PUBLICATIONS DE L'INSTITUT MATHÉMATIQUE
Nouvelle série, tome 36 (50), 1984, pp. 107–109

NOTE ON THE NUMBER OF SEQUENCES WITH GIVEN COMPLEXITY

Dragan Banjević

Abstract. Kolmogorov in 1964 has defined the notion of complexity of a finite binary sequence. In this paper some properties of the number of sequences with given complexity are considered.

Let S be the set of all finite binary sequences, and let $l(x)$, $x \in S$ be the length of x. Let $K(x)$ be the complexity of x with respect to some optimal function F_0^1 and let $K(x/y)$ be the conditional complexity of x given y (see [1] – [3]). It is well known that the complexity of most of the sequences is close to their length (see [2]), or

$$(\exists c)(\forall x)(K(x) \leq l(x) + c), \tag{1}$$

$$|\{x : l(x) = n,\ K(x) \geq n - m\}| \geq 2^n(1 - 1/2^m), \tag{2}$$

($|A|$ is the number of elements in A).

In [3.1.h] it was proved that

$$m - 2\log m \leq \log|\{x : K(x) \leq m\}| \leq m, \tag{3}$$

$$\log|\{x : K(x/m) \leq m\}| \asymp m. \tag{4}$$

$$(F \leq G \Leftrightarrow (\exists c)(\forall x)(F(x) \leq G(x) + c),\ F \asymp G \Leftrightarrow F \leq G \wedge G \leq F,\ \log = \log_2)$$

In this paper we consider some improvements for (3) and (4). Proofs are similar to the proofs for 1. (h) in [3], and they are based on Theorem 1.6 in [2].

(a) *Let*

$$A_n = \{x : l(x) = n,\ K(x) \geq n\}, \qquad A'_n = \{x : l(x) = n, \quad K(x/n) \geq n\},$$
$$B_n = \{x : l(x) = n,\ K(x) \geq n - 1\}, \qquad B'_n = \{x : l(x) = n, \quad K(x/n) \leq n - 1\},$$

Then

$$n - 2\log n \leq \log|A_n| \leq n, \log|A'_n| \asymp n, \tag{5}$$

AMS Subject Classification (1980): Primary 60-04, 60 A 05

$$n - 2\log n \le \log|B_n| \le n, \log|B'_n| \asymp n. \tag{6}$$

We note that this is an improvement of (2), in case $m = 0$.

Proof. We have immediately $|A_n| + |B_n| = 2^n$, $\log|A_n| \le n$, $\log|B_n| \le n$, and by [3, 1. g] $\min\{K(x) : l(x) = n\} \asymp K(n) \le l(n)$, which implies $(\forall n_0)(\exists n \ge n_0)(\exists x)(l(x) = n,\ K(x) < n)$, or $B_n \ne \emptyset$ for $n \ge n_0$. On the other hand $x \in B_n$ means $(\exists p_x)(l(p_x) \le n-1,\ F_0^1(p_x) = x)$ and the number of programs p_x is at the most $2^0 + 2^1 + \ldots + 2^{n-1} = 2^n - 1 < 2^n$, which implies that $0 < |B_n| < 2^n$, $0 < |A_n| < 2^n$, or $A_n \ne \emptyset$ for $n \ge n_0$.

Let, for given $p = \bar{a}b$, the function $F(p)$ be defined in such a way that we first choose the set A of exactly b sequences y for which $K(y) \le a$, and then $x = F(p)$ is the first y such that $y \notin A$. Then for $a = n$, $b = |B_n|$, we have $x = F(p) \notin B_n$, and $n \le K(x) \le K_F(x) \le l(p) \le \log|B_n| + 2\log n$. Further, let $F'(\bar{a}b) = F(\bar{a}(2^a - b))$, and then $F'(\bar{n}|B_n|) = x \notin B_n$, which implies $n \le K(x) \le K_{F'}(x) \le l(\bar{n}|A_n|) \le \log|A_n| + 2\log n$.

In the case concerning B'_n, we can take $F(b,a) = F(\bar{a}b)$, and then $F(|B'_n|, n) = F(\bar{n}|B'_n|) = x \notin B'_n$, or $n \le K(x\|)\| \le K_F^-(x/n) \le l(|B'_n|) \asymp \log|B'_n| \le n$. The proof for $|A'_n|$ is similar.

(b) *Let $g(n)$ be a recursive increasing function, and $g(n) < n-1$. Let*

$$A_n = \{x : l(x) = n,\ K(x) > g(n)\},\quad A'_n = \{x : l(x) = n,\ K(x/n) > g(n)\},$$
$$B_n = \{x : l(x) = n,\ K(x) \le g(n)\},\quad B'_n = \{x : l(x) = n,\ K(x/n) \le g(n)\}.$$

Then

$$\log|A_n| \asymp \log|A'_n| \asymp n, \tag{7}$$
$$g(n) - 2\log n \le \log|B_n| \le g(n),\ \ \log|B'_n| \asymp g(n). \tag{8}$$

Proof. Let c_0 be such that $(\forall n)(\exists x)(l(x) = n,\ K(x) \le \log n + c_0)$, by [3, 1. g]. Let $g(n) - 2\log n \ge c_0$. Then $B_n \ne \emptyset$, and imitating the proof in (a), we can get (8). If $g(n) - 2\log n < c_0$, then (8) is trivial. Equality (7) is a direct consequence of (2), or simply $2^n \ge |A_n| = 2^n - |B_n| \ge 2^n - (2^{g(n)+1} - 1) \ge 2^n - 2^{n-1} = 2^{n-1}$, and $n \ge \log|A_n| \ge n-1$.

(c) *Let*

$$D_n = \{x : l(x) \le n,\ K(x) > n\},\qquad D'_n = \{x : l(x) \le n,\ K(x/n) > n\},$$
$$E_n = \{x : l(x) \le n,\ K(x) \le n\},\qquad E'_n = \{x : l(x) \le n,\ K(x/n) \le n\},$$
$$c_n = \max\{K(x) - n : l(x) = n\},\qquad c'_n = \max\{K(x/n) - n : l(x) = n\},$$
$$B_n = \{x : l(x) = n,\ K(x) = n + c_n\},\quad B'_n = \{x : l(x) = n,\ K(x/n) = n + c'_n\},$$
$$F_n = \{x : l(x) = n,\ K(x) > n\},\qquad F'_n = \{x : l(x) = n,\ K(x/n) > n\}.$$

Then

$$n - 2\log n \le \log|B_n| \le n,\ \log|B'_n| \asymp n, \tag{9}$$

$$n - 2\log n \le \log|D_n| \le n,\ \log|D'_n| \asymp n, \tag{10}$$

$$n - 2\log n \le \log|E_n| \le n,\ \log|E'_n| \asymp n, \tag{11}$$

and for infinitely many n, $c_n > 0$, or $F_n \ne \emptyset$, and in that case

$$n - 2\log n \le \log|F_n| \le n,\ \log|F'_n| \asymp n. \tag{10}$$

PROOF. By definition of c_n, and by (1), $0 \le c_n \le c$, and $B_n \ne \emptyset$, for all n. Similarly to the previous case we can prove (9). In order to prove (10), we shall prove that $D_n \ne \emptyset$. Let $H(n) = 2^n - 1$, and $h(n) = [\sqrt{n}]$ (integer part of $\sqrt{n}$). Then $K(H(H(h(n)))) \asymp K(h(n)) \le l(n)/2$, but $l(H(H(h(n)))) = H(h(n)) > n$, which implies that for every $n \ge n_0$ there exists an x such that $K(x) < n$ and $l(x) > n$, which is equivalent to the existence of an y such that $l(y) \le n$ and $K(y) > n$, which means $D_n \ne \emptyset$ for $n \ge n_0$. Then, it is easy to prove (10). Furthermore, $D_n \ne \emptyset$ for $n \ge n_0$, implies that for infinitely many n we have $F_n \ne \emptyset$, and then (12).

REFERENCES

[1] A.N. Kolmogorov, *Logical basis for information theory and probability theory*, IEEE Trans. Information Theory IT-**14** (1986), 662–664.

[2] А.К. Звонкин, Л.А. Левин, *Сложность конечных обьектов и обоснование понятий информации и случайности с помощью теории алгоритмов*, Успехи Мат. Наук. **25** (1970), 85–127.

[3] D. Banjević, *On some basic properties of the Kolmogorov complexity*, Publ. Inst. Math. (N.S.) (Beograd) **30** (**44**) (1981), 17–23.

Institut za matematiku
Prirodno-matematički fakultet
Studentski trg 16, Beograd

(Received 03 10 1983)