Докл. Акад. Наук СССР
Том 276 (1984), № 3

Soviet Math. Dokl.
Vol. 29 (1984), No. 3

ALGORITHMIC VARIANTS OF THE NOTION OF ENTROPY

UDC 517.11

A. KH. SHEN′

This note presents a general scheme for obtaining various algorithmic variants of the notion of entropy. The scheme uses the notion of f_0-space in the sense of Ershov [1], and it uses the interpretation of logical operations as operations over problems in the sense of Kolmogorov [2]. Special cases of this scheme turn out to be simple and conditional Kolmogorov entropies [3], [4], decision entropy, monotone and prefix entropies [4]–[6], and also the entropy of computable functions, which is equal to the logarithm of the minimum number for an optimal numbering in the sense of Schnorr ([4], p. 151). Also from the point of view of this scheme we consider the notion of a priori probability [5], [6].

1. The notion of f_0-space. This notion was introduced by Ershov. Let us give a definition convenient for our purposes. The triple $\langle X, X_0, \leqslant \rangle$, where $\langle X, \leqslant \rangle$ is an ordered set and $X_0 \subset X$, is called an *f_0-space* provided: 1) X contains a least element $\bot$, which belongs to X_0; 2) any two elements of X_0 that have a common majorant in X have a least upper bound in X which belongs to X_0; and 3) if $x, y \in X$ and $x \nleqslant y$, then there exists $x_0 \in X_0$ such that $x_0 \leqslant x$ and $x_0 \nleqslant y$. Elements of X will be called *objects* of $\langle X, X_0, \leqslant \rangle$. The object $\bot$ will be called the *indeterminate*, and the elements of X_0 will be called *finite objects* or *f-objects*. Objects x and y having a common majorant will be called *concordant*.

Let us call the set $I \subset X_0$ an *ideal* if it is nonempty, and whenever an f-object z belongs to I, then so does every f-object less than z, and, for any two concordant objects $x, y \in I$, $\sup(x, y)$ is also in I. We call the f_0-space *complete* if each ideal is equal to a set $I_x = \{x_0 \in X_0 | x_0 \leqslant x\}$ for some object x. In the sequel we consider only complete f_0-spaces.

Let us describe a few operations over f_0-spaces. The *product* of two f_0-spaces $\langle X, X_0, \leqslant_1 \rangle$ and $\langle Y, Y_0, \leqslant_2 \rangle$ is the space $\langle X \times Y, X_0 \times Y_0, \leqslant_1 \times \leqslant_2 \rangle$ (the product of the orders is defined componentwise). The *sum* of f_0-spaces $\langle X, X_0, \leqslant_1 \rangle$ and $\langle Y, Y_0, \leqslant_2 \rangle$, where X and Y are disjoint, is defined as $\langle X \cup Y \cup \{\bot\}, X_0 \cup Y_0 \cup \{\bot\}, \leqslant \rangle$, where $\bot$ is an element not appearing in either X or Y, and where $\leqslant$ is such that $\bot \leqslant x$ and $\bot \leqslant y$ for each $x \in X$ and $y \in Y$, the order within X and within Y is preserved, and no element of X is comparable with any element of Y. The space of *continuous functions from $\langle X, X_0, \leqslant \rangle$ to $\langle Y, Y_0, \leqslant \rangle$* consists of the everywhere defined functions from X to Y, continuous with respect to the natural topology of f_0-spaces, in which the base open sets are taken to be the sets consisting of all objects greater than a given f-object. The order on the functions is pointwise: $f \leqslant g \Leftrightarrow (\forall x \in X)(f(x) \leqslant g(x))$. The finite objects in the function space are the functions of the form

$$f_{x_0, y_0}(x) = \text{if } x_0 \leqslant x \text{ then } y_0 \text{ else } \bot$$

1980 *Mathematics Subject Classification*. Primary 94A17; Secondary 03D15.

for all f-objects $x_0 \in X_0$ and $y_0 \in Y_0$, and also the least upper bounds of concordant finite collections of such functions. The operations described above when applied to complete spaces yield complete spaces.

Let $\langle X, X_0, \leqslant \rangle$ be an f_0-space, and let ν be an integer numbering of X_0 such that the sets $\{\langle m, n\rangle | \nu(m) \leqslant \nu(n)\}$ and $\{\langle m, n\rangle | \nu(m)$ concordant with $\upsilon(n)\}$ are decidable, and such that there exists a computable function f: $\mathbf{N}^2 \to \mathbf{N}$ for which $\nu(f(m, n)) = \sup(\nu(m), \nu(n))$ whenever $\nu(m)$ and $\nu(n)$ are concordant. In this case, we shall call the quadruple $\langle X, X_0, \leqslant, \nu\rangle$ an *effective* f_0-space. If X and Y are effective f_0-spaces, then on the product $X \times Y$, on the sum $X + Y$, and on the space of continuous functions $C(X, Y)$, a structure of effective f_0-space may be introduced in a natural way.

Let us give some examples of f_0-spaces that are used in the sequel. We denote by $\mathbf{N}_\perp$ the space whose objects are the natural numbers and the symbol $\perp$. All objects are finite, the object $\perp$ is less than the others, and the natural numbers are not pairwise comparable. We denote by Ω the space whose objects are all finite and infinite sequences of the digits 0 and 1. The f-objects are the finite sequences, and $x \leqslant y$ signifies that x appears at the beginning of y. We denote by Ξ the space of partial functions from $\mathbf{N}$ into $\{0,1\}$. The f-objects are the functions with finite domain, and $x \leqslant y$ signifies that y extends x. Upon replacing $\{0,1\}$ by $\mathbf{N}$ we obtain a space which we denote by F. In each of these spaces the structure of an effective f_0-space is introduced in a natural way. All are complete. In the sequel, complete effective f_0-spaces will simply be called spaces, for brevity.

An object x in the space $\langle X, X_0, \leqslant, \nu\rangle$ is *computable* if the set $\{n | \nu(n) \leqslant x\}$ is enumerable. For any space X there exists a computable object from $C(\mathbf{N}_\perp, X)$, which forms the set of all computable objects of X.

A function l which associates natural numbers to f-objects of a space will be called a *volume* if $n \mapsto l(\nu(n))$ is computable and $l(x_1) \leqslant l(x_2)$ whenever $x_1 \leqslant x_2$. The basic examples of volumes for us are the following: on $\mathbf{N}_\perp$ we define a volume such that $l(\perp) = 0$ and $l(n) =$ (integral part of $\log_2(1 + n)) + 1$, on Ω the volume is to coincide with length, and on Ξ the volume of x is equal to the number of elements in the domain of definition of the function x.

2. Problems and their entropy. Let X be a space, and let A be a set of objects of X. We shall call any pair $\langle X, A\rangle$ a *problem*. X is the *space of the problem*, and objects from A are *solutions to the problem* $\langle X, A\rangle$. We interpret it as the problem of determining, from among objects belonging to X, that object entering the set A. We shall call the problem *monotone* if $x \in A$ and $x \leqslant y$ imply $y \in A$, and *solvable* if in A there exists a computable object. Let X and Y be spaces, and let l be a volume on X. By a *mode of description* of objects of Y with the help of objects of X, we shall mean any computable object of $C(X, Y)$. Let there be given a mode of description $f \in C(X, Y)$ and a problem $\alpha = \langle Y, A\rangle$ in the space Y. The number

$$K_f(\alpha) = K_f(\langle Y, A\rangle) = \inf\{ l(x_0) | x_0 \text{ a finite object in } X, f(x_0) \in A\}$$

is called the *complexity of the problem α with respect to the mode of description f*. We shall say that the mode of description $f \in C(X, Y)$ is *more effective* than the mode of description $g \in C(X, Y)$ if there exists a C such that for any problem $\alpha = \langle Y, A\rangle$ in the space Y the inequality $K_f(\alpha) \leqslant K_g(\alpha) + C$ holds. The mode of description $f \in C(X, Y)$ is called *optimal* if it is more effective than any other mode of description in $C(X, Y)$. Let us call a space X with volume l *regular* if for every space Y there exists an optimal mode of description in $C(X, Y)$.

THEOREM 1. *The space X with volume l is greater if and only if there exists a mode of description $f \in C(X, X \times N_\perp)$ for which*

$$(\forall n \in \mathbf{N})(\exists C)(\forall x_0 \in X_0)\big(K_f(\langle X \times \mathbf{N}_\perp, \{\langle x_0, n\rangle\}\rangle) \leqslant l(x_0) + C\big).$$

From this theorem it follows that the spaces $\mathbf{N}$, Ω, and Ξ are regular.

Let the space X with volume l be regular. For any space Y we choose an optimal mode of description $f \in C(X, Y)$, and we define the *entropy* $K_X(\alpha)$ of α in Y with respect to X to be the complexity of α with respect to f. Thus for a given space Y, the entropy of a problem in this space is determined to within an additive bound.

THEOREM 2. *Let X be a regular space with volume, let Y be an arbitrary space, and let α be a problem in Y. Then the entropy $K_X(\alpha)$ is finite if and only if α is solvable.*

Let (X_1, l_1) and (X_2, l_2) be regular spaces with volume, and let f be a monotone increasing function satisfying a Lipschitz condition.

THEOREM 3. *The following properties are equivalent:*

1) *For any space Y there exists a C such that for any problem $\alpha = \langle Y, A\rangle$*

$$K_{X_1,l_1}(\alpha) \leqslant f\big(K_{X_2,l_2}(\alpha)\big) + C.$$

2) *There exists a C such that for any finite object $x_2 \in X_2$*

$$K_{X_1,l_1}(\langle X_2, \{x_2\}\rangle) \leqslant f(l_2(x_2)) + C.$$

The theorem remains valid if in condition 1) "for any monotone problem" is substituted for "for any problem", and in 2) "$\langle X, \Gamma_{x_2}\rangle$, where $\Gamma_{x_2} = \{x \in X_2 | x \geqslant x_2\}$" is substituted for "$\langle X, \{x_2\}\rangle$". Let us call the conditions so obtained 1′) and 2′). If conditions 1′) and 2′) are satisfied by the function $f(n) = n$, then we shall say that X_1 is *no worse than* X_2; if they are satisfied by $f(n) = n + C\log_2 n$ for some C, then we shall say that X_1 is *almost no worse than* X_2.

THEOREM 4. *The relations $\mathbf{N}_\perp \rightleftarrows \Omega \leftarrow \Xi$ are valid, where $X \to Y$ signifies that X is no worse than Y, and $X \rightleftarrows Y$ signifies that X is almost no worse than Y. No other correlations are valid (with the exception of $\Xi \to \mathbf{N}_\perp$, which follows from the stated relations).*

Let us define logical operations on problems. Let $\alpha = \langle X, A\rangle$ and $\beta = \langle Y, B\rangle$ be two problems. We define $\alpha \wedge \beta = \langle X \times Y, A \times B\rangle$, $\alpha \vee \beta = \langle X + Y, A \cup B\rangle$ (X and Y are assumed disjoint), and $\alpha \supset \beta = \langle C(X, Y), \{f | f(A) \subset B\}\rangle$. We shall call the problem $F = \langle P, \varnothing\rangle$, where P contains a single finite object, *false*.

The entropy $K_X(\alpha \supset \beta)$ of the problem $\alpha \supset \beta$ will be called the *conditional entropy of β with respect to the known α*. We designate it $K(\beta|\alpha)$.

Let $\Phi(p_1,\ldots,p_n)$ be a propositional formula containing the signs $\wedge$, $\vee$, $\supset$, and F (falsity). If in place of $p_1,\ldots,p_n$ we substitute the problems $\alpha_1 = \langle X_1, A_1\rangle,\ldots,$ $\alpha_n = \langle X_n, A_n\rangle$, then the problem $\Phi(\alpha_1,\ldots,\alpha_n)$ will arise. The space of this problem is determined by the spaces $X_1,\ldots,X_n$ and does not depend on the A_i; let us designate this space as $\Phi(X_1,\ldots,X_n)$.

THEOREM 5. *Let $\Phi(p_1,\ldots,p_n)$ be deducible in the intuitionistic propositional calculus, and let $X_1,\ldots,X_n$ be spaces.*

Then there exists a computable object in the space $\Phi(X_1,\ldots,X_n)$ which is the solution of the problem $\Phi(\langle X_1, A_1\rangle,\ldots,\langle X_n, A_n\rangle)$ for any $A_i \subset X_i$.

THEOREM 6. *Let $\Phi(p_1,\dots,p_n) \supset \Psi(p_1,\dots,p_n)$ be a formula deducible in the intuitionistic propositional calculus, let $X_1,\dots,X_n$ be spaces, and let X be a regular space with volume.*

Then there exists a C such that for any of the problems $\alpha_1,\dots,\alpha_n$ in spaces $X_1,\dots,X_n$ the inequality $K_X(\Psi(\alpha_1,\dots,\alpha_n)) \leqslant K_X(\Phi(\alpha_1,\dots,\alpha_n)) + C$ is valid.

This theorem implies the inequalities $K_X(\alpha) \leqslant K_X(\alpha \wedge \beta) + O(1)$, $K_X(\alpha|\beta) \leqslant K_X(\alpha) + O(1)$, $K_X(\beta) \leqslant K_X(\alpha \wedge (\alpha \supset \beta)) + O(1)$ and many others.

Let us consider the set Q of all formulas which satisfy the statement of Theorem 5. Let Q be a superintuitionistic logic.

THEOREM 7. *The logic Q does not coincide with either the intuitionistic logic nor the classical logic. It also differs from Medvedev's logic of finitary problems* [7].

THEOREM 8. a) $K_{\mathbf{N}_\perp}(\langle \mathbf{N}_\perp, \{n\}\rangle) =$ (*complexity of n in the sense of* [3]) $+ O(1)$.

b) $K_{\mathbf{N}}(\langle \Omega, \Gamma_x\rangle) =$ (*complexity of the solution of the sequence x in the sense of* [5]) $+ O(1)$.

c) $K_\Omega(\langle N_\perp, \{n\}\rangle) =$ (*prefix entropy of n in the sense of* [6]) $+ O(1)$.

d) $K_\Omega(\langle \Omega, \Gamma_x\rangle) =$ (*monotone entropy of the sequence x in the sense of* [6]) $+ O(1)$.

e) $K(\langle \mathbf{N}_\perp, \{n\}\rangle \supset \langle \mathbf{N}_\perp, \{m\}\rangle) =$ (*conditional complexity of m relative to n in the sense of* [3]) $+ O(1)$.

f) $K_{\mathbf{N}}(\langle F, \{f\}\rangle) =$ (*logarithm of the number of the computable function f for an optimal numbering, in the sense of* [4], *p.* 151) $+ O(1)$.

We recall that Γ_x designates the set $\{y|x \leqslant y\}$.

Let X be an arbitrary space, and $f \in C(\Omega, X)$ a mode of description. With each problem $\alpha = \langle X, A\rangle$, where A is a Borel subset of X (with respect to the topology), we shall compare the number $P_f(\alpha) =$ measure(ω-infinite sequence of digits 0 and $1|f(\omega) \in A$), which is called the *decision probability* of the problem α under the mode of description f. Among all the modes of description there exists one which is optimal, for which $P_f(\alpha)$ is maximal to within a multiplicative constant: for every other method g, there may be found a $C > 0$ such that $P_f(\alpha) \ngeqslant CP_g(\alpha)$ for all problems α in X. Having selected and fixed an optimal mode f, let us call $P_f(\alpha)$ the *a priori probability* of the problem α and denote it by $P(\alpha)$. With $X = \mathbf{N}_\perp$ the a priori probability of the problem $\langle \mathbf{N}_\perp, \{n\}\rangle$ coincides with that introduced in [6], p. 26 (to within a bounded factor, isolated from zero). With $X = \Omega$ the a priori probability of the problem $\langle \Omega, \Gamma_x\rangle$ coincides with that introduced in [5], p. 49 (semi-measure M in Theorem 4.1).

Let P be a measure defined on the Borel subsets of X. Let us call the measure *enumerable* if the set $\{\langle n, r\rangle \in \mathbf{N} \times \mathbf{Q}|r < P(\Gamma_{\nu(n)})\}$ is enumerable. (Here ν is the numbering appearing in the definition of effective space.) An a priori probability is an enumerable measure.

THEOREM 9. *If every pair of finite concordant objects of the space X satisfies $x \leqslant y$ or $y \leqslant x$, then the a priori probability on X is a maximal* (*to within a multiplicative constant*) *enumerable measure. The condition imposed on the space X is essential: in the space Ξ the a priori probability is not a maximal enumerable measure.*

THEOREM 10. a) $-\log_2 P(\alpha) \leqslant K_\Omega(\alpha) + O(1)$, *$O(1)$ depends only upon the space of α.*

b) *The inverse inequality $K_\Omega(\alpha) \leqslant -\log_2 P(\alpha) + O(1)$ fails for problems of the form* $\langle \Xi, \Gamma_x\rangle$.

c) *There exists a regular space M with volume l for which* $K_{M,l}(\alpha) = -\log_2 P(\alpha) + O(1)$ *for all problems of type* $\langle X, \Gamma_x \rangle$, *where X is an arbitrary space (upon which the bound for* $O(1)$ *depends), and x is any finite object in X.*

d) *There does not exist a regular space for which the relation in* c) *holds for all monotone problems in every space X.*

Institute for Problems of Information Transmission
Academy of Sciences of the USSR
Moscow

Received 1/JULY/83

BIBLIOGRAPHY

1. Yu. L. Ershov, Algebra i Logika **11** (1972), 367–437; English transl. in Algebra and Logic **11** (1972).

2. A. Kolmogoroff [A. N. Kolmogorov], Math. Z. **35** (1932), 58–65.

3. ______, Problemy Peredachi Informatsii **1** (1965), no. 1, 3–11; English transl. in Selected Transl. Math. Statist. and Probab., vol. 7, Amer. Math. Soc., Providence, R.I., 1968.

4. V. A. Uspensky [Uspenskiĭ] and A. L. Semenov, Algorithms in Modern Mathematics and Computer Science (Urgench, 1979), Lecture Notes in Computer Sci., vol. 122, Springer-Verlag, 1981, pp. 100–234.

5. A. K. Zvonkin and L. A. Levin, Uspekhi Mat. Nauk **25** (1970), no. 6(156), 85–127; English transl. in Russian Math. Surveys **25** (1970).

6. V. V. V'yugin, Semiotika i Informatika Vyp. 16, VINITI, Moscow, 1981, pp. 14–43. (Russian)

7. Yu. T. Madvedev, Dokl. Akad. Nauk SSSR **142** (1962), 1015–1018; English transl. in Soviet Math. Dokl. **3** (1962), No. 1.

Translated by W. MARGOLIS