Ay, Nihat; Jost, Jürgen; Lê, Hông Vân; Schwachhöfer, Lorenz

Information geometry. (English) Zbl 1383.53002

Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge 64. Cham: Springer (ISBN 978-3-319-56477-7/hbk; 978-3-319-56478-4/ebook). xi, 407 p. (2017).

The epoch-making paper [*C. R. Rao*, Bull. Calcutta Math. Soc. 37, 81–91 (1945; Zbl 0063.06420)] has made use of Fisher information to define a Riemannian metric on a space of probability distributions on finite samples, which enabled Rao to derive the Cramér-Rao inequality. After 30 years, *B. Efron* [Ann. Stat. 3, 1189–1242 (1975; Zbl 0321.62013)] extended Rao's ideas to a higher-order asymptotic theory of statistical inference, defining smooth subfamilies of large exponential families and their statistical curvature, which is no other than the second fundamental form of subfamilies put down as Riemannian submanifolds in the Riemannian manifold of the underlying exponential family endowed with the Fisher metric. In [Proc. R. Soc. Lond., Ser. A 186, 453–461 (1946; Zbl 0063.03050)], *H. Jeffreys* introduced what is now called the Kullback-Leibler divergence, where, for two infinitely close distributions, their Kullback-Leibler divergence was to be written as a quadratic form with elements of the Fisher information matrix as coefficients, the quadratic form being interpreted as the length of a Riemannian manifold while the Fisher information playing the role of the Riemannian metric. This geometrization of the statistical model enabled him to derive his prior distributions as the measures naturally induced by the Riemannian metric. All this is the harbinger of information geometry.

Inspired by Efron's above mentioned paper [loc. cit.], *S. Amari* [Ann. Stat. 10, 357–385 (1982; Zbl 0507.62026)] introduced the notion of α-connections and exhibited its usefulness in the asymptotic theory of statistical estimation, finally realizing Fisher's unfulfilled dream of showing that the maximal likelihood estimator is optimal (see also [*S. Amari*, Differential-geometrical methods in statistics. Berlin etc.: Springer-Verlag. (1985; Zbl 0559.62001); Differential-geometrical methods in statistics. Corr. 2nd. printing. Berlin etc.: Springer-Verlag (1990; Zbl 0701.62008)], [*S. Amari* and *H. Nagaoka*, Methods of information geometry. Transl. from the Japanese by Daishi Harada. Providence, RI: AMS, American Mathematical Society; Oxford: Oxford University Press (2000; Zbl 0960.62005); Methods of information geometry. Translation from the Japanese by Daishi Harada. Reprint of the 2000 edition. Providence, RI: American Mathematical Society (AMS) (2008; Zbl 1146.62001)]). *S. Amari* and *H. Nagaoka* [Dept. Math. Eng. and Instr. Phys., Univ. of Tokyo, Technical Report METR 82-7 (1982)] have introduced the notion of dual connections and developed the general theory of dually flat spaces, which was applied successfully to the geometry of α-connections. Amari's ideas of α-connections were preceded by the Soviet mathematician *N. N. Chentsov* [Sov. Math., Dokl. 5, 1282–1286 (1964; Zbl 0129.10503); translation from Dokl. Akad. Nauk SSSR 158, 543–545 (1964)], who defined an affine flat connection, called the e-connection, on the set $\mathcal{P}_+(\Omega,\mu)$. *N. N. Chentsov* [Dokl. Akad. Nauk SSSR 164, 511–514 (1965)] and *N. Morse* and *R. Sacksteder* [Ann. Math. Stat. 37, 203–214 (1966; Zbl 0158.37105)] independently introduced the category of mathematical statistics almost simultaneously. *N. N. Chentsov* [Statistical decision rules and optimal inference. Transl. from the Russian by the Israel Program for Scientific Translations; ed. by Lev J. Leifman. Providence, R.I.: American Mathematical Society (AMS). (1982; Zbl 0484.62008)] discovered the Amari-Chenstov connections and established their uniqueness by the invariance under sufficient statistics. All this is the emergence of information geometry.

The book under review, devoted to information geometry, consists of six chapters. The first chapter is an introduction.

Chapter 2, consisting of nine sections, deals with basic constructions in the elementary case that the underlying space Ω is of finitely many elements. It introduces and discusses (§2.1) manifolds of finite measures, (§2.2) the Fisher metric, (§2.3) gradient fields, (§2.4) m-connections as well as e-connections, (§2.5) the Amari-Chentsov tensor as well as the α-connections, (§2.6) congruent families of tensors, (§2.7) divergences, (§2.8) exponential families, (§2.9) hierarchical and graphical models. The book is mainly concerned with differential geometry of statistics, but algebraic geometry is also useful in statistics and the corresponding approach is called *algebraic statistics*, which is particularly important for comprehending closures of models. In this regard, the authors present implicit descriptions of exponential families and their closures in §2.8.2, while, in the context of graphical models and their closures [*D. Geiger* et al., Ann. Stat. 34, No. 3, 1463–1492 (2006; Zbl 1104.60007)], they establish the Hammersley-Clifford theorem on the lines of [*S. L. Lauritzen*, Graphical models. Oxford: Oxford Univ. Press (1998; Zbl 0907.62001)] in §2.9.3. Algebraic statistics makes use of computational commutative algebra in order to determine maximum likelihood estimators on the one hand, and provides linear and toric models on the other.

Chapter 3, consisting of three sections, considers a general space Ω, for which a functional analytic framework is to be developed. Parametrized measure models and suitable integrability properties are also discussed. The first section is of a more informal character, paving the way to more formal considerations in the second section, where the authors discuss parametrized measure models based upon [Bernoulli 24, No. 3, 1692–1725 (2018; Zbl 1419.62057)]. The third section investigates the Pistone-Sempi structure [*G. Pistone* and *C. Sempi*, Ann. Stat. 23, No. 5, 1543–1561 (1995; Zbl 0848.62003)], which geometrizes $\mathcal{M}_+(\Omega,\mu_0)$ of finite measures compatible with a fixed measure μ_0 in place of $\mathcal{M}(\Omega)$, providing $\mathcal{M}_+(\Omega,\mu_0)$ with the structure of a Banach manifold.

Chapter 4 addresses differential geometry of statistical models, discussing dualistic structures which consist of a Riemannian metric and two dual connections. When both connections are torsion-free, the structure is to be obtained from potential functions which are convex functions with their second derivatives playing the role of a metric and their third derivatives being regarded as a symmetric 3-tensor. Here we find a pair of dual affine structures, which is the very geometry discovered by Amari and Chentsov. The theory of dually affine structures turn out to be a real analogue of Kähler geometry, as can be seen in [*S.-Y. Cheng* and *S.-T. Yau*, in: Differential geometry and differential equations, Proc. 1980 Beijing Sympos., Vol. 1, 339–370 (1982; Zbl 0517.35020); *S. Y. Cheng*, in: Proc. Int. Congr. Math., Warszawa 1983, Vol. 1, 533–539 (1984; Zbl 0601.53055)]. The analogy stems from the fact that the Kähler form of a Kähler manifold is locally obtained as the complex Hessian of some function, while our metric is given by the real Hessian. The analogy is enjoyed and exploited, by way of example, in [*H. Shima*, The geometry of Hessian structures. Hackensack, NJ: World Scientific (2007; Zbl 1244.53004); *H. Shima* and *K. Yagi*, Differ. Geom. Appl. 7, No. 3, 277–290 (1997; Zbl 0910.53034); *H. Shima*, J. Math. Soc. Japan 47, No. 4, 735–753 (1995; Zbl 0845.53033); *H. Shima*, Sémin. Gaston Darboux Géom. Topologie Différ. 1988–1989, 1–48 (1989; Zbl 0696.53030); Ann. Inst. Fourier 36, No. 3, 183–205 (1986; Zbl 0586.57013); Prog. Math. 14, 385–392 (1981; Zbl 0481.53038); Ann. Inst. Fourier 30, No. 3, 91–128 (1980; Zbl 0424.53023); Osaka J. Math. 15, 509–513 (1978; Zbl 0415.53032); J. Math. Soc. Japan 29, 581–589 (1977; Zbl 0349.53036); *S. Amari* and *J. Armstrong*, Differ. Geom. Appl. 33, 1–12 (2014; Zbl 1325.53028); Lect. Notes Comput. Sci. 9389, 240–247 (2015; Zbl 1396.53048)]. Any statistical model is to be isostatistically immersed into a standard model stemming from an Amari-Chentsov structure, providing the link between differential geometry discussed in Chapter 4 and functional analysis discussed in Chapter 3.

Chapter 5 shows that the Fisher metric and Amari-Chentsov tensor are characterized by their invariance under sufficient statistics, which is to be regarded as one of the main results of information geometry. Therein the authors discuss estimators and derive a general version of the Cramér-Rao inequality.

The concept of Shannon information is to be related to the entropy concept of Boltzmann and Gibbs, thereby yielding a natural connection between information geometry and statistical mechanics. Information geometry lies at the foundations of such important areas of mathematical biology as the theory of replicator equations and population dynamics. The information-geometric approach is particularly efficient in natural gradient methods [*S. Amari*, Information geometry and its applications. Tokyo: Springer (2016; Zbl 1350.94001)], which greatly improve classical gradient-based algorithms in exploitation of the natural geometry of the Fisher metric.

Information geometry abounds with applications. Chapter 6 addresses some fields of applications of information geometry with the authors' flavor emphasized. The chapter consists of four sections, the first of which addresses the famous problem of the whole of a complex system being more than the sum of its parts (usually attributed to celebrated ancient Greek philosopher Aristotle) from an information-geometric viewpoint based upon [*N. Ay*, Entropy 17, No. 4, 2432–2458 (2015; Zbl 1338.94032); Ann. Probab. 30, No. 1, 416–436 (2002; Zbl 1010.62007); *N. Ay* and *A. Knauf*, Kybernetika 42, No. 5, 517–538 (2006; Zbl 1249.82011)]. Its second section addresses evolutionary dynamics, which studies the change in time of relative frequencies of various types in a population. Formally speaking, such relative frequencies are to be regarded as probabilities, leading to dynamical systems on the probability simplex, which naturally involve the natural metric called the Fisher metric in mathematical statistics and the *Shahashahani metric* in evolutionary dynamics [*S. Shahshahani*, Mem. Am. Math. Soc. 211, 34 p. (1979; Zbl 0473.92008)]. These yield systems of ordinary differential equations called *replicator equations*, which, if perturbed by noise, result in what physicists call *Fokker-Plank equations* while mathematicians refer to as *Kolgomorov equations*. Its third section is devoted to applications of the Fisher metric to Monte Carlo methods [*T. Bui-Thanh* and *M. Girolami*, Inverse Probl. 30, No. 11, Article ID 114014, 23 p. (2014; Zbl 1306.65269)]. The last section puts some information-geometric constructions, particularly those discussed in §4.3, into the context of Gibbs families in statistical mechanincs, concluding that the entropy and the free energy are dual to each other (more explicitly, the Lagrange multipliers θ^i are the derivatives of the negative entropy with respect to the expectation values η_j while the expectation values are the derivatives of the negative free energy with respect to the Lagrange multipliers).

The book is concluded with three appendices on measure theory, Riemannian geometry and Banach manifolds.

Last but not least, it is exciting to note that no sooner has the book appeared than a new journal devoted completely to information geometry is published from Springer (Editor in Chief: Shinto Eguchi, ISSN 2511-2481 and 2511-249X). The book as well as [Zbl 1350.94001] will remain standard textbooks on information geometry in the forseeable future and will become the classics in the arena afterwards.

Reviewer: Hirokazu Nishimura (Tsukuba)

MSC:

53-02 Research monographs (differential geometry)
62-02 Research monographs (statistics)
53B99 Local differential geometry

Cited in **7** Documents

Keywords:

Kullback-Leibler divergence; Fisher information; Riemannian geometry; information geometry; Cramér-Rao inequality;statistical curvature; α-connection; dual connections; e-connection; Amari-Chenstov connections; Fisher metric; m-connections; Kähler geometry; Hessian geometry; Shahashahani metric; replicator equations; Fokker-Plank equations; Kolgomorov equations; Monte Carlo methods; Banach manifolds

Full Text: DOI