



Soviet-era science, translated into English

MATHEMATICS

1965

SovietRxiv

View the original and related papers at <https://sovietrxiv.org/items/ru-196501.67159>

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

Abstract

Full Text

MATHEMATICS

N. N. CHENTSOV

CATEGORIES OF MATHEMATICAL STATISTICS

(Presented by Academician A. N. Kolmogorov, 25 II 1965)

Let us agree to denote by $\text{Cap}(\Omega, S)$ the totality of all probability distributions on the measurable space (Ω, S) , i.e., the totality of all normalized measures on the σ -algebra S of measurable subsets of the space Ω of elementary outcomes ω . For any pair of objects $\text{Cap}(\Omega_1, S_1)$ and $\text{Cap}(\Omega_2, S_2)$ there exist Markov (homo-)morphisms $\Pi : \text{Cap}_1 \rightarrow \text{Cap}_2$, assigning to each distribution $P \in \text{Cap}_1$ its distribution $Q \in \text{Cap}_2$ by the formula

$$Q\{\cdot\} = \int_{\Omega_1} \Pi\{\cdot \mid \omega'\} P\{d\omega'\}. \quad (1)$$

The transition measure $\Pi\{A \mid \omega'\}$ in (1) is an S_1 -measurable function on Ω_1 for fixed $A \in S_2$, and a normalized measure (probability distribution) on S_2 for fixed $\omega' \in \Omega_1$.

Markov morphisms form a multiplicative system:

Cat 1. For every object $\text{Cap}(\Omega, S)$ there exists and is unique the identity automorphism. It is given by the transition measure $\Pi\{A \mid \omega\} = \chi_A(\omega)$, the indicator of the variable set A .

Cat 2. The composition of Markov morphisms $\Pi_{12}(\text{Cap}_1 \rightarrow \text{Cap}_2)$ and $\Pi_{23}(\text{Cap}_2 \rightarrow \text{Cap}_3)$ is defined and is, by Fubini's theorem, the Markov morphism $\Pi_{13}(\text{Cap}_1 \rightarrow \text{Cap}_3)$ with transition measure

$$\Pi_{13}\{A \mid \omega'\} = \int_{\Omega_2} \Pi_{12}\{d\omega'' \mid \omega'\} \Pi_{23}\{A \mid \omega''\}.$$

Thus, the objects $\text{Cap}(\Omega, S)$ with the system of Markov morphisms form a category, see ⁽¹⁾, the category CAP of totalities of probability distributions.

The objects of study in mathematical statistics are families $w = \{w_\theta\}$ of probability distributions $P_\theta\{d\omega\}$ on one and the same measurable space (Ω, S) , i.e., all possible subsets of the totalities $\text{Cap}(\Omega, S)$. The following "geometric picture" is obtained. There are "spaces" $\text{Cap}(\Omega, S)$ and a system of mappings. One

studies the properties of the “figure” w that are invariant with respect to the “invertible” mappings of this figure. The relations that arise can be described in the language of categories, since the families w themselves form a category with a system of Markov morphisms*, or, more precisely, even two categories FAM_1 and FAM_0 , depending on whether the method of parametrization is included in the definition of the family w or not. In the category FAM the usual equivalence relation of objects is made concrete as follows.

Definition 1. Two families w_1 and w_2 of probability spaces

$$w_{\theta^{(1)}}^{(1)} = (\Omega_1, \mathcal{S}_1, P_{\theta^{(1)}}), \quad w_{\theta^{(2)}}^{(2)} = (\Omega_2, \mathcal{S}_2, P_{\theta^{(2)}})$$

are called **equivalent** if they are parametrized in the same way and there exist

* The conversion of the collection of objects FAM into a category is achieved at the cost of complicating the multiplicative system of morphisms: each morphism $w_1 \rightarrow w_2$ is indexed by an integral kernel $\Pi\{\cdot | \cdot\}$ and by the object w_1 ; morphisms $w_1 \rightarrow w_2$ that define identical mappings of the elements of w_1 into the elements of w_2 are identified.

transition measures $\Pi_{12}\{d\omega'' | \omega'\}$ and $\Pi_{21}\{d\omega' | \omega''\}$ such that $\mathbf{w}_1\Pi_{12} = \mathbf{w}_2$ and $\mathbf{w}_2\Pi_{21} = \mathbf{w}_1$ ($w_{\theta}^{(1)}\Pi_{12} = w_{\theta}^{(2)}$, $w_{\theta}^{(2)}\Pi_{21} = w_{\theta}^{(1)}$ for all θ).

We shall show that the geometric approach described above in the style of Klein arises naturally in the theory of statistical decisions⁽²⁾, while Definition 1 describes families of distributions with “identical” statistical properties. Let, in addition to the family $\mathbf{w} = \{w_{\theta}\}$ of probability distributions $P_{\theta}\{d\omega\}$ of outcomes ω of observations, a measurable space (\mathcal{E}, B) of inferences ε be given, and let some stochastic law Π of inference be specified—the probability distribution $\Pi\{d\varepsilon | \omega\}$ of inferences ε for a known outcome ω , cf. ⁽³⁾. Each rule Π induces, precisely by formula (1), the family

$$Q_{\theta}^{(\Pi)}\{\cdot\} = \int_{\Omega} \Pi\{\cdot | \omega\} P_{\theta}\{d\omega\} \quad (2)$$

of probability distributions of inferences, or, for short, a certain statistical decision for the problem \mathcal{E} for the family \mathbf{w} . All families $Q_{\theta} \subset \text{Cap}(\mathcal{E}, B)$ representable in the form (2) for the given \mathbf{w} and all possible Π form the class $D(\mathcal{E}, \mathbf{w})$ of all decisions of the problem \mathcal{E} for \mathbf{w} .

Example. When the space of inferences is the set Θ of all values of the parameter θ , the problem Θ for the family $\{w_{\theta}\}$ is the problem of estimating, from the observed outcome ω , the unknown parameter of the distribution law P_{θ} of the observed outcomes.

If one disregards the complexity and cost of implementing the inference procedure, it is natural to assume that:

Definition 2. Two families \mathbf{w}_1 and \mathbf{w}_2 are statistically equivalent when, for every problem \mathcal{E} , the classes $D(\mathcal{E}, \mathbf{w}_1)$ and $D(\mathcal{E}, \mathbf{w}_2)$ are identical.

Theorem 1. *Statistical equivalence of families of distributions is equivalence of objects of the category FAM.*

Let us state a number of propositions of “Markov geometry.”

Theorem 2. *Any two simple families, i.e. each consisting of a single distribution, are statistically equivalent.*

If $\mathbf{w}_1 \Pi_{12} = \mathbf{w}_2$ and $\mathbf{w}_2 \Pi_{21} = \mathbf{w}_1$, then $\mathbf{w}_1 \Pi_{12} \Pi_{21} = \mathbf{w}_1$ and $\mathbf{w}_2 \Pi_{21} \Pi_{12} = \mathbf{w}_2$, i.e. the families \mathbf{w}_1 and \mathbf{w}_2 are subsets of stationary “vectors” of the nonnegative linear transformations $\Pi_{12} \Pi_{21}$ and $\Pi_{21} \Pi_{12}$.

Lemma 1. *Let V_1 and V_2 be lineals of all measures of bounded variation, stationary respectively for the Markov linear operators $\Pi_{12} \Pi_{21}$ and $\Pi_{21} \Pi_{12}$, and let $v_1 = V_1 \cap \text{Cap}_1$, $v_2 = V_2 \cap \text{Cap}_2$. Then the operators Π_{12} and Π_{21} establish a one-to-one linear correspondence between the “points” of the lineals V_1 and V_2 , and also between the “points” of the families v_1 and v_2 .*

For the time being we restrict ourselves to collections of all probability distributions on sets Ω that have only a finite number of outcomes and all subsets of which are measurable. Such collections form a full subcategory in CAP, which we shall denote by CAPF. The multiplicative system of Markov operators of the category CAPF consists of stochastic matrices. The structure of the set of stationary distributions of a finite Markov chain is well known, cf. (4). This makes it possible to give a complete description of equivalent families of the category FAMF.

Theorem 3. *Two families $\mathbf{w}_1 \subset \text{Capf}(\Omega_1)$ and $\mathbf{w}_2 \subset \text{Capf}(\Omega_2)$ of probability distributions are equivalent if and only if the families of factors—the distributions of their minimal sufficient statistics—coincide:*

$$P_\theta^{(1)}\{\omega'\} = R_\theta\{\varepsilon(\omega')\} \cdot Q^{(1)}\{\omega' \mid \varepsilon\},$$

$$P_\theta^{(2)}\{\omega''\} = R_\theta\{\varepsilon(\omega'')\} \cdot Q^{(2)}\{\omega'' \mid \varepsilon\};$$

here $Q^{(i)}$ does not depend on θ , and R_θ does not depend on i .

For the entire category FAM one can formulate only a weaker proposition:

Theorem 3'. *Let the family $\{P_1, P_2\}$ be equivalent to $\{Q_1, Q_2\}$, where $Q_i = P_i \Pi_{12}$, $P_j = Q_j \Pi_{21}$. Denote $P_e = \frac{1}{2}[P_1 + P_2]$, $Q_e = \frac{1}{2}[Q_1 + Q_2]$,*

$$p(\omega') = \frac{dP_1}{dP_e}(\omega'); \quad q(\omega'') = \frac{dQ_1}{dQ_e}(\omega'').$$

Then: 1)

$$P_e p^{-1}\{\cdot\} = Q_e q^{-1}\{\cdot\},$$

2)

$$P_f \Pi_{12} = Q_f; \quad Q_f \Pi_{21} = P_f,$$

where

$$\frac{dP_f}{dP_e}(\omega') = f(p(\omega')); \quad \frac{dQ_f}{dQ_e}(\omega'') = f(q(\omega'')).$$

Analogous propositions are valid for any finite families. Let us now consider, for finite sets Ω , the cones $\text{Conf}(\Omega)$ of all nonnegative measures on the algebra of all subsets of Ω . Nonnegative linear homomorphisms $R : \text{Conf}_1 \rightarrow \text{Conf}_2$

$$\nu\{\cdot\} = \int_{\Omega_1} R\{\cdot \mid \omega'\} \mu\{d\omega'\} \quad (3)$$

already with non-normalized transition measures are specified by arbitrary non-negative matrices $\|r_j^i\|$. In complete analogy with the preceding, the cones $\text{Conf}(\Omega)$ form a category CONF with a system of morphisms of the form (3). In this case, for families of probability distributions, as for “figures” of the object Conf , there arises a broader relation of affine equivalence.

For each measure from the cone $\text{Conf}(\Omega)$ there exists a unique projection onto the “hyperplane” $\text{Capf}(\Omega)$

$$\pi\mu\{\cdot\} = \frac{1}{\mu\{\Omega\}} \mu\{\cdot\}. \quad (4)$$

The projection operator π defines a covariant functor taking objects of the category CONF to objects of the category πCAPF with a multiplicative system of positive projective morphisms of the form πR . Each projective morphism is specified by a transition measure $R\{A \mid \omega\}$, i.e., by a nonnegative matrix $\|r_j^i\|$, determined up to a positive numerical factor.

For a fixed set Cap , the system of its Markov endomorphisms forms a semigroup \mathfrak{M} with identity. It contains the subgroup \mathfrak{G} of Markov automorphisms. For $\text{Capf}(\Omega)$, the group \mathfrak{G} consists only of permutations of the outcomes ω_i of the set Ω . For the cone $\text{Conf}(\Omega)$, the group \mathfrak{P} of positive linear automorphisms is much richer. The identity component in \mathfrak{P} is the commutative simply transitive group \mathfrak{T} of positive diagonal matrices, and the factor group $\mathfrak{P}/\mathfrak{T}$ is isomorphic to \mathfrak{G} .

The cone $\text{Conf}(\Omega)$ contains the cone $C(\Omega)$ of all strictly positive measures on Ω . With respect to the group \mathfrak{T} , $C(\Omega)$ is an affine space, see (5). Denote by $H(\Omega)$ the hyperplane $C(\Omega) \cap \text{Capf}(\Omega)$. $H(\Omega)$ is not invariant with respect to the group \mathfrak{T} of translations. However, the following holds.

Theorem 4. If a figure $w_1 \subset H(\Omega)$ can be continuously moved inside $H(\Omega)$, with equivalence preserved, into a figure $w_2 \subset H(\Omega)$ by a family of Markov

endomorphisms (w_1 is \mathfrak{M} -homotopic to w_2), then this motion can be carried out by a family of translations (w_1 is \mathfrak{T} -homotopic to w_2).

The maximal group of projective endomorphisms of Capf is the factor group of the group \mathfrak{P} by the group Λ of scalar matrices. The identity component in \mathfrak{P}/Λ is simply the transitive commutative group $\mathfrak{G} = \mathfrak{T}/\Lambda$.

Theorem 5. The manifold H with the motion group \mathfrak{G} is a homogeneous locally affine space*. Its natural geometry of linear connection is intrinsic in the sense of Levi-Civita: the geometry of the hyperplane H , equipped in C with pseudonormals of the form OP at a variable point $P \in H$.

* Every homogeneous space of negative curvature has a natural boundary, see (6,7). For spaces of zero curvature, compactifications are not unique. However, if together with the connected group \mathfrak{T}/Λ one considers the full group of \mathfrak{P}/Λ -automorphisms and the continuous semigroup \mathfrak{M} of Markov endomorphisms of the open simplex $H(\Omega)$, then a reasonable compactification of $H(\Omega)$ becomes unique. The corresponding boundary was first constructed in (8), where $H(\Omega)$ was compactified in such a way that continuity of conditional probabilities was preserved.

It follows from Theorem 5 that two \mathfrak{M} -homotopic families $\mathfrak{W}_1, \mathfrak{W}_2 \subset H(\Omega)$ are necessarily \mathfrak{G} -congruent.

Theorem 6. *The unique linear connection of the manifold $H(\Omega)$ that is symmetric with respect to permutations of outcomes and preserves equality of \mathfrak{M} -homotopic tangent vectors is its natural connection.*

Theorem 6 shows the significance, in questions of mathematical statistics, of the natural linear connection on H introduced by us in (5).

Let us return again to arbitrary measurable spaces (Ω, S) . If in (1) one gives up the normalization of the transition measure and does not assume it to be uniformly bounded in ω' , then in (3) one has to consider the cones $\text{Con}(\Omega, S)$ of all nonnegative measures, both finite and infinite, including those not necessarily σ -finite. They form the category CON . Having defined, on finite measures, the projection operation by (4), we obtain a functor that carries objects of the category CON into objects of the category πCAP . Accordingly there arises the category πFAM , with a system of morphisms broader than that of the category FAM :

$$Q_\theta\{\cdot\} = \frac{1}{\lambda(P_\theta)} \int_{\Omega_1} R\{\cdot | \omega'\} P_\theta\{d\omega'\}, \quad \text{where } \lambda(P_\theta) = \int_{\Omega_1} R\{\Omega_2 | \omega'\} P_\theta\{d\omega'\}. \quad (5)$$

This category is closely connected with the theory of stochastic modeling (the theory of the Monte Carlo method). Represent the kernel from (5) in the form $R\{\cdot | \omega'\} = g(\omega')\Pi\{\cdot | \omega'\}$, where $g(\omega') = R\{\Omega_2 | \omega'\}$ is P_θ -almost everywhere

finite, and $\Pi\{\cdot | \omega'\}$ is a probability distribution. Then, for any Q_θ -integrable function $f(\omega'')$,

$$\left[\frac{1}{N} \sum_{i=1}^N g(\omega'_i) f(\omega''_i) \right] \left[\frac{1}{N} \sum_{i=1}^N g(\omega'_i) \right]^{-1} \Rightarrow \int_{\Omega_2} f(\omega'') Q_\theta\{d\omega''\},$$

where $\omega'_i \in \Omega_1$ are independent observations with distribution $P_\theta\{d\omega'\}$, and ω''_i are additional observations simulated according to the random distributions $\Pi\{d\omega'' | \omega'_i\}$.

The parallel between the categories FAM and π FAM can be continued.* However, in contrast to the finite case, in this way one cannot relate a natural linear connection to the object $\text{Cap}(\Omega, S)$. From the very beginning one has to consider narrower cones $C(\Omega, S, I)$ of strictly positive σ -finite measures with a common ideal I of null sets, and correspondingly the narrower objects

$$H(\Omega, S, I) = \text{Cap}(\Omega, S) \cap C(\Omega, S, I).$$

The natural linear connection arising here was described by us in (5).

Received
23 II 1965

REFERENCES

1. A. Grothendieck, *On some questions of homological algebra*, II, 1961.
2. A. Wald, *Sequential Analysis*, Moscow, 1960.
3. R. R. Bahadur, *Ann. Math. Statist.*, **25**, No. 3, 423 (1954).
4. M. G. Krein, M. A. Rutman, *UMN*, **3**, issue 1, 3 (1948).
5. N. N. Chentsov, *DAN*, **158**, No. 3, 543 (1964).
6. I. M. Gelfand, M. I. Graev, *Tr. Mosk. matem. obshch.*, **8**, 321 (1959).
7. F. I. Karpelevich, *Doctoral dissertation*, Moscow, 1963.
8. N. N. Vorob'ev, D. K. Faddeev, *Theory of Probability and Its Applications*, **6**, issue 1, 116 (1961).
9. M. Ya. Antonovskii, V. G. Boltyanskii, T. A. Sarymsakov, *Topological Boolean Algebras*, Tashkent, 1963.

* In particular, one may interpret the notion of equivalence arising here and describe the structure of projectively equivalent families on sets with a finite number of outcomes, and so on.

** The cones $C(\Omega, S, I)$ preserve the following fundamental property of the cones $C(\Omega)$. The ratio of two measures (their Radon-Nikodym derivative) is a positive finite S -measurable function on Ω , defined up to values on a null set. These functions form a positive cone of the corresponding field, cf. (9). With respect to multiplication by such a function, the cone $C(\Omega, S, I)$ is a homogeneous space.

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

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