V. G. VOVK

ON A CRITERION OF RANDOMNESS

(Communicated by Academician A. N. Kolmogorov, 25.01.1986)

This work belongs to algorithmic information theory (see [1]). Theorem 1 shows that if some sequence is random with respect to two computable measures P and Q simultaneously, then these measures asymptotically agree in their forecasts of the behaviour of this sequence. It turns out that in these terms one can give a criterion of randomness with respect to a computable measure Q of a sequence random with respect to a computable measure P (Theorem 3).

1. In this section we will give the main definitions and state some known results of algorithmic information theory in a convenient for us form (a more complete exposition can be found in the survey [2]).

Let X be an arbitrary ensemble in the sense of [3] (in a different terminology, a space of constructive objects), fixed until the end of this note[1]. Denote by X^∞ the set of all infinite sequences $\omega = \omega_0\omega_1\omega_2\ldots$ of elements of the ensemble X, and by X^* the set of all finite sequences $a = a_0a_1\ldots a_{n-1}$ of elements of the ensemble X; $\Lambda \in X^*$ is the empty sequence. For $\omega \in X^\infty$, we denote by ω^n the sequence $\omega_0\omega_1\ldots\omega_{n-1}$ from X^*.

A function $P : X^* \to [0, +\infty[$ is called a *semimeasure* if $P(\Lambda) = 1$ and $P(a) \geq \sum_{x\in X} P(ax)$ for all $a \in X^*$ (by ax we denote the sequence obtained from a by adding another term x). The semimeasure P is called a *measure* if $P(a) = \sum_{x\in X} P(ax)$ for all $a \in X^*$.

We will consider recursively enumerable (r.e.) semimeasures, i.e., semimeasures for which the set $\{(r, a) \mid r \in \mathbb{Q}, a \in X^*, r < P(a)\}$ is r.e. ($\mathbb{Q}$ is the set of all rational numbers). A *computable semimeasure* is an r.e. semimeasure such that the set $\{(r, a) \mid r \in \mathbb{Q}, a \in X^*, r > \sum_{x\in X} P(ax)\}$ is also r.e. and the set $\{a \in X^* \mid P(a) = 0\}$ is decidable. If P is a computable semimeasure, there is an algorithm that computes $P(a)$ and $\sum_{x\in X} P(ax)$ for any $a \in X^*$ with any degree of accuracy.

There exists an r.e. semimeasure M such that, for any r.e. semimeasure P, $P(a) = M(a) \cdot O(1)$, $a \in X^*$. Let us fix one of such semimeasures M—let us call it the a priori *semimeasure*—and fix the notation M for it.

A sequence $\omega \in X^\infty$ is called *random in the sense of Martin-Löf* (in what follows the words "in the sense of Martin-Löf" will be omitted) with respect to an r.e. semimeasure P if $M(\omega^n) = P(\omega^n) \cdot O(1)$. In the case when P is a computable measure, this definition is equivalent to the original, very natural, definition given by Martin-Löf. Martin-Löf's definition requires that ω should satisfy all "efficient" (in a certain exact sense) laws of probability theory.

If P is an r.e. semimeasure and $a \in X^*$, we call the value $\ln(M(a)/P(a))$ the *randomness deficiency* of the sequence a with respect to P and denote it $d(a|P)$. If $\omega \in X^\infty$, the rate of growth of $d(\omega^n|P)$ as $n \to \infty$ reflects the "degree

[1] Examples of ensembles: the set of all natural numbers, the set of all finite binary sequences. One can also take the set $\{0, 1\}$ as X.

of non-randomness" of ω with respect to P. The randomness of $\omega \in X^\infty$ with respect to P is equivalent to $d(\omega^n \mid P) = O(1)$.

2. A *probability distribution* is a function $p : X \to [0, +\infty[$ such that $\sum_{x \in X} p(x) \leq 1$. If p and q are probability distributions, the *Hellinger distance* $\rho(q, p)$ between p and q is defined as $\sum_{x \in X} \left(\sqrt{q(x)} - \sqrt{p(x)} \right)^2$, and the χ^2*-distance*, which we will denote $\rho_2(q, p)$, as[2] $\sum_{x \in X} \left(q(x) - p(x) \right)^2 / q(x)$ (see [4, p. 194]). A probability distribution p is *proper* if $\sum_{x \in X} p(x) = 1$.

The ratio $P(ax)/P(a)$, where $a \in X^*$ and $x \in X$, will be denoted $P(x \mid a)$. Let P be a semimeasure and $\omega \in X^\infty$. By P_n^ω we will denote the probability distribution such that $P_n^\omega(x) = P(x \mid \omega^n)$ for all $x \in X$. If P is a measure and $P(\omega^n) \neq 0$, the probability distribution P_n^ω is proper.

Theorem 1. *Let P and Q be computable semimeasures such that Q is a measure, and $\omega \in X^\infty$. Then*

$$\sum_{i=0}^{n-1} \rho(Q_i^\omega, P_i^\omega) - d(\omega^n \mid P) - O(1) \leq d(\omega^n \mid Q) \leq \sum_{i=0}^{n-1} \rho_2(Q_i^\omega, P_i^\omega) + 2d(\omega^n \mid P) + O(1).$$

Proof. (a) **Lower bound.** Define an r.e. semimeasure R by

$$R(x \mid a) = \frac{\sqrt{P(x \mid a) Q(x \mid a)}}{\sum_{y \in X} \sqrt{P(y \mid a) Q(y \mid a)}},$$

where $x \in X$ and $a \in X^*$. Being an r.e. semimeasure, R satisfies

$$R(\omega^n) = e^{d(\omega^n \mid P)} \cdot P(\omega^n) \cdot O(1).$$

Without loss of generality we suppose $P(\omega^n) \neq 0$ and $Q(\omega^n) \neq 0$, $\forall n$. Writing $P(\omega^n)$ as $\prod_{i=0}^{n-1} P(\omega_i \mid \omega^i)$, $R(\omega^n)$ as $\prod_{i=0}^{n-1} R(\omega_i \mid \omega^i)$, and $R(\omega_i \mid \omega^i)$ according to its definition, after cancellation we obtain

$$\prod_{i=0}^{n-1} \frac{\sqrt{Q(\omega_i \mid \omega^i) / P(\omega_i \mid \omega^i)}}{\sum_{y \in X} \sqrt{P(y \mid \omega^i) Q(y \mid \omega^i)}} = e^{d(\omega^n \mid P)} \cdot O(1).$$

Noticing that $\prod_{i=0}^{n-1} \sqrt{Q(\omega_i \mid \omega^i) / P(\omega_i \mid \omega^i)} = e^{(d(\omega^n \mid P) - d(\omega^n \mid Q))/2}$ and taking logarithms of both sides, it is easy to obtain

$$d(\omega^n \mid P) + d(\omega^n \mid Q) \geq -2 \sum_{i=0}^{n-1} \ln \left(\sum_{y \in X} \sqrt{P(y \mid \omega^i) Q(y \mid \omega^i)} \right) - O(1).$$

[2] We set $\frac{0}{0} := 0$ (and also $\frac{\infty}{\infty} := 0$).

The statement we are proving now follows from

$$2\ln\left(\sum_{y\in X}\sqrt{P(y\mid\omega^i)Q(y\mid\omega^i)}\right)\le 2\ln\left(1-\frac{1}{2}\sum_{y\in X}\left(\sqrt{P(y\mid\omega^i)}\right.\right.$$
$$\left.\left.-\sqrt{Q(y\mid\omega^i)}\right)^2\right)\le-\sum_{y\in X}\left(\sqrt{P(y\mid\omega^i)}-\sqrt{Q(y\mid\omega^i)}\right)^2.$$

(b) **Upper bound.** Let R be a semimeasure such that

$$R(a)\ne 0\quad\Longrightarrow\quad R(x\mid a)=\frac{P^2(x\mid a)}{Q(x\mid a)}\Big/\sum_{y\in X}\frac{P^2(y\mid a)}{Q(y\mid a)}$$

for all $x\in X$, $a\in X^*$. Analogously to part (a) we obtain

$$d(\omega^n\mid Q)-2d(\omega^n\mid P)\le\sum_{i=0}^{n-1}\ln\sum_{y\in X}\frac{P^2(y\mid\omega^i)}{Q(y\mid\omega^i)}+O(1).$$

After this it suffices to notice that

$$\ln\sum_{y\in X}\frac{P^2(y\mid\omega^i)}{Q(y\mid\omega^i)}\le\ln\left(1+\sum_{y\in X}\frac{(P(y\mid\omega^i)-Q(y\mid\omega^i))^2}{Q(y\mid\omega^i)}\right)$$
$$\le\sum_{y\in X}\frac{(P(y\mid\omega^i)-Q(y\mid\omega^i))^2}{Q(y\mid\omega^i)}.$$

□

It is easy to see that the assumption that Q is a measure was used only in the proof of the upper bound.

Theorem 1 shows that if a sequence ω is random with respect to a computable measure P and a computable measure Q is chosen so that $d(\omega^n\mid Q)=o(n)$, then the "mean Hellinger distance" $\frac{1}{n}\sum_{i=0}^{n-1}\rho(Q_i^\omega,P_i^\omega)\to 0$.

Theorem 2. *Let P and Q be computable semimeasures, and let $\omega\in X^\infty$ be random with respect to both P and Q. Then*

$$\sum_{i=0}^{n-1}\left(\frac{P(\omega_i\mid\omega^i)}{Q(\omega_i\mid\omega^i)}-1\right)^2<\infty.$$

Proof. Define a computable semimeasure R by the formula

$$R(x\mid a)=\frac{P(x\mid a)+Q(x\mid a)}{2}$$

for all $x\in X$, $a\in X^*$. The condition of the theorem immediately implies

$$\prod_{i=0}^{n-1}R(\omega_i\mid\omega^i)=\prod_{i=0}^{n-1}P(\omega_i\mid\omega^i)\cdot O(1),$$
$$\prod_{i=0}^{n-1}R(\omega_i\mid\omega^i)=\prod_{i=0}^{n-1}Q(\omega_i\mid\omega^i)\cdot O(1).$$

Writing out $R(\omega_i \mid \omega^i)$ according to its definition, we can obtain

$$\prod_{i=0}^{n-1} \frac{1 + Q(\omega_i \mid \omega^i)/P(\omega_i \mid \omega^i)}{2} = O(1),$$
$$\prod_{i=0}^{n-1} \frac{1 + P(\omega_i \mid \omega^i)/Q(\omega_i \mid \omega^i)}{2} = O(1).$$

Multiplying, we obtain

$$\prod_{i=0}^{n-1} \frac{2 + P(\omega_i \mid \omega^i)/Q(\omega_i \mid \omega^i) + Q(\omega_i \mid \omega^i)/P(\omega_i \mid \omega^i)}{4} = O(1).$$

Notice that each term of this product is ≥ 1. This immediately implies $P(\omega_i \mid \omega^i)/Q(\omega_i \mid \omega^i) \to 1$ as $i \to \infty$. The natural logarithm of the typical term of the product is, asymptotically,

$$\left(\frac{P(\omega_i \mid \omega^i)}{Q(\omega_i \mid \omega^i)} - 1\right)^2 \Big/ 4,$$

which immediately implies the conclusion of the theorem. □

3. From Theorems 1 and 2 we will deduce the following criterion of randomness.

Theorem 3. *Let P and Q be computable semimeasures such that Q is a measure, $\omega \in X^\infty$ be random with respect to P, and $Q(\omega^n) \neq 0$, $\forall n$. Then*

$$\omega \text{ is random with respect to } Q \iff \sum_{i=0}^{\infty} \rho(P_i^\omega, Q_i^\omega) < \infty.$$

Proof. Theorem 1 implies that

$$\sum_{i=0}^{n-1} \rho(P_i^\omega, Q_i^\omega) - O(1) \leq d(\omega^n \mid Q) \leq \sum_{i=0}^{n-1} \rho_2(Q_i^\omega, P_i^\omega) + O(1).$$

The implication "$\Longrightarrow$" is obvious. Let us prove "$\Longleftarrow$". Choose some computable family $\{A(a) \mid a \in X^*\}$ of subsets of the set X such that $P(x \mid a)/Q(x \mid a) \geq 2$ for $x \in A(a)$ and $P(x \mid a)/Q(x \mid a) \leq 3$ for $x \notin A(a)$. Define a computable measure $\overline{Q}$ by the requirement that

$$\overline{Q}(x \mid a) = \begin{cases} P(x \mid a)/2 & \text{if } x \in A(a), \\ Q(x \mid a) \cdot C(a) & \text{if } x \notin A(a), \end{cases}$$

for all $x \in X$ and $a \in X^*$ such that $Q(a) \neq 0$, where the function $C : X^* \to]0, 1]$ is chosen so that $\overline{Q}$ can indeed be a measure. Using the function C we define a

semimeasure $\overline{P}$ by the equality $\overline{P}(x|a) = P(x|a)\cdot C(a), \forall x \in X, a \in X^*$. We will consecutively prove that ω is random with respect to $\overline{P}$, with respect to $\overline{Q}$, and, finally, with respect to Q. The proof will use the fact that $\rho(p,q)$, where p and q are probability distributions, is within a constant factor of $\sum_{x\in X} \frac{(p(x)-q(x))^2}{p(x)\vee q(x)}$ (the symbol $\vee$ denotes the maximum of two numbers).

In order to prove the randomness of ω with respect to $\overline{P}$ it suffices to prove that $\prod_{i=0}^{\infty} C(\omega^i) > 0$. Let us use the convergence of the series

$$\sum_{i=0}^{\infty} \sum_{y\in X} \frac{(P(y \mid \omega^i) - Q(y \mid \omega^i))^2}{P(y \mid \omega^i) \vee Q(y \mid \omega^i)}.$$

Of course, the convergence will not be affected if $\sum_{y\in X}$ is replaced by $\sum_{y\in A(\omega^i)}$. Therefore, $\sum_{i=0}^{\infty} P(A(\omega^i) \mid \omega^i) < \infty$ (here we have used the notation $R(B \mid a) = \sum_{y\in B} R(y \mid a)$, where R is a semimeasure, $B \subseteq X$, $a \in X^*$). It remains to notice that $C(\omega^i) \geq 1 - P(A(\omega^i) \mid \omega^i)/2$, and so $-\ln C(\omega^i) = O(P(A(\omega^i) \mid \omega^i))$. Therefore, ω is random with respect to $\overline{P}$.

The already established convergence of the series $\sum_{i=0}^{\infty} P(A(\omega^i) \mid \omega^i)$ implies the convergence of $\sum_{i=0}^{\infty} \rho_2(\overline{Q}_i^{\omega}, \overline{P}_i^{\omega})$. In conjunction with Theorem 1 this implies that ω is random with respect to $\overline{Q}$.

To prove the randomness of ω with respect to Q it suffices to prove that $\omega_n \in A(\omega^n)$ only finitely often. If $\omega_n \in A(\omega^n)$ were true for infinitely many n, we would have $\overline{Q}(\omega_n \mid \omega^n) = P(\omega_n \mid \omega^n)/2$ for infinitely many n, which would contradict Theorem 2. □

A related result—a criterion of absolute continuity and singularity of probability measures in "predictable" terms—has been obtained in probability theory ([5]; see also [6, p. 516, Theorem 4]). In the case of the probability space $(X, \mathcal{P}(X))^{\infty}$, where $\mathcal{P}(X)$ is the set of all subsets of the set X, it is a simple corollary of our criterion of randomness.

Institute of Problems of Cybernetics
Academy of Sciences of the USSR
Moscow

Received
03.02.1986

LITERATURE

1. Колмогоров А. Н. – Пробл. передачи информации, 1965, vol. 1, No 1, pp. 3–11. 2. Вьюгин В. В. – Семиотика и информатика, 1981, vol. 16, pp. 14–43. 3. Успенский В. А., Семенов А. Л. In: Алгоритмы в современной математике и ее приложениях. Novosibirsk, 1982, vol. 1, pp. 99–342. 4. Боровков А. А. Математическая статистика. Moscow: Nauka, 1984. 472 pp. 5. Кабанов Ю. М., Липцер Р. Ш., Ширяев А. Н. – Матем. сб., 1977, vol. 104 (146), pp. 227–247. 6. Ширяев А. Н. Вероятность. Moscow: Nauka, 1980. 575 pp.

Remarks by Volodya Vovk (6 April 2008)

This paper was published as

В. Г. Вовк. Об одном критерии случайности. Доклады Академии Наук СССР, 294(6):1298–1302, 1987.

Another English translation (I have never seen it) appeared as

V. G. Vovk. On a randomness criterion. *Soviet Mathematics Doklady*, 35(3):656–660, 1987.

In my translation I corrected one misprint. These are English translations of references [1]–[6]:

1. A. N. Kolmogorov. Three approaches to the quantitative definition of information. *Problems of Information Transmission*, 1:1–7, 1965. See also Kolmogorov's *Collected Papers*.

2. V. V. V'yugin. Algorithmic entropy (complexity) of finite objects and its applications to defining randomness and amount of information. *Selecta Mathematica Sovietica*, 13:357–389, 1994.

3. There exists an English translation of [3], but it has been superseded by the English translation of a Russian book extending [3]: V. A. Uspensky and A. L. Semenov. *Algorithms: Main Ideas and Applications*. Norwell, MA: Kluwer Academic Publishers, 1993.

4. A. A. Borovkov. *Mathematical Statistics*. Amsterdam: Gordon and Breach Science Publishers, 1998. Page 194 of the Russian original corresponds to page 177 of the translation.

5. Yu. M. Kabanov, R. Sh. Liptser, and A. N. Shiryaev. To the question of absolute continuity and singularity of probability measures. *Mathematics of the USSR—Sbornik*, 33:203–221, 1977.

6. A. N. Shiryaev. *Probability*. Second edition. New York: Springer, 1996. The theorem mentioned in the paper is Theorem 4 on p. 528 of the translation (Chapter VII, §6).