BULLETIN (New Series) OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 55, Number 1, January 2018, Pages 109–111
http://dx.doi.org/10.1090/bull/1582
Article electronically published on May 26, 2017

Geometric modeling in probability and statistics, by O. Calin and C. Udriste, Springer, Cham, Switzerland, 2014, ISBN 978-3-319-07778-9 (print), 978-3-319-07779-6 (online), $79.99

The geometric ideas developed in the 19th century still exert a fundamental influence on modern mathematics. This is especially true in theoretical and mathematical physics, including the areas of special and general relativity, tensor analysis in mechanics and hydrodynamics, and discrete and continuum groups in solid state and quantum theories.

In the more analytic branches of mathematics, the process of "geometrization" has been much slower. For instance, in the famous paper by A. N. Kolmogrov [4], "On analytic methods in probability theory", the generator for a diffusion process in $\mathbb{R}^d$ was written in the form

$$\mathcal{L} = \sum_{i,j=1}^{d} a_{ij}(x) \frac{\partial^2}{\partial x_i \partial x_j} + \sum_{i=1}^{d} b_i(x) \frac{\partial}{\partial x_i}, \tag{1}$$

where $A(x) = [a_{ij}(x)]$ is called the diffusion matrix and $(b_i(x))$ the drift vector. This terminology is still used today in the overwhelming majority of textbooks on the theory of Markov processes, stochastic differential equations, etc. However, a diffusion process exists independently of the selection of a coordinate system. Thus, its description should be covariant. Physicists understood this fact earlier than mathematicians, and they presented the generator in the so-called Fokker–Planck form,

$$\mathcal{L} = \sum_{i,j=1}^{d} \frac{\partial}{\partial x_i} \left(a_{ij}(x) \frac{\partial}{\partial x_j} \right) + \sum_{i=1}^{d} \tilde{b_i}(x) \frac{\partial}{\partial x_i}. \tag{2}$$

Here $A(x) = [a_{ij}(x)]$ is a tensor, and $B(x) = [\tilde{b}_i(x)]$ is a vector. Note that $A(x)$ can be nonsmooth and even discontinuous, which is important for applications to composite matrices. The theory of parabolic equations $\frac{\partial u}{\partial t} = \mathcal{L}u$ with $\mathcal{L}$ presented in "divergent" form in (2) was developed only in the second half of the 20th century; see J. Nash [3], J. Moser [6], D. Aronson [2]. In many senses, this theory is better than the old theory (using $\mathcal{L}$ in the form (1)). That theory requires regularity on the coefficients (they must be at least of the Hölder class) together, of course, with the symmetry and nondegeneracy of the matrix $A(x)$.

The matrix $A(x) = [a_{ij}(x)]$, under mild regularity conditions, guarantees that we can define the Riemannian metric $ds^2 = \alpha^{ij} dx_i dx_j$ in $\mathbb{R}^d$, and (in the absence of drift $\tilde{b}$ in (2)) the operator $\mathcal{L} = \frac{\partial}{\partial x_i}(a_{ij}(x))\frac{\partial}{\partial x_j}$ is the Laplace–Beltrami operator on the Riemannian mannifold with metric form ds^2.

It is interesting that A. N. Kolmogorov (after an exchange of information with E. Schrödinger; see [7]) published the important paper [5], completely based on the differential geometry approach and generators of the form (2).

Geometric ideas in statistics appeared much later and were related mainly to information theory. In fact the first monographs in the field of geometric modeling

2010 *Mathematics Subject Classification.* Primary 53B20, 53C21, 53C22, 60D99, 62B10.

appeared (in English) only around 2000 (Amari and Nagaoka [1]; Kass and Vos [8]). In the theory of statistical estimators, the starting point was the Cramer–Rao inequality (mid-1940s). In the simplest case of an unbiased estimator for an unknown scalar parameter ξ, it has the following form. Let $X_1, \dots, X_n$ be the sample of i.i.d. random variables with density $p(x, \xi)$, and let $T(X_1, \dots, X_n)$ be the estimator of θ, such that $E_\xi T = \xi$. Then, under some regularity conditions, $E[T-\xi]^2 \geqslant \frac{1}{n} I(\xi)$, where $I(\cdot)$ is the Fisher information given by

$$I(\xi) = E_\xi \left(\frac{\partial \ln(p(X_1, \xi)}{\partial \xi} \right)^2 . \tag{3}$$

The estimator $T(X_1, \dots, X_n)$ is called efficient if $E[T-\xi]^2 = \frac{1}{n} I(\xi)$ and asymptotically efficient if $E[T-\xi]^2 \to \frac{1}{n} I(\xi)$ as $n \to \infty$. (Of course, one must assume that $I(\xi) < \infty$.)

In the case of vector-valued parameter $\vec{\xi} = (\xi^1, \dots, \xi^n)$ and the smooth parametric family $p_{\vec{\xi}}(x)$, one can introduce the similar object: the Fisher information matrix

$$g_{ij}(\xi) = E_\xi \left[\left(\partial_{\xi^i} \ln p_\xi(x, \xi) \right) \left(\partial_{\xi^j} \ln p(x, \xi) \right) \right] . \tag{4}$$

In typical situations (smoothness, nondegeneracy, etc.) the matrix $[g_{ij}(\xi)]$ defines in $\mathbb{R}^n$ a Riemannian metric form with covariant positively definite tensor $g_{ij}(\xi)$

$$ds^2 = g_{ij}(\xi) d\xi^i d\xi^j . \tag{5}$$

As a result, we embed our statistical model into a Riemannian manifold, and we can now define an information distance $d(P, Q)$ for two distributions on our family and define standard geometric quantities including curvature tensor, Levi-Civita connection, etc.

A typical example is given by the normal (Gaussian) distribution $N(\mu, \sigma^2)$ with the density

$$p(x, \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2} \right) . \tag{6}$$

Here $\vec{\xi} = (\mu, \sigma)$ belongs to the upper half-plane with the Riemannian (Introduction) metric:

$$ds^2 = \frac{d\mu^2 + d\sigma^2}{\sigma^2} , \tag{7}$$

i.e., it is the Poincaré model of the hyperbolic plane.

Another source of geometric ideas in statistics is the testing of hypothesis. Assume that we have two probability distributions $P(dw), Q(dw)$ on the same measurable space $(\Omega, \mathcal{F})$. Due to Neyman–Pearson theory, all useful information on the goodness of fit of these two laws is contained in the logarithm of the likelihood ratio $\ln \frac{dP}{dQ}(w)$. The expectation of this ratio is given by

$$D_{KL}(p \| q) = \int_\Omega l(w) P(dw) = \int_\Omega \frac{P(dw)}{Q(dw)} \ln \left(\frac{P(dw}{Q(dw)} \right) Q(dw), \tag{8}$$

and it is known as the Küllback–Leibler relative entropy. In contrast to the notation, this expression is not a metric (due to asymmetry of H_0 and H_1 in the testing of statistical hypotheses).

The book under review is divided into two sections. The first, entitled "The geometry of statistical models", contains six chapters. Together with an introduction to the general topics (including probability spaces, entropy, and information), it contains many examples of the Riemannian metrics associated to classical discrete and continuous parametric models (Gamma and Beta distributions, Geometric and Poisson laws, etc). It also contains detailed derivations of the Fisher information matrices, Küllback–Leibler relative entropy, and related quantities.

The second section, entitled "Statistical manifolds" is purely geometrical. Chapters 7–10 describe standard geometric objects, including Riemannian manifolds and divergence of vector fields. The three remaining chapters, 11–13, contain the theory of so-called "contrast functions", which are distance-like nonnegative functions $D(Q\|P)$ of two distributions P and Q, which are not necessarily symmetric and do not necessarily satisfy the triangle inequality, but vanish if and only if $P = Q$. This is the generalization of Kullback–Leibler relative entropy. Each smooth contrast functional generates an associated Levi-Civita connection and Riemannian metric. This section of the book is illustrated with many examples of statistical manifolds and submanifolds from the first section. Further, each chapter in both sections is followed by numerous exercises, which are interesting not only from a probabilistic but also from a purely analytic point of view.

It is necessary to mention that, while useful, this book is not a comprehensive treatise. For instance, the fundamental Cramer–Rao inequality, which is one of the strongest motivations for the development of this theory, is only briefly mentioned. The authors describe this area as "informational geometry".

In diffusion processes theory, Riemann's geometrical ideas have been highly fruitful. In statistics right now, it looks like a new language to describe parametric models. One hopes that it will find important applications as well.

The author would like to thank Dr. Michael Grabchak for his help in the preparation of this review.

References

[1] S.-I. Amari and H. Nagaoka, *Methods of information geometry*, Translations of Mathematical Monographs, vol. 191, American Mathematical Society, Providence, RI; Oxford University Press, Oxford, 2000. Translated from the 1993 Japanese original by Daishi Harada. MR1800071

[2] D. G. Aronson, *The fundamental solution of a linear parabolic equation containing a small parameter*, Illinois J. Math. **3** (1959), 580–619. MR0107758

[3] J. Nash, *Continuity of solutions of parabolic and elliptic equations*, Amer. J. Math. **80** (1958), 931–954, DOI 10.2307/2372841. MR0100158

[4] A. Kolmogoroff, *Über die analytischen Methoden in der Wahrscheinlichkeitsrechnung* (German), Math. Ann. **104** (1931), no. 1, 415–458, DOI 10.1007/BF01457949. MR1512678

[5] A. Kolmogoroff, *Zur Umkehrbarkeit der statistischen Naturgesetze* (German), Math. Ann. **113** (1937), no. 1, 766–772, DOI 10.1007/BF01571664. MR1513121

[6] J. Moser, *A Harnack inequality for parabolic differential equations*, Comm. Pure Appl. Math. **17** (1964), 101–134, DOI 10.1002/cpa.3160170106. MR0159139

[7] E. Schrödinger, *Uber die Umkehrung der Naturgesetze*. Sitzungsber Press, Acad. Wiss, Phys-math, K1, 12 märz 1931, 144–153 .

[8] R. E. Kass and P. W. Vos, *Geometrical foundations of asymptotic inference*, Wiley Series in Probability and Statistics: Probability and Statistics, John Wiley & Sons, Inc., New York, 1997. A Wiley-Interscience Publication. MR1461540

Stanislav A. Molchanov

Department of Mathematics and Statistics, UNCC, Charlotte, North Carolina 28223; and National Research University "Higher School of Economics", Russian Federation

E-mail address: smolchan@uncc.edu