

Notebooks

# Information Geometry

13 Aug 2012 20:50

---

This is a slightly misleading name for applying [differential geometry](#) to families of [probability distributions](#), and so to [statistical models](#). Information does however play two roles in it: Kullback-Leibler information, or [relative entropy](#), features as a measure of divergence (not quite a metric, because it's asymmetric), and Fisher information takes the role of curvature. One very nice thing about information geometry is that it gives us very strong tools for proving results about statistical models, simply by considering them as well-behaved geometrical objects. Thus, for instance, it's basically a tautology to say that a manifold is not changing much in the vicinity of points of low curvature, and changing greatly near points of high curvature. Stated more precisely, and then translated back into probabilistic language, this becomes the Cramer-Rao inequality, that the variance of a parameter estimator is at least the reciprocal of the Fisher information. As someone who likes differential geometry, and now is interested in statistics, I find this very pleasing.

As a physicist, I have always been somewhat bothered by the way statisticians seem to accept *particular* parametrizations of their models as obvious and natural, and build those parameterizations into their procedures. In linear regression, for instance, it's reasonably common for them to want to find models with only a few non-zero coefficients. This makes my thumbs prick, because it seems to me obvious that if I regressed on arbitrary linear combinations of my covariates, I have exactly the same information (provided the transformation is invertible), and so I'm really looking at exactly the same model --- but in general I'm *not* going to have a small number of non-zero coefficients any more. In other words, I want to be able to do *coordinate-free* statistics. Since differential geometry lets me do coordinate-free physics, information geometry seems like an appealing way to do this. There are various information-geometric model selection criteria, which I want to know more about; I suspect, based purely on this disciplinary prejudice, that they will out-perform coordinate-dependent criteria.

I should also mention that [statistical physics](#), while it does no actual *statistics*, is also very much concerned with probability distributions. Sun-Ichi Amari, who is the leader of a large and impressive Japanese school of information-geometers, has a nice result (in, e.g., his "Hierarchy of Probability Distributions" paper) showing that [maximum entropy distributions](#) are, exactly, the ones with minimal interaction between their variables --- the ones which approach most closely to independence. I think this throws a very interesting new light on the issue of *why* we can assume equilibrium corresponds to a state of [maximum entropy](#) (*pace* Jaynes, assuming independence is clearly not an innocent way of saying "I really don't know anything more"). I also see, via the Arxiv, that people are starting to think about phase transitions in information-geometric terms, which seems natural in retrospect, though I can't comment further, not having read the papers.

See also: [Exponential Families of Probability Measures](#), where the geometry is especially nice; [Filtering and State Estimation](#) for some papers on differential-geometric ideas in statistical state estimation and signal processing; [Partial Identification of Parametric Statistical Models](#)

Recommended, big picture:

- S.-I. Amari, O. E. Barndorff-Nielsen, R. E. Kass, S. L. Lauritzen, and C. R. Rao, *Differential Geometry in Statistical Inference* [Now [free online](#)]
- Sun-Ichi Amari and Hiroshi Nagaoka, *Methods of Information Geometry*
- [Robert E. Kass](#) and Paul W. Vos, *Geometrical Foundations of Asymptotic Inference*
- Rudolf Kulhavy, *Recursive Nonlinear Estimation: A Geometric Approach*

Recommended, close-ups:

- Sun-Ichi Amari, "Information Geometry on Hierarchy of Probability Distributions", *IEEE Transactions on Information Theory* **47** (2001): 1701--1711 [[PDF reprint](#)]
- Vijay Balasubramanian, "Statistical Inference, Occam's Razor, and Statistical Mechanics on the Space of Probability Distributions", *Neural Computation* **9** (1997): 349--368
- Hwan-sik Choi and Nicholas M. Kiefer, "Differential Geometry and Bias Correction in Nonnested Hypothesis Testing" [[PDF preprint via Kiefer](#)]
- Tommi S. Jaakkola and David Haussler, "Exploiting generative models in discriminative classifiers", *NIPS 11* (1998) [[PDF](#)]
- I. J. Myung, Vijay Balasubramanian and M. A. Pitt, "Counting probability distributions: Differential geometry and model selection", *Proceedings of the National Academy of Sciences (USA)* **97** (2000): 11170--11175

To read:

- Khadiga Arwini and C. T. J. Dodson, "Neighborhoods of Independence for Random Processes", [math.DG/0311087](#)
- Nihat Ay
  - "Information geometry on complexity and stochastic interaction" [[preprint](#)]
  - "An information-geometric approach to a theory of pragmatic structuring" [[preprint](#)]
- Nihat Ay, Jürgen Jost, Hồng Vân Lê, Lorenz Schwachhüfer, "Information geometry and sufficient statistics", [arxiv:1207.6736](#)
- O. E. Barndorff-Nielsen and Richard D. Gill, "Fisher Information in Quantum Statistics", [quant-ph/9808009](#)
- Damiano Brigo, "The direct L2 geometric structure on a manifold of probability densities with applications to Filtering", [arxiv:1111.6801](#)
- Xavier Calmet and Jacques Calmet, "Dynamics of the Fisher Information Metric", [cond-mat/0410452](#) = *Physical Review E* **71** (2005): 056109
- Gavin E. Crooks, "Measuring Thermodynamic Length", *Physical Review Letters* **99** (2007): 100602 ["Thermodynamic length is a metric distance between equilibrium thermodynamic states. Among other interesting properties, this metric asymptotically bounds the dissipation induced by a finite time transformation of a thermodynamic system. It is also connected to

the Jensen-Shannon divergence, Fisher information, and Rao's entropy differential metric."]

- Imre Csiszar and Frantisek Matus, "Closures of exponential families", [\*Annals of Probability\* \*\*33\*\* \(2005\): 582--600](#) = [math.PR/0503653](#)
- C. T. J. Dodson and H. Wang, "Iterative Approximation of Statistical Distributions and Relation to Information Geometry", [\*Statistical Inference for Stochastic Processes\* \*\*4\*\* \(2001\): 307--318](#) ["the optimal control of stochastic processes through sensor estimation of probability density functions is given a geometric setting via information theory and the information metric."]
- Tryphon T. Georgiou, "An intrinsic metric for power spectral density functions", [math.PR/0608486](#) [Leads to a Riemannian geometry on stochastic processes, apparently...]
- Paolo Gibilisco and Tommaso Isola, "Uncertainty Principle and Quantum Fisher Information", [math-ph/0509046](#)
- Paolo Gibilisco, Daniele Imparato and Tommaso Isola, "Uncertainty Principle and Quantum Fisher Information II" [math-ph/0701062](#)
- Kazushi Ikeda, "Information Geometry of Interspike Intervals in Spiking Neurons", [\*Neural Computation\* \*\*17\*\* \(2005\): 2719--2735](#)
- Shiro Ikeda, Toshiyuki Tanaka and Shun-ichi Amari, "Stochastic Reasoning, Free Energy, and Information Geometry", [\*Neural Computation\* \*\*16\*\* \(2004\): 1779--1810](#)
- W. Janke, D.A. Johnston and R. Kenna, "Information Geometry and Phase Transitions", [cond-mat/0401092](#) = *Physica A* **336** (2004): 181--186
- G. Lebanon, "Axiomatic Geometry of Conditional Models", [\*IEEE Transactions on Information Theory\* \*\*51\*\* \(2005\): 1283--1294](#)
- M. K. Murray and J. W. Rice, *Differential Geometry and Statistics* [Thanks to [Anand Sarwate](#) for the recommendation]
- Hiroyuki Nakahara and Shun-ichi Amari, "Information-Geometric Measure for Neural Spikes", *Neural Computation* **14** (2002): 2269--2316
- Frank Nielsen, "Chernoff information of exponential families", [arxiv:1102.2684](#)
- J. Pletonen and S. Kaski, "Discriminative Components of Data", *IEEE Transactions on Neural Networks* **16** (2005): 68--83
- Steven T. Smith, "Covariance, Subspace, and Intrinsic Cramer-Rao Bounds", [\*IEEE Transactions on Signal Processing\* \*\*53\*\* \(2005\): 1610--1630](#) [Thanks to Dr. Smith for a reprint]
- R. F. Streater, "Quantum Orlicz spaces in information geometry", [math-ph/0407046](#)
- Masanobu Taniguchi and Yoshihide Kakizawa, *Asymptotic Theory of Statistical Inference for Time Series* [The first few chapters are quite nice, but I haven't gotten to the parts where they actually use much information geometry]
- Marc Toussaint, "Notes on information geometry and evolutionary processes", [nlin.AO/0408040](#)
- Mark K. Transtrum, Benjamin B. Machta, James P. Sethna, "The geometry of nonlinear least squares with applications to sloppy models and optimization", [arxiv:1010.1449](#) [From the abstract, this sounds like a rediscovery of Amari's 1967 paper, but Sethna is someone who usually know what he's doing so I reserve judgement]

- Paolo Zanardi, Paolo Giorda, and Marco Cozzini, "Information-Theoretic Differential Geometry of Quantum Phase Transitions", [\*Physical Review Letters\* \*\*99\*\* \(2007\): 100603](#)
- 

[permanent link for this note](#) [RSS feed for this note](#)

[\*Notebooks\*](#): Hosted, but not endorsed, by the [Center for the Study of Complex Systems](#)