Math. Systems Theory **29**, 271–292 (1996)

Mathematical Systems Theory
© 1996 Springer-Verlag
New York Inc.

Relations Between Varieties of Kolmogorov Complexities*

V. A. Uspensky[1] and A. Shen[2]

[1] Division of Mathematical Logic and the Theory of Algorithms,
Fakultet of Mechanics and Mathematics, Lomonosov Moscow State University,
V-234 Moscow, GSP-3, 119899 Russia
uspensky@viniti.msk.su
vau@uspensky.ras.ru

[2] Laboratory 13, Institute for Problems of Information Transmission,
Ermolovoi 19, K-51 Moscow, GSP-4, 101447 Russia
shen@landau.ac.ru

Abstract. There are several sorts of Kolmogorov complexity, better to say several Kolmogorov complexities: decision complexity, simple complexity, prefix complexity, monotonic complexity, *a priori* complexity. The last three can and the first two cannot be used for defining randomness of an infinite binary sequence. All those five versions of Kolmogorov complexity were considered, from a unified point of view, in a paper by the first author which appeared in Watanabe's book [23]. Upper and lower bounds for those complexities and also for their differences were announced in that paper without proofs. (Some of those bounds are mentioned in Section 4.4.5 of [16].) The purpose of this paper (which can be read independently of [23]) is to give proofs for the bounds from [23].

The terminology used in this paper is somehow nonstandard: we call "Kolmogorov entropy" what is usually called "Kolmogorov complexity." This is a Moscow tradition suggested by Kolmogorov himself. By this tradition the term "complexity" relates to *any* mode of description and "entropy" is the complexity related to an *optimal* mode (i.e., to a mode that, roughly speaking, gives the *shortest* descriptions).

* The research described in this publication was made possible in part by Grant No. MQ3000 from the International Science Foundation. A preliminary version of this paper appeared in April, 1993, as CWI's Technical Report CS-R9329 (CWI is the Dutch National Research Institute for Mathematics and Computer Science). That version was supported by the Netherlands Organization for Scientific Research (NWO). The second author was also supported by AMS fSU Aid fund and Volkswagen Stiftung.

1. Introduction

This paper collects various definitions of algorithmic complexity (entropy) and information about their relations. All those definitions and facts were given in [23] without proofs; here the proofs are given. Many of these proofs are well known; nevertheless, all the proofs are collected here for the reader's convenience and adapted to the uniform terminology.

The paper is organized as follows. We start (Section 2) with the classification of four entropies (two possibilities for objects combined with two possibilities for descriptions) which goes back to [21] and is explained in Sections 1.2 and 1.3 of [23].

Then in Section 3 we look at a different classification of entropies which goes back to [15] and establish the connections between these two classifications mentioned in Section 1.6 of [23].

Finally, in Section 4 we establish some connections between different entropies mentioned in Sections 2.1 and 2.2 of [23].

2. Objects and Descriptions

Any of the four definitions of entropy given in this section follows the same pattern. First, an appropriate notion of "description mode," or "mode of description," is introduced. Each of the four definitions requires a specific class of description modes. Any description mode is a binary relation E on Ξ (the set of all binary words). If $\langle x, y\rangle \in E$, then x is called a *description* of y. When a mode E is fixed, a *complexity* of a binary word y is defined as the length of its shortest description, i.e.,

$$K_E(y) = \min\{|x| \mid \langle x, y\rangle \in E\},$$

where $|x|$ denotes the length of x. Different modes of description lead to different complexity functions K_E; the basic Solomonoff–Kolmogorov theorem (valid for all four entropies of this section) states that among all the functions related to the relevant class of modes there is a minimal one (up to an additive constant). In other words, in the class of modes there is an *optimal* description mode E such that, for any description mode F of the same class,

$$K_E(y) \leq K_F(y) + C$$

for some constant C and for all words y. Finally, *entropy* is defined as K_E for some optimal description mode E.

Now we use this general scheme for four different cases.

2.1. *Simple Kolmogorov Entropy*

When defining simple Kolmogorov entropy, a *mode of description* ("simple description mode") is a binary relation $E \subset \Xi \times \Xi$ such that, for every x, y_1, y_2 in Ξ,

$$\langle x, y_1\rangle \in E \wedge \langle x, y_2\rangle \in E \quad \Rightarrow \quad y_1 = y_2.$$

In other terms, a mode of description is a (partial) function from Ξ into Ξ. Enumerable (i.e., recursively enumerable) modes of descriptions correspond to computable functions; we restrict ourselves to enumerable modes only.

When a mode E is fixed, the *complexity* of a binary word y is defined as the length of its shortest description, i.e.,

$$K_E(y) = \min\{|x| \mid \langle x, y\rangle \in E\}.$$

Different modes of description lead to different complexity functions K_E; the basic Solomonoff–Kolmogorov theorem states that among all these functions a minimal one (up to an additive constant) exists. In other words, there is an *optimal* description mode E such that, for any description mode F,

$$K_E(y) \leq K_F(y) + C$$

for some constant C and for all words y.

To construct an optimal mode of description, assume that $U(m, n)$ is a universal computable function (i.e., the family $\{U_m\}$, where $U_m(x) = U(m, x)$, contains all computable functions, including partial ones, from Ξ to Ξ). By $\overline{z}$ we denote the word z where each letter is repeated twice. An optimal mode of description may be constructed as follows:

$$E = \{\langle \overline{p}01q, r\rangle | U(p, q) = r\}.$$

Now we fix some optimal description mode E and call the corresponding complexity function $K_E(y)$ *simple Kolmogorov entropy*. It is denoted $KS(y)$ in what follows, and the description modes as defined in this section are called "KS-description modes" or "simple description modes."

This definition of simple Kolmogorov entropy appears in Section 1.2 of [23] where the name "(=, =)-entropy" or "$\mathbb{N}$-entropy" is used. Essentially the same definition is given in Section 1.3 of [23]. Indeed, the ordering on the bunch $\mathbb{B}$ is trivial (only equal objects are comparable), therefore conditions 1 and 2 [23, p. 89] are always satisfied. Condition 3 means that E is a graph of a function, and acceptable modes of descriptions are graphs of computable functions. Therefore, "bunch definition" of [23] coincides with the one given above (and with the original Kolmogorov definition from [12]).

2.2. *Decision Entropy*

For the case of decision entropy a description mode ("decision description mode") is defined as a (recursively enumerable) set $E \subset \Xi \times \Xi$ satisfying the following requirements:

(a) If $\langle x, y_1\rangle \in E$ and $\langle x, y_2\rangle \in E$, then one of the words y_1 and y_2 is a prefix of another one.
(b) If $\langle x, y\rangle \in E$, then $\langle x, y'\rangle \in E$ for all prefixes y' of y.

It is easy to see that, for any fixed x, all y's such that $\langle x, y\rangle \in E$ are prefixes of some (finite or computable infinite) binary string. So the mode of description may be naturally considered as a mapping e of Ξ into the set of all finite or computable infinite binary strings, and $\langle x, y\rangle \in E$ means "y is an initial segment of $e(x)$."

Then decision complexity with respect to a given mode E is defined as before, and again the optimal description mode E exists. The corresponding complexity function $K_E(y)$ is called *decision entropy* and is denoted by $KD(y)$.

Again the Solomonoff–Kolmogorov theorem is valid for this case. The construction of the optimal description mode follows. Assume that $U(p, q, n)$ (where p, q are binary words and n is a natural number) is a computable function with 0–1 values universal for the class of all computable functions $\Xi \times \mathbb{N} \to \{0, 1\}$. Then the set

$$\{\langle \bar{p}01q, r\rangle |\ r_i = U(p, q, i) \text{ for all } i \text{ not exceeding } |r|\}$$

(by r_i we denote the ith bit of r) is an optimal description mode. (This description mode follows the original construction of decision entropy, see [17] or [26].)

The above-mentioned requirements (a) and (b) (given as in Section 1.3 of [23], for $X = \mathbb{B}$, $Y = \mathbb{T}$) seem natural if we think of a description mode as a computable mapping in the sense of the Scott–Ershov domain theory (see [21]). However, requirement (b) may in fact be omitted (as in Section 1.2 of [23]). Then we get a broader class of description modes and, theoretically speaking, may get a smaller entropy. However, for any binary relation E satisfying requirement (a) we may consider its extension E':

$$E' = \{\langle x, y\rangle | y \text{ is a prefix of some } y' \text{ such that } \langle x, y'\rangle \in E\}.$$

It is easy to check that this extension is enumerable if E is, that E' satisfies both requirements (a) and (b), and that the corresponding complexity function does not exceed the complexity function corresponding to E.

The decision entropy is called $(=, \gamma)$-entropy, or $\mathbb{N}\Xi$-entropy in Section 1.2 of [23].

2.3. *Monotonic Entropy*

Here by the description mode ("monotonic description mode") we mean a (recursively enumerable) set $E \subset \Xi \times \Xi$ satisfying the following requirements (see Section 1.3 of [23]):

(a) If $\langle x, y\rangle \in E$, then $\langle x, y'\rangle \in E$ for all prefixes y' of y.
(b) If $\langle x, y\rangle \in E$, then $\langle x', y\rangle \in E$ for all x' having x as a prefix.
(c) If $\langle x, y'\rangle \in E$ and $\langle x, y''\rangle \in E$, then one of the words y', y'' is a prefix of another one.

Then the complexity (for a given mode) is defined in the usual way, as the length of the shortest description.

The optimal description mode does exist; corresponding complexity is called *monotonic entropy* and is denoted by $KM(y)$

Here to prove the existence of an optimal description mode is slightly more difficult than in the previous cases. The reason is that we should construct the "universal computable mapping" for the family of all "computable monotone mappings" from Ξ into Ξ. This is explained in the general case (for semantic domains, or f_0-spaces) in [21]; a very detailed description of what happens for the case of monotonic entropy is given in Sections 3.1 and 3.2 of [24].

Again, the requirements for the description mode may be weakened. Namely, we may require only (as in Section 1.2 of [23]) that if

$$\langle x_1, y_1\rangle \in E \quad \text{and} \quad \langle x_2, y_2\rangle \in E$$

and one of the words x_1, x_2 is a prefix of another one, then one of the words y_1 and y_2 is a prefix of another one.

It is easy to check that this requirement is a consequence of requirements (a)–(c) above (we may replace x_1 and x_2 by the longest of them), but not vice versa. However, if E satisfies the latter requirement, then its extension E' defined as

$$E' = \{\langle x, y\rangle \mid \text{there are } x' \leq x \text{ and } y' \geq y \text{ such that } \langle x', y'\rangle \in E\}$$

(here $p \leq q$ means that a binary word p is a prefix of a binary word q) satisfies requirements (a)–(c). Using this extension, it is easy to check that both versions of monotonic entropy definition lead to functions which differ only by a bounded additive term.

Monotonic entropy is called (γ, γ)-entropy, or $\Xi\Xi$-entropy, in Section 1.2 of [23].

2.4. *Prefix Entropy*

Here the requirements for the description mode ("prefix description mode") are as follows (see Section 1.3 of [23]):

(a) If $\langle x, y\rangle \in E$, then $\langle x', y\rangle \in E$ for any x' such that x is a prefix of x'.
(b) If $\langle x, y_1\rangle \in E$ and $\langle x, y_2\rangle \in E$, then $y_1 = y_2$.

(As everywhere, E is supposed to be recursively enumerable.) They can be replaced by the weaker requirement (see Section 1.2 of [23]): if $\langle x_1, y_1\rangle \in E$ and $\langle x_2, y_2\rangle \in E$ and x_1 is a prefix of x_2, then $y_1 = y_2$. This requirement, though being weaker, leads to the same entropy. Indeed, if some E satisfies this requirement, then its extension

$$E' = \{\langle x, y\rangle | \langle x', y\rangle \in E \text{ for some } x' \text{ being a prefix of } E\}$$

satisfies both requirements (a) and (b) and gives the same complexity function.

The existence of an optimal description mode may be proved by enumerating all description modes (in other terms, all "computable mappings" from Ξ to $\mathbb{N}$). Its existence follows from the general facts about semantic domains (see [21]) and can also be proved directly. We omit this proof because the existence of an optimal mode is a by-product of the coincidence of the definition given above and the encoding-free definition (see the next section).

The complexity with respect to an optimal description mode in the sense of this section is called *prefix entropy* and is denoted by $KP(x)$.

2.5. *Historical Remarks*

The different versions of entropy described above (as well as some other versions) were invented independently by different people. If we attribute those versions according to the first publication date, the list would be as follows:

- Simple entropy KS: 1965, Kolmogorov [12, Section 3]; and (even earlier but in some nebulous form) 1964, Solomonoff [22].
- Decision entropy KD: 1969, Loveland [17].
- *A priori* entropy KA (see below): 1973, Levin (see [26, no. 3.3] and [13]).
- Monotonic entropy KM: 1973, Levin [13].
- Prefix entropy KP: 1974, Levin [14], see also [8].

Some other historical remarks:

- In 1966 Chaitin published his paper [2], where a complexity measure was defined in terms of Turing machine parameters. This definition, however, does not provide the optimal complexity measure, which appeared in a subsequent paper published in 1969 [3]. (According to [16], p. 86, those papers were submitted in October 1965 and November 1965, respectively.) In a publication of 1975 Chaitin also reinvented the prefix entropy (see [4]). See also [5] and [6].

  In [7] Chaitin writes: “I have been the main intellectual driving force behind both AIT_1 and AIT_2.” As to AIT_1 and AIT_2, in [7] there is a declaration that Algorithmic Information Theory “appeared in two installments,” and AIT_1, AIT_2 stand for those installments. Here is the opinion of one of the leading experts in the field: “Chaitin has done more than others to popularize some aspects of algorithmic information theory. The benefits of this activity are offset by his somewhat narrow interests $\langle \dots \rangle$ and the way he ascribes all major achievements to himself” [10].
- In 1964 Markov, Jr. [18], proposed a complexity measure similar to decision entropy. It was based on so-called “normal algorithms.” However, his definition did not provide an optimal complexity measure.
- Monotonic entropy was defined (in its present form) in Levin’s paper [13] together with the characterization of randomness in terms of that entropy. At the same time Schnorr [19] independently provided a similar characterization, but his notion of entropy (“process complexity” according to Schnorr) was slightly different. Later Schnorr [20] discarded his notion and used the same notion of monotonic entropy as given in Levin’s paper.

The complete account of the history of different notions related to Kolmogorov complexity may be found in the recently published monograph [16].

3. Encoding-Free Definitions

3.1. *Simple Kolmogorov Entropy*

The simple Kolmogorov entropy can be characterized as a minimal (up to a constant) enumerable from above function $f\colon\ \Xi \to \mathbb{N} \cup \{\infty\}$ satisfying the following condition (which, in an equivalent form, is called ($\mathbf{C}\mathbb{B}$) in Section 1.5 of [23]):

- There is at most 2^n different y such that $f(y) = n$.

(A function $f\colon\ \Xi \to \mathbb{N} \cup \{\infty\}$ is called *enumerable from above* if the set of all pairs $\langle x, n\rangle$ such that $n > f(x)$ is recursively enumerable.)

Remark. If we replace 2^n by $C \cdot 2^n$ (see condition ($\mathbf{C}'$) in Section 1.5 of [23]) we get the same (up to a constant) entropy: $C \cdot 2^n = 2^{n+\log C}$, therefore this factor C corresponds to an additive constant in the exponent. We may also replace “$f(y) = n$” by “$f(y) \le n$”; if there is at most 2^n objects y such that $f(y) = n$, then the number of objects y such that $f(y) \le n$ does not exceed $1 + 2 + \dots + 2^n < 2 \cdot 2^n$.

To prove this characterization of simple Kolmogorov entropy (as defined in Section 2.1) we should prove that:

- A simple Kolmogorov entropy function $KS(x)$ satisfies this condition.
- For any enumerable from above function f satisfying this condition a simple description mode E can be constructed such that the complexity function corresponding to E exceeds f by not more than a constant.

The first claim is trivial: different objects have different descriptions, and objects y such that $KS(y) = n$ have descriptions of length n. Therefore, the number of those y's does not exceed the total number of descriptions having length n, i.e., 2^n.

The second claim is also simple. We reserve words of length n to be descriptions of objects y such that $f(y) < n$. The total number of these objects does not exceed $1 + 2 + \cdots + 2^{n-1} < 2^n$, therefore we cannot exhaust all reserved words. The function f is by assumption enumerable from above. Thus, the set of all pairs $\langle y, n\rangle$ such that $f(y) < n$ is enumerable. When a new pair $\langle y, n\rangle$ appears during the enumeration process, we allocate one of the unused words e of length n to be a description of y. The set E of all pairs $\langle e, y\rangle$ generated in this way is enumerable; E is a function graph (because each e may be allocated only once), therefore, E is a simple description mode. Evidently, the corresponding complexity function does not exceed $f + 1$.

A by-product of this argument is the existence of a minimal (up to an additive constant) enumerable from above function satisfying our condition.

3.2. *Decision Entropy*

To get the characterization of decision entropy KD we should look for the minimal (up to a constant) function $f\colon\ \Xi \to \mathbb{N} \cup \{\infty\}$ which is enumerable from above and satisfies the following condition:

- If M is a finite set of incomparable words (there is no word in M which is a prefix of another word in M) and $M \subset f^{-1}(n)$, then the cardinality of M does not exceed 2^n.

(The equivalent condition is called ($\mathbf{C}\mathbb{T}$) in [23]). As in the previous section, to prove this characterization we should prove that:

- A decision entropy function $KD(x)$ satisfies this condition.
- For any enumerable from above function f satisfying this condition, a decision description mode E can be constructed such that the complexity function corresponding to E exceeds f by not more than a constant.

We start with the first claim. Assume that M is prefix-free (no word in M is a prefix of another one in M) set of words having decision entropy n. That means that all these words have descriptions of length n. All these descriptions must be different (otherwise the conditions for the description mode are violated). Thus, the number of descriptions (and the cardinality of M) does not exceed 2^n.

Now consider the second claim. As well as in the previous section we reserve words of length n to be descriptions of objects y such that $f(y) < n$. Now the total number of objects y such that $f(y) = n$ is not limited; however, any subset of pairwise incomparable

y's such that $f(y) = n$ has cardinality not greater that 2^n (two words are *comparable* if one of them is a prefix of another one). Therefore, any set of pairwise incomparable objects with f-values less than n contains no more than $1+2+\cdots+2^{n-1} < 2^n$ objects. The function f is by assumption enumerable from above. Thus, the set of all pairs $\langle y, n\rangle$ such that $f(y) < n$ is enumerable. Assume that a new pair $\langle y, n\rangle$ appears during the enumeration process. For each already allocated description e we look at the longest object $z(e)$ in the set of all objects having e as a description. (All other objects in this set will be prefixes of the longest one.) If any of these objects $z(e)$ is comparable with y, then the corresponding e is declared to be a description of y. If not, we allocate a new description for y. (There is a free description because all $z(e)$ together with y are incomparable and therefore the number of used e's is less than 2^n.) The set of all pairs $\langle e, y\rangle$ generated in this way is an enumerable decision description mode (i.e., satisfies the conditions of Section 2.2). Evidently, the corresponding complexity function does not exceed $f + 1$.

3.3. *A Priori Entropy*

In the case of monotonic entropy, situations differs: monotonic entropy has no exact characterization of the same type as in Section 3.1 and 3.2. However, it is connected closely with another complexity measure, called *a priori* probability. We reproduce its original definition from Section 3 of [26], where it is called a "universal semicomputable measure." (This notion is discussed in details in Chapter V of [24].)

A *semimeasure* (in this section!) is a function m defined on Ξ with nonnegative real values satisfying the following conditions:

- $m(\Lambda) = 1$ (here Λ denotes an empty word).
- $m(x0) + m(x1) \leq m(x)$ for any word x.

A semimeasure is called *enumerable from below* if the set of all pairs $\langle x, r\rangle$ such that r is a rational number less than $m(x)$ is enumerable. There is a maximal (up to a constant factor) enumerable from below semimeasure $M(x)$ called *a priori probability* (see [24]). Its logarithm is called *a priori entropy* and is denoted by KA.

Another definition of *a priori* entropy is given in Section 1.5 of [23]. Namely, *a priori* entropy is defined there under the name of $\Sigma\mathbb{T}$-entropy as a minimal enumerable from above function $f\colon \Xi \to \mathbb{N} \cup \{\infty\}$ such that

$$(\Sigma\mathbb{T}) \qquad \sum_{y\in M} 2^{-f(y)} \leq 1 \qquad \text{for any finite prefix-free set} \quad M \subset \Xi$$

("prefix-free" means that no word in M is a prefix of another word in M).

We explain shortly why these two definitions are equivalent. The main role is played by the following two facts:

- If $m(x)$ is a semimeasure, then $f(k) = [\text{minimal } k \text{ such that } 2^{-k} < m(x)]$ satisfies the condition $(\Sigma\mathbb{T})$.
- If a function f satisfies the condition $(\Sigma\mathbb{T})$, then the function $m(x)$ defined as $\max \sum_{x\in D} 2^{-f(x)}$, where maximum is taken over all finite prefix-free sets D such that x is a prefix of each word in D, is a semimeasure. (Technically speaking, we should also change the value of m on Λ and assume that $m(\Lambda) = 1$.)

These facts establish an approximate (up to a factor of 2) correspondence between semimeasures and functions satisfying the condition $\Sigma\mathbb{T}$ which preserves enumerability and allows us to prove the coincidence mentioned above.

There is one more assertion concerning the definition of *a priori* entropy using the $(\Sigma\mathbb{T})$ condition. It is called "Muchnik's theorem" on p. 93 of [23]. It can be stated as follows. Assume that function φ is defined on binary words and all $\varphi(x)$ are real numbers between 0 and 1. We consider any binary word x as a vertex in a complete binary tree and $\varphi(x)$ as its label. Assume that, for each C, we can find a finite set of pairwise incomparable words with the sum of labels exceeding C. Then an infinite set of pairwise incomparable words with an infinite sum of labels exists.

The scheme of the proof is as follows. For each binary word x (each vertex of the tree) consider all sets D of pairwise incomparable words having x as a prefix. For each D compute the sum of all labels of vertices from D and take a supremum over all D's; this supremum (finite or infinite) depends on x. We call a vertex *bad* if that supremum is infinite. By assumption, the tree root is bad. We should find an infinite set of pairwise incomparable words with an infinite sum of labels. Bad vertices form a subtree in the full binary tree; this subtree has no leaves (if x is bad, at least one of the words $x0$ and $x1$ is bad). Now we consider two cases:

- There is a bad vertex x such that its bad descendants form a path (any two bad descendants of x are comparable).
- For any bad vertex x there are two incomparable bad descendants of x.

In both cases it is possible to find the required infinite set of vertices with an infinite sum of labels.

3.4. *Prefix Entropy*

The prefix entropy with its encoding-free definition given in this section is probably the most technically interesting among all the four entropies. It is discussed in detail in [25]; however, an English translation of this paper has not been published yet, so we try to give a self-contained description of what happened in this case.

We start with the another definition of a semimeasure. The corresponding notion differs from the notion of semimeasure used in the previous section. The underlying reason for this difference is that in the previous section binary words were considered as vertices of a binary tree; now this structure is ignored and all the word are "placed on the same level," so Ξ is treated not as a tree but as a "bunch."

In this section a *semimeasure* is a (total) function m defined on the set Ξ of all binary words with nonnegative real values such that $\sum_x m(x) \leq 1$.

A semimeasure m is called *enumerable from below* if the set of all pairs $\langle x, r\rangle$ such that r is a rational number less than $m(x)$ is enumerable.

Enumerable from below semimeasures correspond to probabilistic machines which have no input but have an output where a binary word may appear (after it appears, the machine terminates). Namely:

- If M is a probabilistic machine of this type, the function $P^s_M(y)$ = the probability of the event "machine M stops with output y" is a semimeasure enumerable from below.

- For any semimeasure m enumerable from below a probabilistic machine M can be constructed such that $m(x) = P^s_M(x)$ for all x.

The first claim is almost evident. Indeed, the sum $\sum_x P^s_M(x)$ is the probability of the event "machine M stops" and therefore does not exceed 1. Function P^s_M is also enumerable from below: trying to emulate the computation process of M for all possible random bits, we get more and more cases where the output is known and therefore may generate the lower bounds for $P^s_M(x)$.

Now we proceed to the second claim. We give only a sketch of a proof. Assume that a semimeasure $m(x)$ enumerable from below is given and we are looking at the process of enumeration of all rational lower bounds for all $m(x)$. Assume that $m_k(x)$ is a current lower bound for $m(x)$ at the kth step. We may assume that for each k the value $m_k(x)$ differs from 0 only for finitely many x's, that $m_k(x)$ increases when k increases and converges to $m(x)$. Our probabilistic space is the set of all infinite 0–1 sequences. At step k we allocate the part of it having measure $m_k(x)$ to the output x; this part increases when k and $m_k(x)$ increase. (End of sketch.)

There is an enumerable from below semimeasure $M(x)$ which is maximal in the following sense: for any enumerable from below semimeasure $m(x)$ there is a constant c such that $m(x) \le e \cdot M(x)$ for all words x.

This fact can be proved as follows: enumerate all probabilistic machines and construct a "universal" machine which chooses a natural number i at random (probabilities p_i to choose i are assumed to be positive) and then simulates the ith machine. If m_i is a semimeasure corresponding to the ith machine and M is a semimeasure corresponding to the universal machine, then $M(x) \ge p_i \cdot m_i(x)$. Therefore, M is maximal.

Semimeasures are connected with functions $f\colon\ \Xi \to \mathbb{N} \cup \{\infty\}$ satisfying the following condition:

$$(\Sigma\mathbb{B}) \qquad \sum_x 2^{-f(x)} \le 1.$$

Namely:

- If f is a function satisfying condition $(\Sigma\mathbb{B})$, then $m(x) = 2^{-f(x)}$ is a semimeasure.
- If m is a semimeasure, then the function $f(x)$ = minimal k such that $2^{-k} < m(x)$ satisfies condition $(\Sigma\mathbb{B})$.

Therefore we can go back and forth between semimeasures and functions satisfying condition $(\Sigma\mathbb{B})$ and for the round-trip we pay at most factor 2 (or additive constant 1). It is easy to see that enumerable from below semimeasures correspond to enumerable from above functions. Therefore, the existence of a maximal enumerable from below semimeasure $M(x)$ implies the existence of a minimal enumerable from above function satisfying $(\Sigma\mathbb{B})$ and this function coincides with $-\log_2 M(x)$ up to an additive constant.

It turns out that the minimal function from the preceding paragraph (or logarithm of the maximal semimeasure) coincides with prefix entropy. So prefix entropy may be defined as a *minimal enumerable from above function f satisfying condition* $(\Sigma\mathbb{B})$.

To prove this coincidence we should prove two assertions:

- For any prefix description mode E the corresponding complexity function Compl_E satisfies condition ($\Sigma\mathbb{B}$) and is recursively enumerable from above.
- If a recursively enumerable from above function f satisfies condition ($\Sigma\mathbb{B}$), then a prefix description mode E exists such that the corresponding complexity function Compl_E exceeds f not more than by a constant.

The first assertion is almost trivial. If $M = \{m_1, \dots, m_k\}$ is a finite set of words and $e_1, \dots, e_k$ are their descriptions, then e_i are pairwise incomparable. Therefore, the corresponding intervals in the Cantor space (of all infinite 0–1 sequences) do not overlap and the total measure $\sum 2^{-e_i}$ does not exceed 1. Therefore, condition ($\Sigma\mathbb{B}$) is fulfilled.

The main role in the proof of the second assertion is played by the following construction. Consider the segment $[0, 1]$ divided into two equal parts $[0, \frac{1}{2}]$ and $[\frac{1}{2}, 1]$, each part is divided into two equal parts, etc. At level k we have 2^k parts of length 2^{-k} each. Assume that we get a sequence of natural numbers $n_1, n_2, \dots$ and each number s of this sequence is considered as a request to allocate a segment of level s (one of the 2^s segments of length 2^{-s}). The segments allocated by different requests should not overlap.

Of course, this goal may be achieved only if $\sum_i 2^{-n_i} \le 1$. It turns out that this condition is not only necessary but also sufficient. The simple allocation algorithm maintains the following invariant relation: all free space is represented as a union of nonoverlapping segments which belong to different levels (two segments of the same length should not appear in this union). The following allocation algorithm maintains this relation: if a segment of the required length is present in this union, allocate it; if not, take the smallest segment in the union whose length is sufficient and cut it into half + quarter $+ \cdots$ until a segment of required length appears.

This construction allows us to finish the proof of the second assertion. Assume that f is an enumerable from above function satisfying condition ($\Sigma\mathbb{B}$). Consider the set S of all pairs $\langle x, k\rangle$ such that $k > f(x)$. The set S is enumerable. If we add up all 2^{-k} for all pairs $\langle x, k\rangle \in S$, the sum does not exceed 1. Indeed, when we group all pairs $\langle x, k\rangle \in S$ with the same x we get

$$2^{-f(x)-1} + 2^{-f(x)-2} + 2^{-f(x)-3} + \cdots \le 2^{-f(x)},$$

and the sum $\sum_x 2^{-f(x)}$ does not exceed 1.

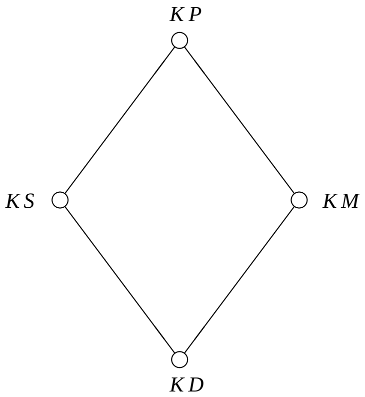
Now each pair $\langle x, k\rangle \in S$ will be considered as a request to allocate a segment of length 2^{-k}. These requests can be fulfilled (see the discussion above). Segments of level k may be indexed by k-bit 0–1 words in a natural way; allocating the segment with index e according to the request $\langle x, k\rangle \in S$, we declare e to be a description of the object x. The allocated segments do not overlap, therefore the descriptions of different objects are incomparable and the requirement of Section 2.4 (in its weakened form) is fulfilled. It is easy to see also that the minimal length of a description of an object x is $f(x) + 1$; therefore, the complexity function exceeds f by not more than 1.

This argument also implies that there is an optimal description mode (i.e., a description mode corresponding to the minimal function f which in its turn corresponds to a maximal semimeasure).

4. Inequalities between Entropies

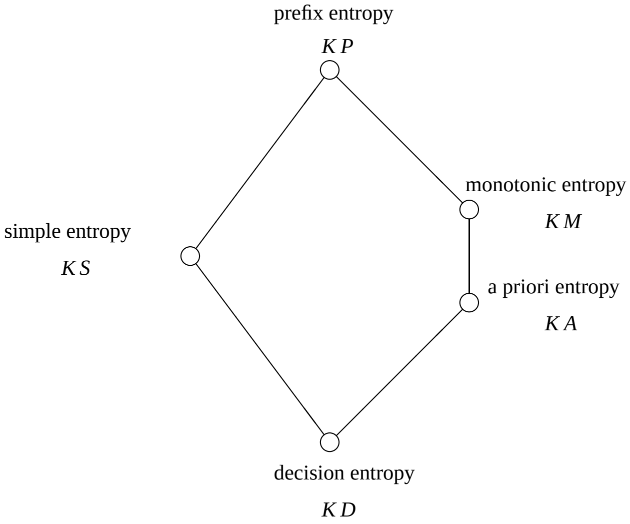
4.1. *Entropies Pentagon*

Four entropies form a diamond:



It is easy to see that restrictions for description modes become weaker when we go down along the sides of this parallelogram: each prefix description mode is a monotonic description mode and at the same time a simple description mode, etc. Weaker restrictions correspond to broader classes of description modes, so the entropy (defined as a minimum taken over all description modes) may only decrease. We shall see later that entropies do decrease when we go down.

So we get a picture where vertices correspond to entropies and edges correspond to inequalities between entropies. The entropy KA (*a priori* entropy) may be added to this picture:



Indeed,

- decision entropy does not exceed *a priori* entropy;
- *a priori* entropy does not exceed monotonic entropy.

We check that:

First, decision entropy may be defined as a smallest function satisfying the condition of Section 3.2. That condition is weaker than the similar condition for *a priori* entropy in Section 3.3—therefore, decision entropy does not exceed *a priori* entropy.

Second, we should prove that *a priori* entropy does not exceed monotonic entropy. This is explained in detail in [24], see Section 5.3; here we give only a short comment. Assume that we have an optimal monotonic description mode E. A semimeasure $m(z)$ can be defined as follows. Consider the set P_z of all infinite sequences $\omega = \omega_0\omega_1\ldots$ such that E contains a pair $\langle x, y\rangle$ such that x is a prefix of ω and z is a prefix of y. Define $m(z)$ as a uniform Bernoulli measure of the set P_z. It is easy to see that m is a semimeasure in the sense of Section 3.3 and that $m(z) \geq 2^{-KM(z)}$, where KM is a complexity function corresponding to description mode E.

In fact both inequalities mentioned above are strict: the difference between decision entropy and *a priori* entropy (as well as between *a priori* entropy and monotonic entropy) is unbounded, see below.

4.2. *Entropies and Lengths*

Any of the entropies $KS(x)$, $KM(x)$, $KA(x)$, $KD(x)$ does not exceed $|x| + C$ for some constant C. (Indeed, we may consider a description mode $E = \{\langle x, x\rangle \mid x \in \Xi\}$.) This upper bound cannot be improved significantly; we have $KD(x) \geq |x|$ for infinitely many x's. (KD is the smallest of the four entropies mentioned, so it is also true for other entropies.) Indeed, consider all the words y of a given length n. They are incomparable, therefore their KD descriptions should be different. If all these descriptions have length smaller than n, the total number of descriptions does not exceed

$$1 + 2 + 4 + 8 + \cdots + 2^{n-1} = 2^n - 1 < 2^n$$

—too few to provide descriptions for all n-bit words.

For prefix entropy the situation is more difficult. Consider the following divergent series (all logarithms are binary; we ignore the difficulties with $\log 0$, $1/\log 1$, etc.):

$$\sum \frac{1}{n}, \quad \sum \frac{1}{n\log n}, \quad \sum \frac{1}{n\log n \log\log n} \ldots.$$

At the same time the series

$$\sum \frac{1}{n^{1+\varepsilon}}, \quad \sum \frac{1}{n(\log n)^{1+\varepsilon}}, \quad \sum \frac{1}{n\log n(\log\log n)^{1+\varepsilon}} \ldots,$$

converge. Let us see how these series provide upper and lower bounds for prefix KP (see inequalities (2) and (3) on p. 99 in [23]). Enumerate all binary words in the lexicographic order (empty, 0, 1, 00, 01, 10, 11, etc.) and identify each word with its number. The series

$$\sum \frac{1}{n^{1+\varepsilon}}$$

converges. Therefore function $n \mapsto 1/n^{1+\varepsilon}$ is a semimeasure (in the sense of Section 3.4) when multiplied by some constant (to make the sum not exceed 1). Therefore, the prefix entropy of n (i.e., prefix entropy of binary word number n) does not exceed $(1+\varepsilon)\log n + O(1)$. So $KP(x)$ does not exceed $(1+\varepsilon)|x| + O(1)$. The convergent series

$$\sum \frac{1}{n(\log n)^{1+\varepsilon}}, \sum \frac{1}{n \log n (\log\log n)^{1+\varepsilon}} \dots$$

provide the upper bounds

$$KP(x) \le |x| + (1+\varepsilon)\log|x|, \qquad KP(x) \le |x| + \log|x| + (1+\varepsilon)\log\log|x|,$$

etc.

Now for the lower bounds. All of them are *weak* lower bounds, i.e., lower bounds valid for infinitely many arguments but not necessarily for all the arguments. Assume, for instance, that the (weak) lower bound

$$KP(y) \ge |y| + \log|y| \qquad \text{for infinitely many } y\text{'s}$$

is *not* valid. Then for all y (except a finite number of y's) we have

$$KP(y) < |y| + \log|y|$$

and, therefore,

$$2^{-KP(y)} > 2^{-(|y|+\log|y|)}.$$

Summing over all y's, we see that the left-hand side series converge (see Section 3.4); therefore, the right-hand side series should converge also. However, recalling that a binary word y is identified with its number n (which is of the same order as $2^{|y|}$) we recognize the series

$$\sum \frac{1}{n \log n}$$

in the right-hand side.

Similar arguments can be used to prove stronger lower bounds:

$$KP(y) \ge |y| + \log|y| + \log\log|y|,$$
$$KP(y) \ge |y| + \log|y| + \log\log|y| + \log\log\log|y|,$$

etc. (valid for infinitely many y's).

The upper bound for $KP(x)$ can be explained also in a more explicit way. The description mode "each binary word is a description of itself" is valid for simple Kolmogorov entropy (or monotonic entropy, or decision entropy) but is not valid for prefix entropy (i.e., KP), because the description mode in this case should be prefix-free: the descriptions of different objects should not be prefixes of each other. We can obtain a prefix-free description if we consider the word

$$\overline{\text{binary representation of } |x|}01x$$

as a description of x. Here $\overline{z}$ denotes the word z where each letter is repeated twice. This encoding is prefix-free, because the position of the 01-group is determined uniquely, and therefore we may reconstruct the length of x. This encoding leads to an upper bound

$$KP(x) \leq |x| + 2\log|x| + O(1)$$

and we can repeat the trick: the encoding

$$\overline{\text{b.r. of }\left|\text{b.r. of }|x|\right|}\,01(\text{b.r. of }|x|)x$$

(b.r. stands for "binary representation") leads to an upper bound

$$KP(x) \leq |x| + \log|x| + 2\log\log|x| + O(1).$$

This trick can be iterated.

4.3. *Differences Between Entropies*

The similar (though a little more subtle) considerations allow us to establish bounds for differences of entropies (stated in Section 2.2 of [23]).

4.3.1. $KP - KD$: *Upper Bound.* We start with the bound $KP(y) - KD(y)$. Assume that $\sum q_n$ in one of the convergent series mentioned above. We should prove that

$$KP(y) \leq KD(y) + \log|y| + (-\log q_{|y|}) + O(1)$$

or, in other words (recall the encoding-free definition of KP in Section 3.4), that the series

$$\sum 2^{-KD(y)} \cdot \frac{1}{|y|} \cdot q_{|y|}$$

converges. We classify all y according to two integer parameters: its length n and its KD-entropy k. It is easy to see that the number of y's of length n and entropy k does not exceed 2^k; each of them contributes

$$2^{-k} \cdot \frac{1}{n} \cdot q_n$$

to the sum; so all n-k elements contribute at most

$$\frac{1}{n} \cdot q_n$$

(for any k). Now we sum over n and k; summing over k we consider only k not exceeding $n + O(1)$ (because $KD(y) \leq |y| + O(1)$), therefore, summing over k means multiplying by $n + O(1)$ and the sum does not exceed $O(1) \cdot q_n$. It remains to recall that $\sum q_n < \infty$.

4.3.2. $KP - KD$: *Lower Bound.* The corresponding lower bound states that if $\sum q_n$ is one of the divergent series mentioned above, then

$$KP(y) > KD(y) + \log|y| + (-\log q_{|y|}) + O(1).$$

for infinitely many n.

To prove it, it is enough to prove that the series

$$\sum 2^{-KD(y)} \cdot \frac{1}{|y|} \cdot q_{|y|}$$

diverges. Consider the decision description mode where x is a description of all words $x10\ldots0$. Consider the set $A_{n,k}$ of all words of length n having this form for some x of length k (assuming that $k < n$). All words from $A_{n,k}$ have decision complexity not exceeding k; the total number of words in $A_{n,k}$ is 2^k. They contribute to the sum at least

$$2^k \cdot 2^{-k} \cdot \frac{1}{n} \cdot q_n = \frac{q_n}{n};$$

summing over k first, we get the sum $\sum q_n = +\infty$.

4.3.3. *$KS - KA$, $KS - KM$: Upper Bounds.* Now we consider another difference (see paragraph (2) on p. 100 of [23]) and prove that

$$KS(y) - KA(y) \leq \log|y| + O(1)$$

(all logarithms are binary logarithms). In other words, we should prove that

$$KS(y) \leq KA(y) + \log|y| + O(1).$$

According to the encoding-free definition of KS (Section 3.1) it is enough to show that the set

$$Y = \{y | KA(y) + \log|y| < n\}$$

contains $O(2^n)$ elements: $\#Y = O(2^n)$. The set Y is prefix-closed (all prefixes of an element of Y belong to Y too); in other words, Y is a subtree of the complete binary tree. We consider the set Y' of all leaves of this subtree, i.e., all maximal elements of Y (having no continuations in Y). Each element of Y is a prefix of some maximal element, and it is easy to see that

$$\#Y \leq \sum_{y \in Y'} |y|$$

(each element y has $|y|$ prefixes). For any element $y \in Y'$ we have

$$KA(y) + \log|y| < n,$$

or

$$KA(y) < n - \log|y|,$$

or

$$2^{-KA(y)} > \frac{|y|}{2^n}.$$

All elements $y \in Y'$ are incomparable, therefore

$$\sum_{y \in Y'} 2^{-KA(y)} < O(1)$$

and, consequently,

$$\sum_{y\in Y'} \frac{|y|}{2^n} < O(1)$$

and we get the upper bound for $\sum |y|$ that we need.

4.3.4. $KS - KA$, $KS - KM$: *Lower Bounds*. To obtain the matching (weak) lower bound, consider the sequence 0^n (n zeros). We have

$$KA(0^n) = O(1), \qquad KM(0^n) = O(1), \quad \text{and} \quad KS(0^n) = KS(n) + O(1)$$

(we identify n and the nth binary word as before). It remains to prove that $KS(n) \geq \log_2 n$ for infinitely many n which could be done by an easy counting argument (see above).

4.3.5. $KA - KS$, $KM - KS$, $KP - KS$: *Upper Bounds*. Next differences (see paragraph (3) on p. 100 of [23]) are $KM(y) - KS(y)$ and $KA(y) - KS(y)$. The upper bounds follow from the following upper bound for $KP(y) - KS(y)$ (mentioned on p. 101 of [23]): assume that $\sum q_n$ is any of the convergent series considered above; then

$$KP(y) \leq KS(y) + (-\log q_{|y|}).$$

According to the encoding-free definition of KP (Section 3.4), we should prove that

$$\sum 2^{-KS(y)} q_{|y|} < \infty.$$

We consider all terms with $KS(y) = k$; the number of such terms is about 2^k, each term is $2^{-k} q_{|y|}$. We may replace $q_{|y|}$ by q_k because q_i is monotone and because $k = KS(y)$ does not exceed $|y|$ (up to a constant, as usual). Then we get the sum $\sum q_k$ which is finite by our assumption.

4.3.6. $KA - KS$, $KM - KS$, $KP - KS$: *Lower Bounds*. To get the complementary lower bound for $KA(y) - KS(y)$ we start with the bound for $KP(y) - KS(y)$ (it is easier, because $KA \leq KP$). Assume that $\sum q_i$ is any of the divergent series mentioned above. We prove that

$$KP(y) - KS(y) \geq -\log q_{|y|}$$

for infinitely many y. Indeed, $KS(y) \leq |y|$ (we ignore $O(1)$ terms) and, as we have seen before,

$$KP(y) \geq |y| + \log q_{|y|}$$

for infinitely many y. Now we show how to transform a lower bound for $KP - KS$ into a lower bound for $KA - KS$. For any binary word x consider the binary word $t(x) = \hat{x}01$. All words $t(x)$ are incomparable. It is easy to show that $KM(t(x)) = KA(t(x)) = KP(t(x))$ (up to $O(1)$ terms). Indeed, these words $t(x)$ form a "bunch embedded into a tree" (see Section 3.4). It is also easy to see that $KS(t(x)) = KS(x)$. Now the lower bound for $KP - KS$ can be rewritten as

$$KA(t(y)) - KS(t(y)) \geq -\log q_{|t(y)|}$$

and it remains to mention that $t(y)$ is only twice as long as x so it does not matter whether we have $q_{|t(y)|}$ or q_y under the logarithm.

4.3.7. *$KM - KD$, $KA - KD$: Upper Bounds.* Now we prove the upper bound for $KM(y) - KD(y)$ (and therefore for $KA(y) - KD(y)$). When defining $KD(y)$ we used an optimal description mode G which may be considered as a "computable mapping" of type $\mathbb{N} \to \Xi$ in the sense of the Scott–Ershov domain theory (here Ξ is a tree, i.e., a domain where binary words are ordered by a relation "to be a prefix," and $\mathbb{N}$ is a bunch, i.e., a domain where all binary words lie on the same level). Now an optimal prefix description mode F (corresponding to the prefix entropy KP) may also be considered as a "computable mapping" of type $\Xi \to \mathbb{N}$. So we get a diagram

$$\Xi \xrightarrow{F} \mathbb{N} \xrightarrow{G} \Xi$$

with two description modes. Their composition H is a mapping of type $\Xi \to \Xi$ and is a monotone description mode, or, if you do not like references to domain theory, just consider a set

$$H = \{\langle x, z\rangle \mid \exists y(\langle x, y\rangle \in F \text{ and } \langle y, z\rangle \in G\}.$$

Therefore, the KM-entropy of some $y \in \Xi$ does not exceed the KP-entropy of the shortest G-description z of an object y:

$$KM(y) \le KP(z) + O(1) \quad \text{and} \quad |z| = KD(y).$$

Now the inequality for the prefix entropy, e.g.,

$$KP(z) \le |z| + \log|z| + (1+\varepsilon)\log\log|z| + O(1),$$

can be applied to get

$$\begin{aligned} KM(y) &\le |z| + \log|z| + (1+\varepsilon)\log\log|z| + O(1) \\ &= KD(y) + \log KD(y) + (1+\varepsilon)\log\log KD(y) + O(1) \\ &\le KD(y) + \log|y| + (1+\varepsilon)\log\log|y| + O(1) \end{aligned}$$

(the last step uses that $KD(y)$ does not exceed $|y|$). More elaborate inequalities for prefix entropy may be used in the same way, and we get

$$\begin{aligned} KM(y) &\le KD(y) + \log|y| + \log\log|y| + (1+\varepsilon)\log\log\log|y| + O(1), \\ KM(y) &\le KD(y) + \log|y| + \log\log|y| + \log\log\log|y| \\ &\qquad + (1+\varepsilon)\log\log\log\log|y| + O(1), \end{aligned}$$

etc.

Remark. Replacing in the diagram above, the rightmost space Ξ by $\mathbb{N}$ we get the upper bound for the difference $KP(y) - KS(y)$ that we have already proved.

4.3.8. *$KM - KD$, $KA - KD$: Lower Bounds.* The lower bound for $KA - KD$ (and therefore for $KM - KD$) can be obtained from the lower bound for $KP - KS$ mentioned above. Indeed,

$$KP(y) - KS(y) = KA(t(y)) - KD(t(y)) + O(1)$$

(here t is an embedding of the bunch into a tree explained above). In this way we obtain (weak) lower bounds like

$$KA(y) > KD(y) + \log|y|,$$
$$KA(y) > KD(y) + \log|y| + \log\log|y|,$$

etc.

4.3.9. $KS - KD$: *Upper Bound.* Assume that a decision description mode F is used to define KD. Construct a simple description mode G as follows: if x is an F-description of y, then

$$\overline{\text{binary representation of } |y|}01x$$

is a G-description of y. Therefore,

$$KS(y) \leq KD(y) + 2\log|y| + O(1).$$

Iterating the trick (using the binary representation of the length of the binary representation of y, etc.) we get stronger inequalities of that sort:

$$KS(y) \leq KD(y) + \log|y| + 2\log\log|y| + O(1),$$
$$KS(y) \leq KD(y) + \log|y| + \log\log|y| + 2\log\log\log|y| + O(1),$$

etc.

4.3.10. $KS - KD$: *Lower Bound.* We prove that

$$KS(y) \geq KD(y) + \log|y| + \log\log|y|$$

for infinitely many y's (the proof of the lower bound with more logarithms is similar). As usual, assume that it is *not* valid, i.e., that

$$KS(y) < KD(y) + \log|y| + \log\log|y|$$

for almost all y. We take y's of the form $x10^{j-1}$ and get

$$KS(x10^{j-1}) < |x| + \log(|x| + j) + \log\log(|x| + j).$$

Now we should count all pairs $\langle x, j\rangle$ where the right-hand side does not exceed some n and see that the number of such pairs is *not* $O(2^n)$. (This would be a contradiction, because different pairs correspond to different words.) We restrict ourselves to x and j such that

$$|x| \leq n \quad \text{and} \quad n \leq j \leq \frac{2^n}{n^2}.$$

In this case we may replace $\log(|x| + j)$ by $\log j$ (ignoring an additive constant) and obtain a sum

$$\sum_{j=n}^{2^n/n^2} \#\{x \mid \log j + \log\log j + |x| \leq n\} \approx \sum_{j=n}^{2^n/n^2} 2^{n-\log j-\log\log j} \approx 2^n \int_n^{2^n/n^2} \frac{dj}{j\log j};$$

the integral tends to infinity when $n \to \infty$.

4.3.11. *$KP-KA$, $KP-KM$: Upper Bounds.* Assume that q_n is one of the convergent series considered above. We prove that

$$KP(y) \le KA(y) + (-\log_2 q_{|y|}).$$

According to the encoding-free definition of KP (Section 3.4), it is enough to prove that

$$2^{-KA(y)+\log_2 q_{|y|}} = q_{|y|}2^{-KA(y)}$$

is finite. Indeed, if we consider the sum over all y's of a given length n, we get $q_n \cdot O(1)$ (these y's are incomparable), and the series $\sum q_n$ is convergent.

The upper bound for $KP - KM$ follows from the upper bound for $KP - KA$ because KM is bigger that KA.

4.3.12. *$KP-KA$, $KP-KM$: Lower Bounds.* The (weak) lower bound for $KP-KA$ is a consequence of the lower bound for $KP - KM$ which in its turn is a consequence of the lower bound for $KP(y) - |y|$ because $KM(y) \le |y| + O(1)$. The lower bound for $KP(y) - |y|$ is established in Section 4.2.

4.3.13. *$KM - KA$: Upper and Lower Bounds.* This difference is of special interest. The very fact that these entropies differ by more than a bounded additive term is disappointing. This fact was discovered by Gács [9]. (The Hungarian surname "Gács" is pronounced approximately as English "garch.") In his paper he considered sequences of natural numbers instead of binary words, and the bounds become much weaker if we restrict ourselves to binary words. As he writes: "Therefore for binary strings, the lower bound obtainable from the proof of Theorem 1.1 is only the inverse of some version of Ackermann's function" [9, p. 75]. As is known, Ackermann's function is a function with natural arguments and values growing faster than any primitive recursive function. Its inverse f^{-1} (defined as $f^{-1}(a) = \min\{z\colon\ f(z) \ge a\}$), therefore grows extremely slowly. Gász's proof is rather technical. Here is a quotation from [11]:

> **Formulation.** For any function $\varphi(\ )$ let us define $h(j, t, \varphi)$ by the following recursion:
>
> $$h(0, t, \varphi) = t,$$
> $$h(j + 1, t, \varphi) = \varphi(h(j, t, \varphi)).$$
>
> Thus, $h(j, t, \varphi)$ is essentially the j-fold iteration of φ. Now we define
>
> $$t^k(i, r) = \lceil 2^r(2^{-k-2}\log i + 8)\rceil,$$
> $$f^k(0, r) = r,$$
> $$f^k(i + 1, r) = h(t^k(i, r), \lceil\log t^k(i, r)\rceil, \lambda s f^k(i, s)).$$
>
> Let
>
> $$L(k) = 2^{2^{k+7}}, \qquad F(k) = L(k)\log f^k(L(k), 3).$$
>
> Then, for large enough n, there is a binary string x of length $\le n$ with
>
> $$KM(x) - KA(x) \ge F^{-1}(n)/2.$$

(In the last line notation is changed because Gács uses another notation: his Km is our KM, his KM is our KA.)

As to upper bounds, the authors know nothing except the trivial consequences of bounds for $KM - KD$ or $KP - KA$. The gap between upper and lower bounds, therefore, is rather big, and it may be interesting to find tighter bounds.

Acknowledgments

The authors thank N. K. Vereshchagin who explained some of the upper and lower bounds proofs to them and J. Tromp who read the papers very thoroughly and made helpful comments and suggestions; however, the authors should be blamed for all errors and omissions. The authors thank CWI (Amsterdam) and LaBRI (Bordeaux) for hospitality.

References

[1] C. Calude, *Information and Randomness. An Algorithmic Perspective*. Springer-Verlag, Auckland, in press.

[2] G. J. Chaitin, On the length of programs for computing finite binary sequences, *J. Assoc. Comput. Mach.*, **13** (1966), 547–569.

[3] G. J. Chaitin, On the length of programs for computing finite binary sequences: statistical considerations, *J. Assoc. Comput. Mach.*, **16** (1969), 145–159.

[4] G. J. Chaitin, A theory of program size formally identical to information theory, *J. Assoc. Comput. Mach.*, **22** (1975), 329–340.

[5] G. J. Chaitin, *Algorithmic Information Theory*, Cambridge University Press, Cambridge, 1987.

[6] G. J. Chaitin, *Information, Randomness and Incompleteness. Papers on Algorithmic Information Theory*, World Scientific, Singapore, 1987; expanded second edition in 1992.

[7] G. J. Chaitin, Foreword to [1].

[8] P. Gács, On the symmetry of algorithmic information, *Soviet Math. Dokl.*, **15** (1974), 1477–1480. (Translated from the Russian version.)

[9] P. Gács, On the relation between descriptional complexity and algorithmic probability. *Theoret. Comput. Sci.*, **22** (1983), 71–93.

[10] P. Gács, A review of [5], *J. Symbolic Logic*, **54**(2) (1989), 624–627.

[11] P. Gács, Personal communication, April, 1993.

[12] A. N. Kolmogorov, Three approaches to the quantitative definition of information, *Problems Inform. Transmission*, **1**(1) (1965), 1–7. (Translated from the Russian version.)

[13] L. A. Levin, On the notion of a random sequence, *Soviet Math. Dokl.*, **14** (1973), 1413–1416. (Translated from the Russian version.)

[14] L. A. Levin, Laws of information conservation (non-growth) and aspects of the foundation of probability theory, *Problems Inform. Transmission*, **10**(3) (1974), 206–210. (Translated from the Russian version.)

[15] L. A. Levin, Various measures of complexity for finite objects (axiomatic description), *Soviet Math. Dokl.*, **17** (1976), 522–526. (Translated from the Russian version.)

[16] M. Li and P. Vitányi, *An Introduction to Kolmogorov Complexity and Its Applications*, Springer-Verlag, New York, 1993.

[17] D. W. Loveland, A variant of the Kolmogorov concept of complexity, *Inform. and Control*, **15** (1969), 602–619.

[18] A. A. Markov, On normal algorithms which compute Boolean functions, *Soviet Math. Dokl.*, **5** (1964), 922–924. (Translated from the Russian version.)

[19] C. P. Schnorr, Process complexity and effective random tests, *J. Comput. System Sci.*, **7** (1973), 376–388.

[20] C. P. Schnorr, A survey of the theory of random sequences. In R. E. Butts and J. Hintikka (eds.), *Basic Problems in Methodology and Linguistics*, Reidel, Dordrecht, 1977, pp. 193–210.

[21] A. Kh. Shen, Algorithmic variants of the notion of entropy, *Soviet Math. Dokl.*, **29**(3) (1984), 569–573. (Translated from the Russian version.)

[22] R. Solomonoff. A formal theory of inductive inference, Part I, *Inform. and Control*, **7**(1964), 1–22.

[23] V. A. Uspensky, Complexity and entropy: an introduction to the theory of Kolmogorov complexity. In O. Watanabe (ed.), *Kolmogorov Complexity and Computational Complexity*, Springer-Verlag, New York, 1992.

[24] V. A. Uspensky, A. L. Semenov, and A. Kh. Shen, Can an individual sequence of zeros and ones be random?, *Russian Math. Surveys*, **45**(1) (1990), 121–189. (Translated from the Russian version.)

[25] V. V. V'yugin. Algorithmic entropy (complexity) of finite objects and its application to defining randomness and amount of information, *Semiotika i Informatika*, **16** (1981), 14–43. (English translation: *Selecta Mathematica formerly Sovietica*, **13**(4) (1994), 357–389.)

[26] A. K. Zvonkin and L. A. Levin, The complexity of finite objects and the developments of the concepts of information and randomness by means of the theory of algorithms, *Russian Math. Surveys*, **25**(6) (1970), 83–124. (Translated from the Russian version.)

Received April 20, 1993, *and in final form June* 9, 1994.