



Soviet-era science, translated into English

MATHEMATICS

1964

SovietRxiv

View the original and related papers at <https://sovietrxiv.org/items/ru-196401.65708>

Source: Math-Net.Ru and CyberLeninka. Machine translation. Verify with the original.

Abstract

Full Text

MATHEMATICS

N. N. Chentsov

GEOMETRY OF THE “MANIFOLD” OF PROBABILITY DISTRIBUTIONS

(Presented by Academician A. N. Kolmogorov, 9 IV 1964)

1°. Let Ω be an arbitrary set of elementary outcomes ω with a σ -algebra S of measurable subsets. Let, further, I be an ideal of the algebra S . Consider the maximal “manifold” H of probability distributions $P\{d\omega\}$, for each of which the ideal I is the class of all events of zero probability. Our aim is to introduce an “affine connection” in H and to study the “geometry” of H . When Ω consists of a finite number of atoms, H is a manifold in the exact sense of the word. Generally speaking, the geometry under study is infinite-dimensional.

2°. Embed H in the space W of all measures of σ -finite variation, necessarily annihilating I , with cone C of strictly positive measures annihilating only I . By $V \subset W$ denote the space of measures of finite variation with positive cone $K = V \cap C$. The operation of Radon–Nikodym differentiation, for fixed $\mu \in C$, assigns to each measure $\nu \in W$ a function of a point—its derivative

$$\nu\{\cdot\} \leftrightarrow \frac{d\nu}{d\mu}(\cdot), \quad (1)$$

defined up to values on a set of μ -measure zero; see ⁽¹⁾. If measures are called vectors, and equivalence classes modulo I of functions are called covectors—elements of the space W^* —then the bilinear operation

$$\langle f, \nu \rangle = \int_{\Omega} f(\omega) \nu\{d\omega\}, \quad (2)$$

which on H becomes the definition of mathematical expectation, may be called the “inner product” of the covector f and the vector ν . The “manifold” H is singled out from the cones C and K by the condition

$$\langle 1, P \rangle = 1. \quad (3)$$

With the “surface” H at each “point” P are associated the spaces $L(P) = \{f : \langle |f|, P \rangle \text{ finite}\}$; $N(P) = \{f : \langle f, P \rangle = 0\}$.

The tangent space $\Lambda(P)$ to H at the “point” P , generated by the differences $P_\alpha - P$, does not depend on the choice of P , $\Lambda = \{R : \langle 1, R \rangle = 0\}$.

3°. Since the linear operator (1) puts every vector in correspondence with its covector and conversely, it is natural to say that it defines a twice covariant tensor. Further, we may consider two quadratic forms

$$(Q, R)_P = \int \frac{dQ}{dP}(\omega) \frac{dR}{dP}(\omega) P\{d\omega\}, \quad (4)$$

$$(f, g)^P = \int f(\omega)g(\omega)P\{d\omega\}, \quad (5)$$

which define scalar products associated with the point P . It seems tempting (see (2, 3)) to declare H a “Riemannian space” with “metric” (4). However, such an approach does not capture all the “statistical” properties

families of probability distributions. For the moment we shall only use the isomorphism

$$W \sim W^*, \quad C \sim C^*, \quad V \sim L(P), \quad \Lambda \sim N(P), \quad P \sim 1, \quad (6)$$

established by the operator (1), in order to decree the line $\lambda \cdot P\{\cdot\}$ to be normal to the “surface” H at the point P .

4°. To each parallel translation of the space $W^* : f(\cdot) \xrightarrow{g} f(\cdot) + g(\cdot)$ there corresponds an automorphism of the cone C

$$\mu\{\cdot\} \xrightarrow{g} \nu\{\cdot\}, \quad \text{where } \frac{d\nu}{d\mu}(\omega) = \exp(g(\omega)). \quad (7)$$

The commutative group G of all “translations” of the cone C is simply transitive, for $\mu\{\cdot\} \xrightarrow{g} \nu\{\cdot\}$ if and only if $g(\omega) = \ln \frac{d\nu}{d\mu}(\omega)$. The cone K is not invariant with respect to G . But, since for every g the intersection $K \cap g(K)$ is nonempty, G is a pseudogroup of (generalized) “automorphisms” of K , see (4). H is a “hypersurface” in K . The “automorphism” of the tangent space Λ , induced in the Levi-Civita sense (see (5)) by the translation $P_1\{\cdot\} \rightarrow P_2\{\cdot\}$, is written by the formula

$$dR_1 \rightarrow dR_2 = \left(\frac{dR_1}{dP_1} - c \right) dP_2,$$

$$R_2(B) = \int_B \frac{R_1\{d\omega\}P_2\{d\omega\}}{P_1\{d\omega\}} - P_2\{B\} \int_\Omega \frac{R_1\{d\omega\}P_2\{d\omega\}}{P_1\{d\omega\}}. \quad (8)$$

If in (8) one passes to densities, the formula is simplified:

$$\frac{dR_1}{dP_1} \rightarrow \frac{dR_1}{dP_1} - c = \frac{dR_2}{dP_2}. \quad (9)$$

The conjugate transformation preserving the “inner product” of the dual subspace $N(P)$ is written as $f_1(\omega) \rightarrow f_2(\omega) = f_1(\omega) \frac{dP_1}{dP_2}(\omega)$.

The transitive pseudogroup of “automorphisms” (7) generates a fibered space with base H and fibers $\Lambda(P)$. In the “affine connection” induced in H by this fibration the parallelism of vectors is absolute (see (5)), so that the connection is flat and even locally affine (see (6)).

5°. Fix an arbitrary measure $\mu \in H$, and write $p(\omega) = \frac{dP}{d\mu}(\omega)$. By virtue of (9), the condition of parallel transfer of a tangent vector along a geodesic line $p(\omega; t)$ is written

$$\lambda(t) \frac{p'_t(\omega; t)}{p(\omega; t)} = \frac{p'_t(\omega; 0)}{p(\omega; 0)} - \psi(t). \quad (10)$$

By a known change of variable (see (5, 6)) one may pass to the canonical affine parameter s

$$q(\omega; s) = \frac{\partial}{\partial s} \ln p(\omega; s) = q(\omega; 0) - \varphi(s), \quad (11)$$

where s is defined up to a linear substitution $s \rightarrow as + b$. From the “first integral” (11) it is not hard to obtain an explicit expression for the “geodesic family” $p(\omega; s)$. We shall use, however, the fact that the geometry of H is the “projective geometry” of the pencil of lines of the cone C with translations (7), since a direct calculation verifies that the “projective” definition of the parallel transport of infinitely small tangent vectors coincides with (8). Geodesics in C —the trajectories of one-parameter subgroups of the group G —have the form

$$f(\omega; s) = f(\omega; 0) \exp[s \cdot g(\omega)]. \quad (12)$$

Geodesics in H must be “logarithmic projections” onto H of these trajectories, i.e., differ from $f(\omega; s)$ by a normalizing factor. Thus, a geodesic passing through a given point $p(\omega; 0)$ in a given direction is described by the formula

$$p(\omega; s) = \frac{1}{a(s)} p(\omega; 0) \exp[s \cdot q(\omega; 0)], \quad (13)$$

where $a(s)$ is the normalizing constant

$$a(s) = M_0 \exp[s \cdot q(\omega; 0)] = \int \exp[s \cdot q(\omega; 0)] p(\omega; 0) d\mu, \quad (14)$$

related to $\varphi(s)$ by the relation

$$\varphi(s) = \frac{d}{ds} \ln a(s) = M_s q(\omega; 0). \quad (15)$$

A geodesic passing through two points is described in barycentric affine coordinates by the formula

$$p(\omega; \tau) = \frac{1}{b(\tau)} [p(\omega; 0)]^{1-\tau} [p(\omega; 1)]^\tau, \quad (16)$$

where $b(\tau)$ is again a normalizing constant

$$b(\tau) = \int [p(\omega; 0)]^{1-\tau} [p(\omega; 1)]^\tau d\mu. \quad (17)$$

6°. For some values of s the trajectory may leave the cone K , and the corresponding densities (and measures) cannot be normalized. The normalizing constant $A(s)$ of the trajectory (12) coincides with the moment-generating function of the distribution of the values of the function $g(\omega)$ with respect to the measure with density $f(\omega; 0)$ (see (7))* . The following proposition is valid:

If the densities $f_0(\omega)$ and $f_1(\omega)$ correspond to measures $\nu_0 \in K$ and $\nu_1 \in K$, then all points ν_z (with densities $f_0^{1-z} \cdot f_1^z$) of the segment $0 \leq z \leq 1$ (of the region $0 \leq \operatorname{Re} z \leq 1$) of the trajectory passing through ν_0 and ν_1 also belong to K . The normalizing constant $A(z) = \langle 1, \nu_z \rangle$ is an analytic function of z in the strip $0 < \operatorname{Re} z < 1$ and is continuous for $0 \leq \operatorname{Re} z \leq 1$.

A simple consequence of this is the theorem:

“The manifold H is geodesically convex.”

By virtue of convexity, the geodesic in H corresponding to a given trajectory in C may be: a) the whole affine line, b) a ray, c) a segment, d) a point, e) the empty set, with the boundary values in cases b) and c) possibly excluded. Within the region of existence the geodesic is an analytic line. At boundary points the analytic formulas valid inside may lose their meaning.

7°. The quantity $a(s)$ first decreases monotonically to $a(0) = 1$, then increases monotonically, no more slowly than an exponential; $a^{(n)}(s) = a(s) M_s [q(\omega; 0)]^n = a(s) \int [q(\omega; 0)]^n p(\omega; s) d\mu$. For $s = 0$, $a'(0) = 0 = \varphi(0)$; $a''(0) = \varphi'(0) = D_0 q(\omega; 0) = I_F$; the latter quantity is called the Fisher information quantity ^(2,8). The quantity $b(\tau)$ changes in a similar way:

$$b(\tau) = \int \left[\frac{dP_1}{dP_0}(\omega) \right]^\tau P_0\{d\omega\} = \int \left[\frac{dP_0}{dP_1}(\omega) \right]^{1-\tau} P_1\{d\omega\}.$$

At first $b(\tau)$ decreases monotonically to a certain minimal value $b(\tau_0) = \rho(P_0, P_1) < 1$, $0 < \tau_0 < 1$, then increases monotonically. In (7) the quantity $I_D = -\ln \rho$ was proposed for use as a measure of the difference between distributions (hypotheses) P_0 and P_1 . If we denote $l(\omega) = \ln p(\omega; 1) - \ln p(\omega; 0)$, then $b^{(n)}(\tau) = b(\tau)M_\tau[l(\omega)]^n$. The quantity $-b'(0) = -M_0l(\omega) = I_C$ is called the Chernoff information quantity on the difference of P_0 from P_1 (8). This concept was generalized in (9), where information quantities of order τ were introduced:

$$I_R(\tau) = \frac{1 - b(\tau)}{\tau},$$

* If $f(\omega; 0)$ is the density of the probability distribution of random outcomes ω , $A(is)$ is the characteristic function of the random variable $g(\omega)$.

$I_R(0) = I_C$. As is known (see (8)), under the corresponding limiting transition $I_C \rightarrow I_F$, when $P_1 \rightarrow P_0$ along a geodesic. Formulas (13) and (16) are related by the relations $q(\omega; 0) = l(\omega) - M_0l(\omega)$; $\varphi(\tau) = M_\tau l(\omega) - M_0l(\omega)$.

For the n -th semi-invariant $\varkappa_s^n q(\omega)$ of the distribution of the values of the function $q(\omega)$ with respect to the distribution of outcomes ω with density $p(\omega; s)$,

$$\varphi^{(k)}(s) = \varkappa_s^k q(\omega; 0) = \varkappa_s^k q(\omega; s), \quad k \geq 1, \quad (18)$$

whence the canonical equation of the geodesic has the form

$$\frac{\partial q(\omega; s)}{\partial s} + D_s q(\omega; s) = \frac{\partial q(\omega; s)}{\partial s} + \int [q(\omega; s)]^2 p(\omega; s) d\mu = 0. \quad (19)$$

8°. By virtue of local affinity, H decomposes into completely geodesic surfaces. For example, the geodesic surface passing through three points is described by the formula

$$p(\omega; \alpha, \beta, \gamma) = \frac{1}{b(\alpha, \beta, \gamma)} [p_1(\omega)]^\alpha [p_2(\omega)]^\beta [p_3(\omega)]^\gamma, \quad \alpha + \beta + \gamma = 1. \quad (20)$$

Any geodesic passing through two "points" of this family also belongs to it. The two-dimensional analogue of formula (13) has the form $\ln p(\omega; s, t) = -\ln a(s, t) + tq_t(\omega; 0, 0) + sq_s(\omega; 0, 0)$.

A geometric characterization of the rank r of a family of mutually absolutely continuous probability distributions $p(\omega; \theta)$ (see (10)) is given by the theorem:

If a family $p(\omega; \theta)$ has rank r , then the convex geodesic hull of the family is a domain of some r -dimensional geodesic surface, and conversely.

9°. In H there exist metric connections compatible with the affine connection (8). It suffices to specify a differential quadratic form at one point μ and then transport it parallelly to all the others. For example, if at the point the metric tensor is the Radon–Nikodym tensor (1), then

$$(Q, R)_P = \text{Cov}_\mu \left(\frac{dQ}{dP}, \frac{dR}{dP} \right) = \int \left[\frac{dQ}{dP} - \int \frac{dQ}{dP} d\mu \right] \left[\frac{dR}{dP} - \int \frac{dR}{dP} d\mu \right] d\mu. \quad (21)$$

It follows from this, in particular, that tensor (1) is not stationary. The length of a curve in any of these metrics is a measurement of the corresponding affine parameter.

10°. Many properties of remarkable families of probability distributions are connected with the fact that they are geodesic families. These include Gaussian and Poisson distributions. It is interesting to note that all Pearson distributions (¹¹), given by the equation

$$y'(x) = t \frac{x - \alpha}{(x - \beta)(x - \gamma)} y(x),$$

naturally decompose into geodesics $y(x; t)$. The scope of this note does not allow us to touch upon other questions where the geometric approach also proves useful.

In conclusion, the author considers it a pleasant duty to express gratitude to E. A. Morozova for a valuable discussion.

Received
9 IV 1964

CITED LITERATURE

1. P. Halmos, *Measure Theory*, IL, 1953.
2. S. Kullback, *Ann. Math. Stat.*, **25**, No. 4, 745 (1954).
3. N. N. Chentsov, *DAN*, **147**, No. 1, 45 (1962).
4. S. E. Ehrešmann, S. S. Chern, *Géométrie différentielle*, Colloq. intern. du C.N.R.S., Paris, 1953.
5. A. P. Norden, *Spaces of Affine Connection*, M.–L., 1950.

6. J. Favard, *Course of Local Differential Geometry*, II, 1960.
7. H. Chernoff, *Ann. Math. Stat.*, **23**, No. 4, 493 (1952).
8. S. Kullback, R. A. Leibler, *Ann. Math. Stat.*, **22**, No. 1, 79 (1951).
9. A. Rényi, *Proc. 4-th Berkeley Simpos. Math. Stat. and Prob.*, 1, 1961, p. 547.
10. E. B. Dynkin, *UMN*, **6**, issue 1, 66 (1951).
11. G. Cramér, *Mathematical Methods of Statistics*, II, 1948.

Note: Figure translations are in progress. See original paper for figures.

Source: Math-Net.Ru and CyberLeninka. Machine translation. Verify with the original.