RESEARCH PAPER

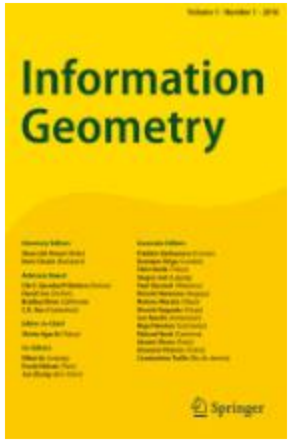
Information geometry of Wasserstein statistics on shapes and affine deformations

Shun-ichi Amari[1,2] · Takeru Matsuda[2,3]

Received: 31 July 2023 / Revised: 13 February 2024 / Accepted: 16 June 2024 /
Published online: 15 July 2024

Abstract
Information geometry and Wasserstein geometry are two main structures introduced in a manifold of probability distributions, and they capture its different characteristics. We study characteristics of Wasserstein geometry in the framework of [32] for the affine deformation statistical model, which is a multi-dimensional generalization of the location-scale model. We compare merits and demerits of estimators based on information geometry and Wasserstein geometry. The shape of a probability distribution and its affine deformation are separated in the Wasserstein geometry, showing its robustness against the waveform perturbation in exchange for the loss in Fisher efficiency. We show that the Wasserstein estimator is the moment estimator in the case of the elliptically symmetric affine deformation model. It coincides with the information-geometrical estimator (maximum-likelihood estimator) when the waveform is Gaussian. The role of the Wasserstein efficiency is elucidated in terms of robustness against waveform change.

Keywords Elliptically symmetric · Information geometry · Optimal transport · Robustness · Wasserstein distance



Volume 7
Issue 2
Article 83

RESEARCH

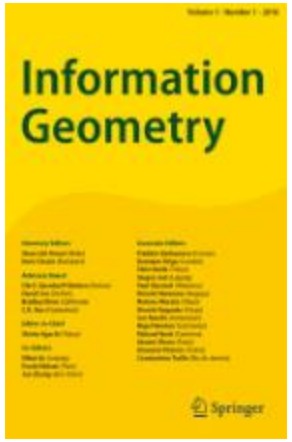
On closed-form expressions for the Fisher–Rao distance

Henrique K. Miyamoto[1] · Fábio C. C. Meneghetti[2] · Julianna Pinele[3] · Sueli I. R. Costa[2]

Received: 29 February 2024 / Revised: 31 July 2024 / Accepted: 1 August 2024 /
Published online: 16 September 2024

Abstract
The Fisher–Rao distance is the geodesic distance between probability distributions in a statistical manifold equipped with the Fisher metric, which is a natural choice of Riemannian metric on such manifolds. It has recently been applied to supervised and unsupervised problems in machine learning, in various contexts. Finding closed-form expressions for the Fisher–Rao distance is generally a non-trivial task, and those are only available for a few families of probability distributions. In this survey, we collect examples of closed-form expressions for the Fisher–Rao distance of both discrete and continuous distributions, aiming to present them in a unified and accessible language. In doing so, we also: illustrate the relation between negative multinomial distributions and the hyperbolic model, include a few new examples, and write a few more in the standard form of elliptical distributions.

Keywords Fisher metric · Fisher–Rao distance · Geodesic distance · Parametric distributions · Statistical manifolds



Volume 7
Issue 2
Article 84

PERSPECTIVE

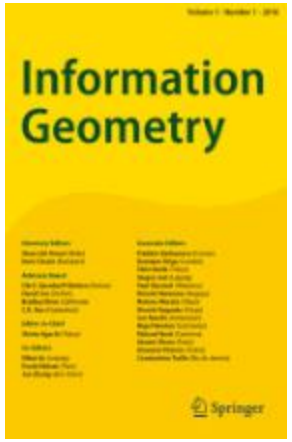
Maximal co-ancillarity and maximal co-sufficiency

H. S. Battey[1]

Received: 10 June 2024 / Revised: 10 August 2024 / Accepted: 12 August 2024 /
Published online: 17 September 2024

Abstract
The purpose of this exposition is to provide some new perspectives on conditional inference through a notional idealised separation within the minimal sufficient statistic, allowing a geometric account of key ideas from the Fisherian position. The notional idealised separation, in terms of an ancillary statistic and what I call a maximal co-ancillary statistic, provides conceptual insight and clarifies what is sought from an approximate conditional analysis, where exact calculations may not be available. A parallel framework applies in the Fisherian assessment of model adequacy. Both aspects are discussed and illustrated geometrically through examples.

Keywords Ancillary · Conditional inference · Inferential separations · Information · Minimal sufficiency · Model adequacy



Volume 7
Issue 2
Article 85

RESEARCH PAPER

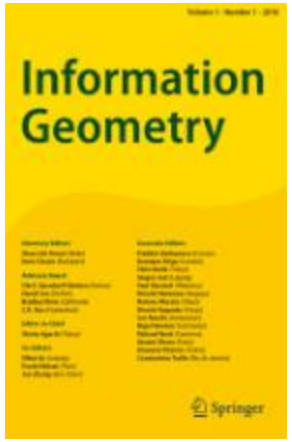
Reverse em-problem based on Bregman divergence and its application to classical and quantum information theory

Masahito Hayashi[1,2,3]

Received: 12 April 2023 / Revised: 15 September 2024 / Accepted: 17 September 2024 /
Published online: 3 October 2024

Abstract
The recent paper (IEEE Trans. IT 69, 1680) proposed an analytical calculation method for the channel capacity without iteration. However, it has restrictions for its application. Also, it does not explain why the channel capacity can be solved analytically in this special case. To expand the applicability of this method, we recall the reverse em-problem, which was proposed by Toyota (Information Geometry, 3, 1355 (2020)) as the repetition of the inverse map of the em iteration to calculate the channel capacity, which is given as the maximum of the mutual information. However, it left several open problems. We formulate the reverse em-problem based on Bregman divergence, and solve these open problems. Using these results, we convert the reverse em-problem into em-problems, and derive a non-iterative formula for the reverse em-problem, which can be considered as a generalization of the above analytical calculation method. Additionally, this derivation explains the information geometrical structure of this special case.

Keywords Maximization · Bregman divergence · Information geometry · Channel capacity



Volume 7
Issue 2
Article 86

BRIEF REPORT

On the statistical Lie groups of normal distributions

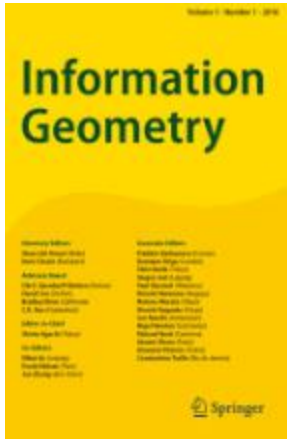
Jun-ichi Inoguchi[1]

Received: 31 May 2024 / Revised: 11 September 2024 / Accepted: 22 September 2024 /
Published online: 5 October 2024

Abstract

The α-connections (other than the Levi-Civita connection of the Fisher metric) on the statistical Lie group of the normal distributions can not be the Levi-Civita connection of any left invariant semi-Riemannian metrics.

Keywords Normal distribution · Statistical Lie groups · Lorentz metrics



Volume 7
Issue 2
Article 87

RESEARCH PAPER

Simple variational inference based on minimizing Kullback–Leibler divergence

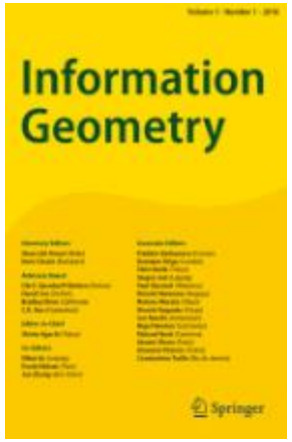
Ryo Nakamura[1] · Tomooki Yuasa[2] · Takafumi Amaba[3] · Jun Fujiki[3]

Received: 14 July 2023 / Revised: 13 May 2024 / Accepted: 22 September 2024 /
Published online: 13 October 2024

Abstract

We introduce a new methodology of estimation of the true distribution. The procedure of getting estimated distribution is constructed from Bayesian statistical models in which the statistical model is fixed and prior distributions for the parameters are varied. Then we consider the Kullback–Leibler divergence of the true distribution from the estimated one and derive variational formulae of the Kullback–Leibler divergence over prior distributions. Next, we propose a Newton–Raphson method for simulating the prior distribution, which is the critical point, based on the Riemannian geometry of the probability simplex. The method can run once a sample from the true distribution is obtained without any other knowledge of the true distribution. For the geometry, we employ the Riemannian metric induced from the characteristic function, which appears in Vinberg's theory of homogeneous convex cones. As a by-product of the geometry, we derive an interpretation that the Kullback–Leibler divergence is the logarithm of a gauge transformation. From this, we obtain a viewpoint that envisaging the true distribution is nothing but gauge fixing, and this depiction as gauge theory seems to match the context of statistics. Also, we show some numerical results on how well our methodology works.

Keywords Prior distribution · Kullback–Leibler divergence · Cross entropy · Variational formula · Vinberg's Riemannian metric · Gauge



Volume 7
Issue 2
Article 88

RESEARCH

The Bayesian central limit theorem for exponential family distributions: a geometric approach

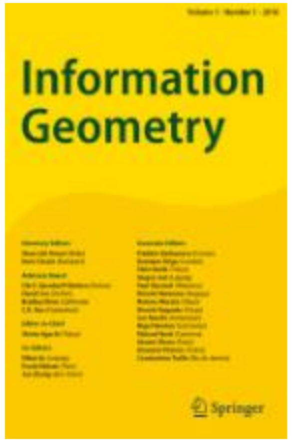
Geoff Goehle[1]

Received: 23 April 2024 / Revised: 23 September 2024 / Accepted: 30 September 2024 /
Published online: 15 October 2024

Abstract

The Bernstein-von Mises theorem, also known as the Bayesian Central Limit Theorem (BCLT), states that under certain assumptions a posterior distribution can be approximated as a multivariate normal distribution as long as the precision parameter is large. We derive a special case of the BCLT for the canonical conjugate prior of a regular exponential family distribution using the machinery of information geometry. Our approach applies the core approximation for the BCLT, Laplace's method, to the free entropy (i.e., log-normalizer) of an exponential family distribution. Additionally, we formulate approximations for the Kullback–Leibler divergence and Fisher-Rao metric on the conjugate prior manifold in terms of corresponding quantities from the likelihood manifold. We also include an application to the categorical distribution and show that the free entropy derived approximations are related to various series expansions of the gamma function and its derivatives. Furthermore, for the categorical distribution, the free entropy approximation produces higher order expansions than the BCLT alone.

Keywords Information geometry · Bayesian inference · Probability · Differential geometry



Volume 7
Issue 2
Article 89

RESEARCH PAPER

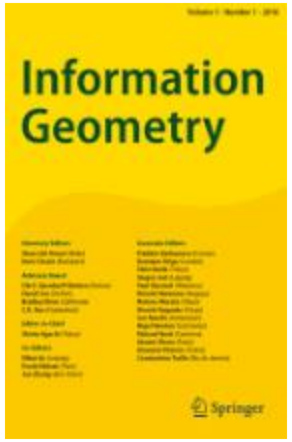
An algorithm for learning representations of models with scarce data

Adrian de Wynter[1]

Received: 29 March 2021 / Revised: 30 March 2024 / Accepted: 11 October 2024 /
Published online: 30 October 2024

Abstract
We present an algorithm for solving binary classification problems when the dataset is not fully representative of the problem being solved, and obtaining more data is not possible. It relies on a trained model with loose accuracy constraints, an iterative hyperparameter searching-and-pruning procedure over a search space Θ, and a data-generating function. Our algorithm works by reconstructing up to homology the manifold on which lies the support of the underlying distribution. We provide an analysis on correctness and runtime complexity under ideal conditions and an extension to deep neural networks. In the former case, if $|\Theta|$ is the number of hyperparameter sets in the search space, this algorithm returns a solution that is up to $2(1 - 2^{-|\Theta|})$ times better than simply training with an enumeration of Θ and picking the best model. As part of our analysis we also prove that an open cover of a dataset has the same homology as the manifold on which lies the support of the underlying probability distribution, if and only said dataset is learnable. This latter result acts as a formal argument to explain the effectiveness of contemporary data expansion techniques.

Keywords Data augmentation · Semi-supervised learning



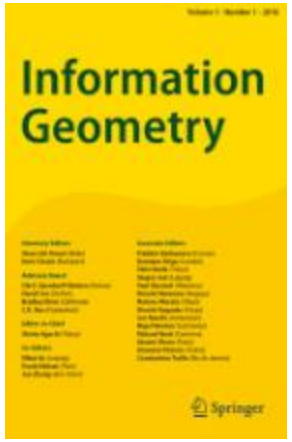
Volume 7
Issue 2
Article 90

Check for updates

An embedding structure of determinantal point process

Hideitsu Hino[1] · Keisuke Yano[1]

Received: 2 July 2024 / Revised: 8 October 2024 / Accepted: 18 October 2024 /
Published online: 7 November 2024



Volume 7
Issue 2
Article 91

Abstract
This paper investigates the information geometrical structure of a determinantal point process (DPP). It demonstrates that a DPP is embedded in the exponential family of log-linear models. The extent of deviation from an exponential family is analyzed using the e-embedding curvature tensor, which identifies partially flat parameters of a DPP. On the basis of this embedding structure, an information-geometrical relationship between a marginal kernel and an L-ensemble kernel is discovered.

Keywords Curved exponential family · Discrete statistical model · L-ensemble kernel · Partially ordered set · Statistical curvature

RESEARCH

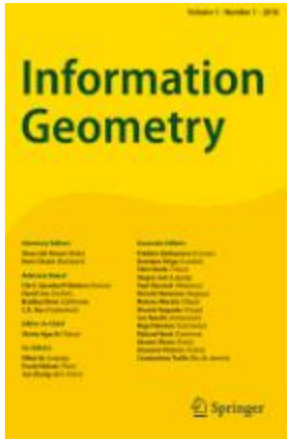
On statistics which are almost sufficient from the viewpoint of the Fisher metrics

Kaori Yamaguchi[1] · Hiraku Nozawa[1]

Received: 24 August 2024 / Revised: 1 November 2024 / Accepted: 10 November 2024 /
Published online: 22 November 2024

Abstract
Given a statistical model, a statistic on the model is sufficient if the Fisher metric of the induced model coincides with the original Fisher metric, according to the definition by Ay-Jost-Lê-Schwachhöfer. We introduce and study its quantitative version: for $0 < \delta \leq 1$, we call a statistic δ-almost sufficient if $\delta^2 \mathfrak{g}(v, v) \leq \mathfrak{g}'(v, v)$ for every tangent vector v of the parameter space, where $\mathfrak{g}$ and $\mathfrak{g}'$ are the Fisher metric of the original and the induced model, respectively. By the monotonicity theorem due to Amari-Nagaoka and Ay-Jost-Lê-Schwachhöfer, the Fisher metric $\mathfrak{g}'$ of the induced model for such a statistic is bi-Lipschitz equivalent to the original one $\mathfrak{g}$, which means that the information loss of the statistic is uniformly bounded. We characterize such statistics in terms of the conditional probability or by the existence of a certain decomposition of the density function in a way similar to the characterizations of sufficient statistics due to Ay-Jost-Lê-Schwachhöfer and Fisher-Neyman.

Keywords Information geometry · Sufficient statistics · Fisher metrics · Statistical manifolds · Binomial distribution



Volume 7
Issue 2
Article 92