on tangent planes (using a smoothly varying inner product).

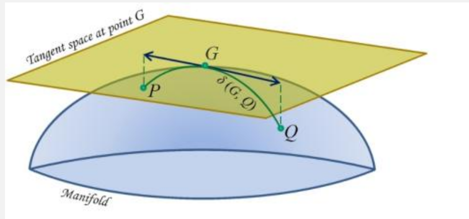

Figure 1: The arc PQ aimed to be approximated by the tangent vector at the point G [1]

Riemannian metrics

Riemann later abstracted Gauss' insight to arbitrary manifolds, defining a Riemannian metric as a smoothly varying inner product on tangent spaces, providing a framework for measuring lengths, angles, and distances on curved spaces.

Take S to be our surface and T_pS the tangent plane to S at the point p.

A Riemannian metric $g_p: T_pS \times T_pS \to \mathbb{R}$, is a smoothly varying inner product, that is symmetric, positive-definite and bilinear – it also only acs on "tangent vectors".

Symmetric: $g_p(v,w) = g_p(w,v)$
Positive definite: $g_p(v,v) > 0, v \neq 0$
Bilinear: $g_p(\lambda_1 v + \lambda_2 w, x) = \lambda_1 g_p(v,x) + \lambda_2 g_p(w,x)$ $(\lambda_1, \lambda_2 \in \mathbb{R})$

The geometry the metric imposes can be summarised as defined such that that the "cosine rule" holds for all v, w in the tangent space. That is,

$$g_p(v+w, v+w) = g_p(v,v) + g_p(v,w) + g_p(w,v) + g_p(w,w)$$
$$= g_p(v,v) + 2g_p(v,w) + g_p(w,w)$$
$$\|v+w\|^2_{g_p} = \|v\|^2_{g_p} + 2\|v\|_{g_p}\|w\|_{g_p}\cos(\theta) + \|w\|^2_{g_p}$$

Note: Every result in classical (Euclidean) Geometry can be derived from the (Euclidean) inner product.

ds^2 – A helpful informal summary of a metric

To understand different Riemannian metrics, it is useful to present how the Riemannian metric acts under some simple coordinate representation.

Standard sphere metric – Cartesian Form
We may express the standard sphere metric on S as
$g_p: T_pS \times T_pS \to \mathbb{R}$,

$$g_p(v,w) = v^{\mathsf{T}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} w$$

Note: v, w here must belong to the tangent space of S and specifically expressed in cartesian form for the above to hold.

If it is clear which local coordinates have been imposed on the surface, a Riemannian metric is often "summarised" with the informal notation ds^2

Informally, the standard sphere metric can be summarised as:

$$ds^2 = (dx\ dy\ dz)\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} dx \\ dy \\ dz \end{pmatrix} = dx^2 + dy^2 + dz^2$$

The notation ds^2 allows us to extract quickly how the metric "weights" our coordinate directions.

Standard sphere metric – Spherical Coordinates

The same metric looks different under the spherical coordinate parametrisation of the unit sphere

$$\Sigma(\varphi,\theta) = (\sin(\theta)\cos(\varphi)\quad \sin(\theta)\sin(\varphi)\quad \cos(\theta))$$

$$ds^2 = (d\varphi\ d\theta)\begin{pmatrix} \sin^2(\theta) & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} d\varphi \\ d\theta \end{pmatrix} = \sin^2(\theta)\, d\varphi^2 + d\theta^2$$

For interest, the relationship between the metric tensors and the Jacobian of Σ is simply $g_{\text{spherical}} = J^{\mathsf{T}} g_{\text{cart}} J$.

What is the distance from A to B along γ?

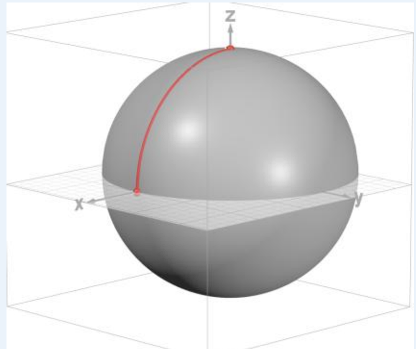

Figure 2: A unit sphere with an arc from the "North Pole" (A) to a point on the "equator" (B) shown in red

Given a curve $\gamma(t)$ on the surface, its "velocity" $\dot{\gamma}(t)$ is a tangent vector at each point. The metric assigns a norm to each tangent vector, $\|\dot{\gamma}(t)\|_{g_p}$, and integrating these norms/"speeds" gives us total arc length.

$$length(\gamma) = \int_a^b \|\dot{\gamma}(t)\|_{g_p}\, dt$$
(assuming $\gamma(a) = A, \gamma(b) = B$)

Note: The length of the arc is invariant under different parametrisations.

Cartesian coordinates

$$\gamma: \left[0, \tfrac{\pi}{2}\right] \to \mathbb{R}, \gamma(t) = (\sin(t), 0, \cos(t))$$

$\|\dot{\gamma}(t)\|^2_{g_p} = g_p(\dot{\gamma}(t), \dot{\gamma}(t)) = 1.$

$$length(\gamma) = \int_a^b \|\dot{\gamma}(t)\|_{g_p}\, dt = \int_0^{\frac{\pi}{2}} dt = \frac{\pi}{2}$$

Spherical coordinates

$$\gamma: \left[0, \tfrac{\pi}{2}\right] \to \mathbb{R}, \gamma(t) = (0, t)$$

As $\dot{\gamma}(t) = (0,1)$,
$\|\dot{\gamma}(t)\|^2_{g_p} = (0,1)\begin{pmatrix} \sin^2(\theta) & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 0 \\ 1 \end{pmatrix} = 1.$

$$length(\gamma) = \int_a^b \|\dot{\gamma}(t)\|_{g_p}\, dt = \int_0^{\frac{\pi}{2}} dt = \frac{\pi}{2}$$

Intrinsic nature of the metric

A Riemannian metric is an intrinsic object. Once fixed on a manifold, it does not depend on the choice of coordinates, and changing coordinates transforms the metric tensor in the standard Jacobian way in order to preserve certain geometric quantities (e.g. arc length)

Which geometric quantities are preserved even through reparametrisation?

Distances along a path are preserved through coordinate changes (see above computations). As such, the "geodesics" (which are the shortest paths through the manifold) also remain the same – though their coordinate equation may change.

Another quantity that is preserved is Gaussian curvature, which roughly measures how the geometry deviates from flat Euclidean space.

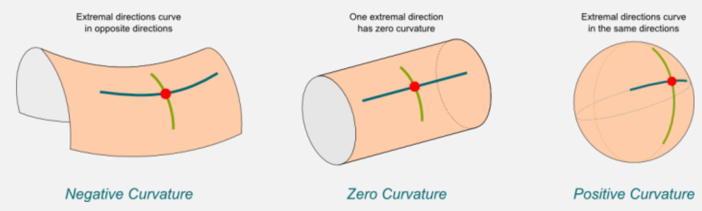

Figure 3: An illustration of what Gaussian curvature aims to capture [2]

Fisher Information (Fisher-Rao metric) – Starting Information Geometry

A statistical manifold arises when a family of probability distributions is parametrized smoothly, with each parameter value corresponding to a point on a manifold

$$S_1 = \{N(\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma^2 > 0\}$$

$$S_2 = \{Bern(\theta) = \theta^x(1-\theta)^{1-x} : \theta \in [0,1]\}$$

The **Fisher information matrix** from statistics already creates a natural Riemannian metric on the parameter space Θ,, for a large class of probability distributions.

It is given by

$$g_{ij}(\theta) = \mathbb{E}_\theta\left[\frac{\partial \log p(X \mid \theta)}{\partial \theta_i}\frac{\partial \log p(X \mid \theta)}{\partial \theta_j}\right]$$

Distances computed using the Fisher metric reflect "meaningful" or intrinsic differences between probability distributions rather than differences between parameter values alone, which is not invariant under coordinate transformations.

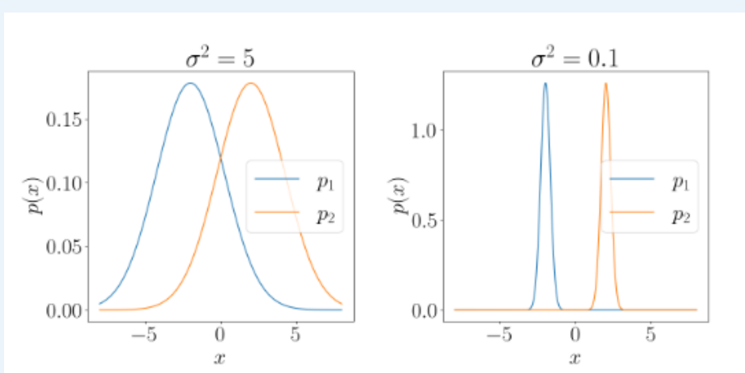

Figure 4: These 4 Gaussians can be represented by points in S_1. From left to right, these points are $(\mp 2, \sqrt{5})$, $(\mp 2, \sqrt{0.1})$ [3]

Under the Euclidean metric these two pairs of Gaussians have the same distance between them (4 units). However, under the Fisher-Rao metric, the Gaussians on the left are much "closer" to one another than those on the right. (~1.69 units compared to ~6.23 units).

Results: Exploring Geometry Induced by the Fisher–Rao Metric

We show in this section that, under the Fisher–Rao metric, common probability distribution families inherit natural geometric structures such as hyperbolic, Euclidean, and spherical geometry.

Hyperbolic Geometry of the Normal Family – Negative Curvature

Consider $S = \{N(\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma^2 > 0\}$

For clarity and spacing, we will present the metrics in line element form

$$ds^2_{\text{Normal}} = \frac{1}{\sigma^2}d\mu^2 + \frac{2}{\sigma^2}d\sigma^2$$

This representation also makes a fruitful change of variables much easier to spot. The well-studied hyperbolic metric has similar form

$$ds^2_{\text{Hyperbolic}} = \frac{1}{y^2}(dx^2 + dy^2)$$

Under the change of variables $x = \mu$, $y = \sqrt{2}\,\sigma$, we get

$$ds^2_{\text{Normal}} = \frac{2}{y^2}(dx^2 + dy^2) = 2ds^2_{\text{Hyperbolic}}$$

Thus, the Fisher metric here is exactly the Hyperbolic metric! (up to scale), and so all intrinsic geometric quantities (e.g. distances, geodesics and curvature) of S can be determined from hyperbolic space.

Curvature: As $ds^2_{\text{Normal}} = 2ds^2_{\text{Hyperbolic}}$, the Gaussian curvature of S is scaled by ½, Thus, the Gaussian curvature of S is -1/2 .

Geodesics: In the hyperbolic plane, geodesics are known to be semi-circles that intersect the x-axis at right angles, however considering our change of variables, they will appear as ellipse-shaped curves in the (μ, σ) plane

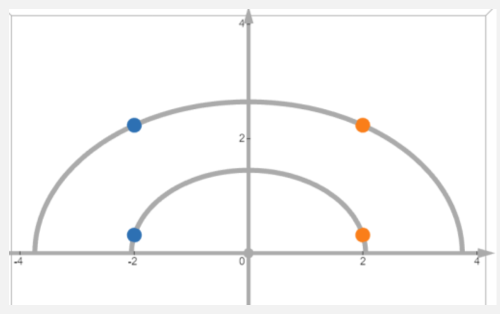

Figure 5: Points: $(\mp 2, \sqrt{5})$, $(\mp 2, \sqrt{0.1})$ and the shortest path (geodesic) between them, these are aligned with those in Figure 5 by choice.

Flat Geometry of the Poisson Family – No Curvature

Consider $P = \{Poi(\lambda_1) \otimes Poi(\lambda_2) : \lambda_1, \lambda_2 > 0\}$

$$ds^2 = \frac{d\lambda_1^2}{\lambda_1} + \frac{d\lambda_2^2}{\lambda_2} = dx^2 + dy^2 \quad (\text{Letting } x = 2\sqrt{\lambda_1} \text{ and } y = 2\sqrt{\lambda_2})$$

Thus, under the square root embedding, this is in fact Euclidean space.

Spherical Geometry of the Categorical Family – Positive Curvature

A categorical distribution assigns probabilities to one of k discrete outcomes.

Consider $C_3 = \{p = (p_1, p_2, p_3) : p_i > 0,\ p_1 + p_2 + p_3 = 1\}$

$$ds^2 = \frac{dp_1^2}{p_1} + \frac{dp_2^2}{p_2} + \frac{dp_3^2}{p_3} = 4(dq_1^2 + dq_2^2 + dq_3^2) \quad (\text{Letting } q_i = \sqrt{p_i})$$

Constraints affecting geometry

Due to the constraint $q_1^2 + q_2^2 + q_3^2 = 1$, our whole space of points, (q_1, q_2, q_3), should lie on the unit sphere.

Accordingly, the Fisher-Rao metric here is exactly our standard sphere metric (though scaled by 4) – but due to the further constraint $p_i > 0$ – we may only consider an "octant" of the sphere.

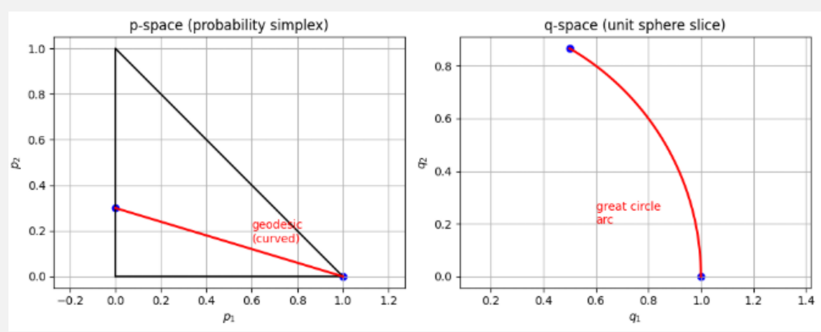

Figure 6: Geodesics for the Categorical distribution (shown in different spaces)

Curvature: The Categorical distribution has constant positive curvature of ¼ , since the sphere has Gaussian curvature 1.

Distance between 2 points: $s = (s_1, s_2, s_3)$, $t = (t_1, t_2, t_3)$
$d(s,t) = 2\arccos\left(\sum_{i=1}^{3}\sqrt{s_i t_i}\right)$

Geodesics: In the q-space, the geodesics follow a great circle arc

References
[1] M. Congedo, *Introduction to Riemannian Geometry*, online lecture notes, 2024. https://marco-congedo.github.io/PosDefManifold.jl/v0.3/introToRiemannianGeometry/
[2] Science4All, *The Brazuca: Geometry on a Football*, Science4All.org, 2014. https://www.science4all.org/article/brazuca/
[3] A. C. Jones, *Natural Gradients*, personal blog, 2020. https://andrewcharlesjones.github.io/journal/natural-gradients.html

pdjoshi@student.unimelb.edu.au