proceedings

MDPI

Proceedings

Information Geometry Conflicts With Independence †

John Skilling

Maximum Entropy Data Consultants Ltd., Kenmare, Ireland; skilling@eircom.net

† Presented at the 39th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering, Garching, Germany, 30 June–5 July 2019.

Published: 5 December 2019

check for updates

Abstract: Information Geometry conflicts with the independence that is required for science and for rational inference generally.

Keywords: Information Geometry; metric space; probability distribution

PACS: 02.50Cw; 02.70Rr

---

1. Introduction

Information Geometry [1] assigns a geometrical relationship between probability distributions, using the local curvature (Hessian) of the Kullback-Leibler formula

$$H(\mathbf{p};\mathbf{q}) = \sum_i p(i)\log\frac{p(i)}{q(i)} \tag{1}$$

as the covariant geometrical metric tensor [2,3] between $\mathbf{q}$ and $\mathbf{p}$. On a n-dimensional manifold $\mathbf{p}(\theta)$ specified by parameters $\theta^1,\ldots,\theta^n$, this $n\times n$ Riemannian metric g is

$$g_{jk}(\theta) = \sum_i p(i\mid\theta)\frac{\partial\log p(i\mid\theta)}{\partial\theta^j}\frac{\partial\log p(i\mid\theta)}{\partial\theta^k} \qquad \left(\text{or} \int dt\, p(t\mid\theta)\ldots \text{ in continuum form}\right) \tag{2}$$

Geodesic lengths ℓ and invariant volumes V follow from $(d\ell)^2 = \sum g_{jk}d\theta^j d\theta^k$ and $dV = \sqrt{\det g}\, d^n\theta$.

Necessarily, lengths are symmetric $\ell(\mathbf{p},\mathbf{q}) = \ell(\mathbf{q},\mathbf{p})$ between source and destination, so cannot be isomorphic to H which is from-to asymmetric. Yet (1) is the only connection which preserves independence of separate distributions, $H(\mathbf{x}\times\mathbf{p}\,;\mathbf{y}\times\mathbf{q}) = H(\mathbf{x};\mathbf{y}) + H(\mathbf{p};\mathbf{q})$. Specifically, when H is used to assign an optimal $\mathbf{p}$ (meaning minimally distorted from $\mathbf{q}$) under constraints, that "maximum entropy" selection also depends on separate optimisation $\mathbf{x}$-from-$\mathbf{y}$ unless H has the form (1) [4,5].

It follows that any imposed geometrical connection must introduce interference between supposedly separate distributions. That behaviour is incompatible with the practice of scientific inference, and is confirmed by a counter-example.

2. Counter-Example

Consider the 2-parameter family of probability distributions [6]

$$\mathbf{p}_{vw}(t \ (\mathrm{mod}\ 1)) = \begin{cases} f\left(\dfrac{t-v}{w}\right) & \text{for } v < t < v+w, \\ f\left(\dfrac{1+v-t}{1-w}\right) & \text{for } v+w < t < v+1, \end{cases} \tag{3}$$

Parameters v (location) and w (width) lie between 0 and 1. The function f (Figure 1) is monotonically increasing so that it rises from $f(0)$ at $t = v$ to $f(1)$ at $t = v{+}w$ (mod 1) before falling back to $f(0)$ at $t = v{+}1$ (mod 1). It is positive and normalised to $\int_0^1 f(u)du = 1$ so that the $p_{vw}(\cdot)$'s can be probability distributions on the interval (0,1) — which could model growth and decay in a periodic system.

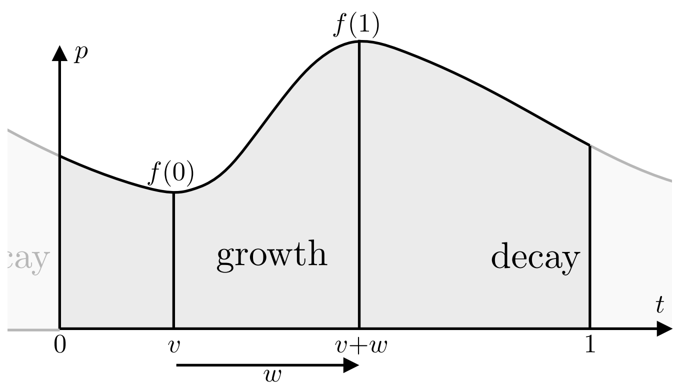

Figure 1. Does v affect w?

2.1. Two Parameters v and w

The 2×2 information-geometry metric evaluates to

$$\begin{bmatrix} g_{vv} & g_{vw} \\ g_{wv} & g_{ww} \end{bmatrix} = \frac{1}{w(1-w)} \int_0^1 \begin{bmatrix} 1 & u \\ u & u^2 \end{bmatrix} \frac{f'(u)^2}{f(u)}\, du = \begin{bmatrix} A & B \\ B & C \end{bmatrix} / \, w(1-w) \tag{4}$$

where A, B, C are constants. The table shows their values for two example functions. The first is easy to integrate while the second has vanishing slope $f'(0) = f'(1) = 0$ at the joins (as in Figure 1).

$f(u)$	$e^u/(e-1)$	$(8+6u^2-4u^3)/9$
A	$e{-}1 = 1.71828$	$\frac{11}{6}\log 2 + \frac{5}{6}\log 5 - \sqrt{15}(\arctan\frac{5}{\sqrt{15}} - \arctan\frac{1}{\sqrt{15}}) = 0.05945$
B	$1 = 1.00000$	$\frac{89}{12}\log 2 - \frac{25}{12}\log 5 - \frac{\sqrt{15}}{6}(\arctan\frac{5}{\sqrt{15}} - \arctan\frac{1}{\sqrt{15}}) - \frac{4}{3} = 0.02909$
C	$e{-}2 = 0.71828$	$\frac{251}{24}\log 2 + \frac{5}{24}\log 5 + \frac{13\sqrt{15}}{12}(\arctan\frac{5}{\sqrt{15}} - \arctan\frac{1}{\sqrt{15}}) - \frac{31}{3} = 0.01636$

The invariant volume element follows as

$$dV = \sqrt{\det g}\, dv\, dw = \frac{\sqrt{AC - B^2}}{w(1-w)}\, dv\, dw \tag{5}$$

where, by construction, $AC - B^2 > 0$. The total invariant volume is infinite.

$$V = \int_0^1 dv \int_0^1 dw \sqrt{\det g} = \infty \tag{6}$$

2.2. One Parameter w

If v had been fixed, $\mathbf{p}$ would have been confined to a submanifold $\mathbf{p}_w(\cdot)$ parameterised by w alone. The information-geometry metric reduces to

$$g_{ww} = \frac{1}{w(1-w)} \int_0^1 u^2 \frac{f'(u)^2}{f(u)}\, du = \frac{C}{w(1-w)} \tag{7}$$

The invariant volume element follows as

$$dV = \sqrt{g_{ww}}\, dw = \left(\frac{C}{w(1-w)}\right)^{1/2} dw \tag{8}$$

where, by construction, $C > 0$. The total invariant volume is finite.

$$V = \int_0^1 \sqrt{g_{ww}}\, dw = \pi C^{1/2} \tag{9}$$

2.3. Comparison of One and Two Parameters

Both shape ((5) versus (8)) and integral ((6) versus (9)) over w differ qualitatively according to whether or not v is held fixed.

$$\boxed{\text{Treatment of } v \text{ influences invariant volumes over } w} \qquad \text{[Geometry]}$$

That is a **mathematical fact** of information geometry.

2.4. Science

For scientific application, (3) defines a wraparound translation-invariant model in which v does not affect w.

$$\boxed{\text{Treatment of } v \text{ should not influence inference about } w} \qquad \text{[Science]}$$

That is a **science requirement**. Any observational consequence of information-geometry's invariant volumes would be rejected by the informed scientist. If there were such consequence, then observation of width w could be used to infer something about location v, contrary to the intention of the formulation.

3. Conclusions

Information geometry is not science. It denies the independence of separate parameters even though such independence is a fundamental requirement of scientific inquiry. The assumption of a geometrical connection between distributions is unnecessary for science and it fails under test.

Information geometry is a self-consistent mathematical structure which (like any other piece of mathematics) may find specialised application within science, but is not fundamental to it. The only fundamental connection is the Kullback-Leibler, which is from-to asymmetric hence not geometric.

Funding: This research received no external funding.

Acknowledgments: This investigation has been refined by many conversations, in particular with Ariel Caticha.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Amari, S. Differential-geometrical methods in statistics. In *Lecture Notes in Statistics*; Springer-Verlag: Berlin, Germany, 1985.
2. Fisher, R. A. Theory of statistical estimation *Proc. Camb. Philos. Soc.* **1925**, *122*, 700–725.
3. Rao, C.R. Information and the accuracy attainable in the estimation of statistical parameters. *Bull. Calcutta Math. Soc.* **1945**, *37*, 81–89.
4. Shannon, C.F. A Mathematical theory of Communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423, 623–656.
5. Knuth, K.H.; Skilling, J. Foundations of Inference. *Axioms* **2012**, *1*, 38–73.
6. Skilling, J. Critique of Information Geometry. *AIP Conf. Proc.* **2013**, *1636*, 24–29.