Докл. Акад. Наук СССР
Том 271 (1983), № 6

Soviet Math. Dokl.
Vol. 28 (1983), No. 1

THE CONCEPT OF (α, β)-STOCHASTICITY IN THE KOLMOGOROV SENSE, AND ITS PROPERTIES

UDC 517.11

A. KH. SHEN′

We investigate properties of Kolmogorov's concept of a finite stochastic object. Informally, stochastic objects are "elements in general position" in simple sets. Precise definitions will be given.

We regard the natural numbers as *finite objects*. In speaking of the *entropy* (= complexity) of a number x we shall have in mind its simple Kolmogorov entropy, introduced in [1]. The entropy of a number x is denoted by $K(x)$. We also need to speak of the entropy of finite sets of natural numbers. For this we fix some natrual enumeration of the finite sets (for example, that described in [2]) and understand the *entropy of a set* to be the entropy of its index. The entropy of a set A is denoted by $K(A)$.

DEFINITION (KOLMOGOROV). Let α and β be natural numbers. A number x will be called (α, β)-*stochastic* if there exists a finite set $A \subset \mathbf{N}$ such that

$$x \in A, \qquad K(A) \leqslant \alpha, \qquad K(x) \geqslant \log_2|A| - \beta;$$

here $|A|$ denotes the number of elements in the set A.

The first inequality (if α is not too large) means that A is sufficiently simple. The second (if β is not too large) means that the element x is an "element in general position" in A. Indeed, if x had some features peculiar to only a very small part Q of A, then these could be used for a simple description of x by determining its ordinal number in a list of all the elements in Q, which would require $\log_2|Q|$ bits, i.e., many fewer than $\log_2|A|$.

We establish a connection between the concept of (α, β)-stochasticity and the foundations of mathematical statistics. Suppose that we carry out some probabilistic experiment whose result can be a priori any natural number. Suppose that the result of this experiment turns out to be the number x. Knowing x, we want to recover the probability distribution P on the set $\mathbf{N}$ of all natural numbers. It is reasonable to require that, first, P has a simple description and, second, x would be a "typical" outcome of an experiment with the probability distribution P. (In practice the specific nature of the problem frequently suggests beforehand a possible form of P, and it remains to select some of its parameters; however, we assume that our only existing information about P is the value of x obtained.)

Let us make this precise. As *probability distributions* we consider functions $P\colon \mathbf{N} \to \mathbf{Q}$ defined everywhere and with nonnegative values such that $P(x) = 0$ except for finitely many numbers x and $\Sigma_x P(x) \leqslant 1$. (We allow for the possibility that $\Sigma P(x) < 1$, considering that our experiment may not give a result.) To speak of the entropy of such functions we fix some natural enumeration of them by the natural numbers, and in speaking of the *entropy of a function* we shall have in mind the (simple Kolmogorov) entropy of its index. The entropy of a distribution P is denoted by $K(P)$. The requirement

1980 *Mathematics Subject Classification*. Primary 60H99; Secondary 28D20.

that the distribution P be simple now becomes the requirement that its entropy be small. The requirement that x be "typical" for the distribution P is made precise as follows:

$$K(x) \text{ must not be much less than } -\log_2 P(x).$$

For example, if P assigns probability $1/2^n$ to all the numbers from 0 to $2^n - 1$, then those $x \in \{0,\ldots, 2^n - 1\}$ for which the entropy $K(x)$ is close to n are "typical".

Note that $K(x)$ cannot greatly exceed $-\log_2 P(x)$ if P is a sufficiently simple distribution. Namely, for any x

$$K(x) \leqslant -\log_2 P(x) + K(P) + O(\log_2(-\log_2 P(x) + K(P))).$$

Indeed, suppose that $1/2^{k+1} \leqslant P(x) \leqslant 1/2^k$. Consider the set of all t such that $P(t) \geqslant 1/2^{k+1}$. There are no more than 2^{k+1} elements in it, and x is one of them. To specify x if suffices to determine this set and the ordinal number of the element x in it. To determine the set it suffices to determine P and the number k; determination of the ordinal number requires no more than $k + 1$ bits. This implies the inequality we have just written. The requirement that x be "typical" quarantees that this inequality is close to an equality (if the entropy of P is not too large).

The next definition singles out those x for which it is possible to find a distribution P with the properties described.

DEFINITION. Let α and β be natural numbers. A number is said to be (α, β)-*quasistochastic* if there exists a distribution P (in the class described) such that

$$K(P) \leqslant \alpha, \qquad K(x) \geqslant -\log_2 P(x) - \beta.$$

The concepts of stochasticity and quasistochasticity turn out to be very close. Namely,

THEOREM 1. *There exist constants C_1 and C_2 such that for any number x:*
a) *if x is (α, β)-stochastic, then x is $(\alpha + C_1, \beta)$-quasistochastic; and*
b) *if x is (α, β)-quasistochastic and $x \in \{0,\ldots, 2^n - 1\}$, then x is $(\alpha + C_1 \log_2 n, \beta + C_2)$-stochastic.*

This theorem shows that stochasticity and quasistochasticity "coincide to within $\log_2 n$."

PROOF. Assertion a) is easily proved. Suppose that $x \in A$, (entropy of A) $\leqslant \alpha$, and $K(x) \geqslant \log_2|A| - \beta$. Consider the distribution P assigning the same probability $1/|A|$ to all the elements of A, and zero probability to all the remaining numbers. Obviously, the entropy of P exceeds the entropy of A by not more than a constant, and $\log_2|A| = -\log_2 P(x)$. What is required follows from this.

The proof of b) is scarcely more complicated. Suppose that the number x is (α, β)-quasistochastic. Then there exists a distribution P whose entropy is at most α, and $K(x) \geqslant -\log_2 P(x) - \beta$. Let k be the number such that $2^{-(k+1)} \leqslant P(x) < 2^{-k}$. Then $K(x) \geqslant k - \beta$. From this inequality and the inequality $K(x) \leqslant n + O(1)$ it follows that $k \leqslant n + \beta + O(1)$ (this estimate is needed in what follows). Consider now the set A consisting of all the $y \in \mathbf{N}$ such that $P(y) \geqslant 2^{-(k+1)}$. To specify A it suffices to determine P and k; therefore, the entropy of A does not exceed $\alpha + C\log_2(n + \beta)$. The set A contains at most 2^{k+1} elements, and so $K(x) \geqslant \log_2|A| - (\beta + 1)$. Thus, x is $(\alpha + C\log_2(n + \beta), \beta + 1)$-stochastic. If $\beta \leqslant n$, then b) is proved. But if $\beta > n$, then any number from 0 to $2^n - 1$ is $(C \cdot \log_2 n, \beta)$-stochastic (it suffices to take A to be the set of all numbers from 0 to $2^n - 1$). Theorem 1 is proved.

We now turn to the question of values of α and β such that there exist numbers between 0 and $2^n - 1$ that are not (α, β)-stochastic. The answer is given by

THEOREM 2. a) *There exists a constant C such that, for any n and any α and β with $\alpha \geqslant \log_2 n + C$ and $\alpha + \beta \geqslant n + 4\log_2 n + C$, all the numbers from 0 to $2^n - 1$ are (α, β)-stochastic.*

b) *There exists a constant C such that, for any n and any α and β with $2\alpha + \beta < n - 6\log_2 n - C$, not all the numbers from 0 to $2^n - 1$ are (α, β)-stochastic.*

PROOF. a) Suppose first that $\beta \leqslant n$. We partition the numbers from 0 to $2^n - 1$ into $2^{n-\beta}$ sets with 2^β elements in each (for example, by putting the numbers from $2^\beta i$ to $2^\beta(i + 1)$ into the ith set). To specify any of these sets it is necessary to specify n, β, and the number from 0 to $2^{n-\beta}$ which indicates its order in our partition. Therefore, the entropy of any of the sets of the partition does not exceed $2\log_2 n + 2\log_2 \beta + (n - \beta) + C$, i.e., is $\leqslant n - \beta + 4\log_2 n + C$ (here C is a constant not depending on n, α, nor β). If $\alpha + \beta \geqslant n + 4\log_2 n + C$, then the entropy of any of the sets in the partition does not exceed α; therefore, all the numbers from 0 to $2^n - 1$ are (α, β)-stochastic. (The second inequality in the definition of (α, β)-stochasticity holds because $\log_2 2^\beta - \beta = 0$ stands on its right-hand side.) But if $\beta \geqslant n$, then with the set $\{0, 1, \ldots, 2^n - 1\}$ as A we see that its entropy is at most $\log_2 n + C$ (and, consequently, is at most α), and all its elements are (α, β)-stochastic.

b) Let α be fixed. Consider the list $A_1, \ldots, A_s$ of all the finite sets whose entropy does not exceed alpha. Obviously, $s \leqslant 2^{\alpha+1}$. We want to estimate the entropy of the family $A_1, \ldots, A_s$. To specify this family it suffices to determine (besides α) that one of the descriptions of the sets $A_1, \ldots, A_s$ which requires the greatest number of steps in its processing by the chosen method of description. Therefore, the entropy of this family does not exceed $\alpha + 2\log_2 \alpha + C_1$, where C_1 is a constant not depending on α. Consider those of the sets $A_1, \ldots, A_s$ with fewer than $2^{n-\alpha-1}$ elements. We take the smallest number x not contained in their union. This number is less than 2^n, since s does not exceed $2^{\alpha+1}$, while each of the sets $A_1, \ldots, A_s$ has fewer than $2^{n-\alpha-1}$ elements.

To specify x we must determine $A_1, \ldots, A_s$, α, and n. Therefore, its entropy does not exceed

$$\alpha + 2\log_2 \alpha + 2\log_2 \alpha + 2\log_2 n + C'$$

and, all the more so, $\alpha + 6\log_2 n + C'$ (here C' is a constant not depending on n nor α). We prove that if

$$\beta < n - 6\log_2 n - (C' + 1) - 2\alpha,$$

then the x we constructed is not (α, β)-stochastic. This will imply the assertion of the theorem with $C = C' + 1$. Indeed, if x is (α, β)-stochastic, then $x \in A_i$ and $K(x) \geqslant \log_2|A_i| - \beta$ for some i. The set A_i must contain no fewer than $2^{n-\alpha-1}$ numbers (otherwise x would not belong to it); therefore, $K(x) \geqslant n - \alpha - 1 - \beta$. But $K(x) \leqslant \alpha + 6\log_2 n + C'$, whence

$$\alpha + 6\log_2 n + C' \geqslant n - \alpha - 1 - \beta \quad \text{and} \quad \beta \geqslant n - 6\log_2 n - 2\alpha - 1 - C'.$$

Theorem 2 is proved.

Theorem 2 indicates a boundary for α/n and β/n such that the last nonstochastic objects disappear when it is crossed. This boundary (for the case $\alpha = \beta$) is somewhere between $1/2$ and $1/3$.

The next theorem answers the question about the fraction of (α, β)-stochastic numbers among all the numbers from 0 to $2^n - 1$.

THEOREM 3. *There exists a constant C such that, for all n and all α and β such that $\alpha \geqslant C \cdot \log_2 n$, the cardinality of the set of numbers from 0 to $2^n - 1$ that are not (α, β)-stochastic is between the numbers*

$$[2^{n-2\alpha-\beta-C\log_2 n}] \quad \textit{and} \quad 2^{n-\alpha-\beta+C\log_2 n},$$

$[a]$ is the integer part of a number a.

PROOF. Let us first get an upper estimate for the cardinality of the set of nonstochastic numbers. As in the proof of Theorem 2, we partition the set of numbers from 0 to $2^n - 1$ into 2^p parts of 2^{n-p} numbers each. The entropy of each part is at most $p + O(\log_2 n)$, and so by choosing $p = \alpha - C\log_2 n$ with a suitable C we can ensure that the entropy of any part does not exceed α. Here all the numbers whose entropy is greater than $n - p - \beta$ are (α, β)-stochastic. Therefore, the cardinality of the set of nonstochastic numbers does not exceed

$$2^{n-p-\beta} = 2^{n-\alpha-\beta+C\log_2 n}.$$

The upper estimate has been obtained.

To get a lower estimate we consider all the sets whose entropy does not exceed α, while the number of elements does not exceed $2^{n-\alpha-2}$. The entropy of the list of all such sets is at most $\alpha + O(\log_2 n)$. The union of all the sets in this list contains no more than half of all the numbers from 0 to $2^n - 1$. Let a_i denote the ith (in increasing order) number not in this union (with $i < 2^{n-1}$). By the foregoing, $a_i < 2^n$ for any $i < 2^{n-1}$. The entropy of a_i is at most

$$\alpha + O(\log_2 n) + O(\log_2 \alpha) + \log_2 i;$$

only those among the a_i whose entropy exceeds $n - 2 - \alpha - \beta$ can be (α, β)-stochastic, i.e.,

$$\alpha + O(\log_2 n) + \log_2 i \geqslant n - \alpha - 2 - \beta \quad \text{and} \quad \log_2 i \geqslant n - 2\alpha - \beta - O(\log_2 n).$$

Therefore, there are at least $[2^{n-2\alpha-\beta-O(\log_2 n)}]$ nonstochastic numbers. This proves Theorem 3.

It shows that (to within $\log_2 n$) the fraction of (α, β)-stochastic numbers among the numbers from 0 to $2^n - 1$ is included between $1 - 1/2^{\alpha+\beta}$ and $1 - 1/2^{2\alpha+\beta}$.

From the point of view of our statistical interpretation it is of interest to know what the probability is of nonstochastic numbers appearing in a probabilistic experiment. More precisely, suppose that P is a probability distribution on the set of numbers from 0 to $2^n - 1$. What can be said about $P(Q)$, where Q is the set of all nonstochastic numbers (for given α and β)? It is natural to want $P(Q)$ to be small. If nothing is required of P, then this cannot be achieved: for example, P can assign probability 1 to some nonstochastic number. However, if P has small entropy, then we can obtain the desired estimate.

THEOREM 4. *There exists a constant C such that, for any probability distribution P whose entropy does not exceed α, the quantity $P(Q)$, where Q is the set of all numbers from 0 to $2^n - 1$ that are not $(\alpha + C\log_2 n, \beta)$-stochastic, is at most $2^{-\beta+C\log_2 n}$.*

PROOF. Using Theorem 1, we can prove the assertion with $(\alpha + C \cdot \log_2 n, \beta)$-stochasticity replaced by (α, β)-quasistochasticity. This is done as follows. For all x that are not (α, β)-quasistochastic we have that $K(x) < -\log_2 P(x) - \beta$ or $P(x) < 2^{-K(x)-\beta}$.

Hence

$$\sum_{x \text{ not } (\alpha,\beta)\text{-quasistochastic}} P(x) < 2^{-\beta} \sum_{x \in \{0,\dots,2^n-1\}} 2^{-K(x)};$$

the first part is at most $n + O(1)$, because the number of those x such that $K(x) = a$ is at most 2^a. The statement of Theorem 4 is obtained from this.

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

Received 26/DEC/82

BIBLIOGRAPHY

1. A. N. Kolmogorov, 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.

2. Hartley Rogers, Jr., *Theory of recursive functions and effective computability*, McGraw-Hill, 1967.

Translated by H. H. McFADEN