Relations between varieties of Kolmogorov complexities

V.A. Uspensky, A. Shen

Computer Science/Department of Algorithmics and Architecture

Report CS-R9329 April 1993

CWI is the National Research Institute for Mathematics and Computer Science. CWI is part of the Stichting Mathematisch Centrum (SMC), the Dutch foundation for promotion of mathematics and computer science and their applications. SMC is sponsored by the Netherlands Organization for Scientific Research (NWO). CWI is a member of ERCIM, the European Research Consortium for Informatics and Mathematics.

P.O. Box 4079, 1009 AB Amsterdam (NL)
Kruislaan 413, 1098 SJ Amsterdam (NL)
Telephone +31 20 592 9333
Telefax +31 20 592 4199

Relations between Varieties of Kolmogorov Complexities

Vladimir A. Uspensky
uspensky@int.glas.apc.org

Alexander Shen
shen@sch57.msk.su

CWI
P.O. Box 4079, 1009 AB Amsterdam, The Netherlands

and

Dept. of Mathematical Logic and the Theory of Algorithms
Faculty of Mechanics and Mathematics
Moscow Lomonosov University
V-234 Moscow, 119899 Russia

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 O. Watanabe's book [5], and which is included as a supplement of this report, with the kind permission of Springer-Verlag. Upper and lower bounds for those complexities and also for their differences were announced in that paper without proofs. The purpose of this paper is to give proofs for those bounds.

In this paper, the word "entropy" (not in a physical sense) is used instead of "complexity". This is a Moscow tradition suggested by A. N. 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).

1991 Mathematics Subject Classification: 68Q30
1991 CR Categories: F.2, F.4
Keywords and Phrases: Kolmogorov complexity, Entropy, Algorithmic information theory, Simple entropy (complexity), Decision entropy (complexity), Monotonic entropy (complexity), Prefix entropy (complexity), A priori entropy (complexity)
Note: This work was supported by the Netherlands Organization for Scientific Research (NWO), and sponsored by the Committee for Cooperation with Eastern Europe.

Report CS-R9329
ISSN 0169-118X
CWI
P.O. Box 4079, 1009 AB Amsterdam, The Netherlands

Table of Contents

1. Introduction

This paper is a supplement to the paper [5]. Here the proofs of some results stated there are given. All these proofs are well known; they are collected here for reader's convenience and adapted to the terminology and notation used in [5]. We assume that reader has a copy of [5].

The paper is organized as follows. We start (sect. 2) with the classification of four entropies (two possibilities for objects combined with two possibilities for descriptions) which goes back to [4] and is explained in sections 1.2 and 1.3 of [5]. Looking closely on each entry of the table of four entropies, we show that definitions of sections 1.2 and 1.3 coincide with the classical definitions of the corresponding entropies.

Then in sect. 3 we look at a different classification of entropies which goes back to [2] and establish the connections between these two classifications mentioned in section 1.6 of [5]

Finally in sect. 4 we establish some connections between different entropies mentioned in sect. 2.1 and 2.2 of [5].

2. Objects and Descriptions

2.1 Simple Kolmogorov Entropy

This entropy is called (=,=)-entropy or $\mathbb{NN}$-entropy in [5], sect. 1.2. When defining this entropy, a *mode of description* is a binary relation $E \subset \Xi \times \Xi$ (here Ξ denotes the set of all binary words) such that for every x_1, x_2, y_1, y_2 in Ξ

$$\langle x_1, y_1\rangle \in E \wedge \langle x_2, y_2\rangle \in E \wedge x_1 = x_2 \Rightarrow y_1 = y_2.$$

In other terms, mode of description is a (partial) function from Ξ into Ξ. Recursively enumerable modes of descriptions correspond to computable functions.

Now look at the definition from sect. 1.3 of [5]. The ordering on the bunch $\mathbb{B}$ is trivial (only equal objects are comparable), therefore conditions 1 and 2 ([5], p. 89) are always satisfied. The condition 3 means that E is a graph of a function, and acceptable modes of descriptions are graphs of computable functions. Therefore, this definition coincides with that of sect. 1.2 (and in fact with the Kolmogorov definition from [1]).

To prove the Solomonoff—Kolmogorov theorem in this case means to construct an optimal mode of description. Assume that $U(x, y)$ is an universal computable function (i.e., the family $\{U_x\}$, where $U_x(y) = U(x, y)$, contains all computable functions). By $\widehat{x}$ we denote the word x where each letter is repeated twice. An optimal mode of description may be constructed as follows:

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

2.2 Decision Entropy

Now let us look at the $(=, \gamma)$-entropy, or $\mathbb{N}\Xi$-entropy (according to the notation of [5], sect. 1.2) The requirement of sect. 1.2 says that if

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

then one of the words y_1 and y_2 is a prefix of another one. Therefore, for any fixed x all y such that $\langle x, y\rangle \in E$ are prefixes of some finite or infinite binary string.

The requirements of sect. 1.3 of [5] (for $X = \mathbb{B}$, $Y = \mathbb{T}$) are slightly different: in particular, they say that if $\langle x, y\rangle \in E$ then $\langle x, y'\rangle \in E$ for all prefixes y' of y. Therefore, for any fixed x all y such that $\langle x, y\rangle \in E$ form the set of *all* prefixes of some finite or infinite binary string.

The definition of sect. 1.2 gives a broader class of description modes and, theoretically speaking, may lead to a smaller entropy. However, for any binary relation E satisfying the requirements of sect. 1.2 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 recursive enumerable if E is, that E' satisfies the requirements of sect. 1.3 and the corresponding complexity function does not exceed the complexity function corresponding to E. Therefore, the definitions of sections 1.2 and 1.3 give the same entropy (i.e., difference between entropies is bounded).

It remains to say why there exists an optimal description mode in this case. Assume that $U(x, y, n)$ (where x, y 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

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

(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 [3] or [8].)

2.3 Monotonic Entropy

This is a name for (γ, γ)-entropy, or $\Xi\Xi$-entropy (according to the notation of [5], sect. 1.2) The requirement of sect. 1.2 says 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.

The requirements of sect. 1.3 of [5] (for $X = \mathbb{T}$, $Y = \mathbb{T}$) are slightly different:

- if $\langle x, y\rangle \in E$ then $\langle x, y'\rangle \in E$ for all prefixes y' of y;
- if $\langle x, y\rangle \in E$ then $\langle x', y\rangle \in E$ for all x' having x as a prefix;
- 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.

It is easy to check that the requirements of sect. 1.2 are consequences of the requirements of sect. 1.3 (we may replace x_1 and x_2 by the longest of them), but not vice versa. However, if E satisfies the requirements of sect. 1.2, then its extension E' defined as

$$E' = \{\langle x, y\rangle | \text{there are } x' \le x \text{ and } y' \ge y \text{ such that } \langle x', y'\rangle \in E\ \}$$

(here $p \le q$ means that a binary word p is a prefix of a binary word q) satisfies the requirements of sect. 1.3. Therefore, the entropies coincide (see the previous section).

It remains to show that there exists an optimal description mode. It 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 [4]; a very detailed description of what happens for the case of monotonic entropy, is given in [6].

The universal "simple" machine of the previous section gives an optimal "simple" mode of description, i.e. an optimal mode for the case of simple entropy. Similar to that, a universal "monotonic" machine can be constructed which will give an optimal "monotonic" mode of description, i.e. an optimal mode for the case of monotonic entropy. In fact the new machine behaves very much like the previous one, finding from the start of its input the encoding of an arbitrary monotonic machine, and simulating that on the remainder of the input. The only difference is that monotonic machines have no blanks in their tape alphabet and hence lack a notion of the "end of input".

2.4 Prefix Entropy

This is a name for $(\gamma, =)$-entropy, or $\Xi\mathbb{N}$-entropy (according to the notation of [5], sect. 1.2) In this case (as well as in all other) it is easy to show that entropies defined as in sections 1.2 and 1.3 of [5] coincide. In fact the definitions of sections 1.2 and 1.3 are different versions of the same definition; the really different (encoding-free) definitions are given in sect. 1.5.

The requirements of sect. 1.2 say that 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 $y_1 = y_2$. The requirements of sect. 1.3 of [5] (for $X = \mathbb{T}$, $Y = \mathbb{B}$) are slightly different:

- 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';
- if $\langle x, y_1\rangle \in E$ and $\langle x, y_2\rangle \in E$ then $y_1 = y_2$.

It is easy to see that the requirements of sect. 1.2 follow from the requirements of sect. 1.3. Moreover, if some E satisfies the requirements of 1.2, 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 the requirements of sect. 1.3. Therefore, the corresponding entropies coincide.

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 [4]) and also can be proved directly. We omit this proof because the existence of an optimal mode is a byproduct of the coincidence of the definition given above and the encoding-free definition (see the next section).

Remark. The prefix entropy relates to the modes of description fulfilling the $(\gamma, =)$-property of [5], which means that if two words describe the same object then neither of them is the beginning of the other. Thus, any mode with that property can be called prefix-free. It is easy to create a family $\pi_0, \pi_1, \ldots$ of straightforward prefix-free encodings (i.e. modes of description) of increasing efficiency. The simplest encoding, π_0, is the unary one: it maps a string x to the string 1^i0 where i is the index of x in the lexicographic ordering of all strings. The encoding π_{n+1} maps x to $\pi_n(u)x$ where the index of u in the lexicographic ordering is the length $|x|$ of x. A variation of π_1 could be seen in the construction of an optimal mode in Section 2.1.

3. Encoding-free definitions

3.1 Simple Kolmogorov entropy

The simple Kolmogorov entropy ($\mathbb{NN}$-entropy) can be characterized as a minimal (up to a constant) enumerable from above function $f : \Xi \to \mathbb{N} \cup \{\infty\}$ satisfying the following condition ($\mathbf{C}\mathbb{B}$) from sect. 1.5 of [5] (we give an equivalent form of it):

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

Remark. If we replace 2^n by $C \cdot 2^n$ (see the condition ($\mathbf{C}'$) in sect. 1.5) 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 also may replace $=$ in ($\mathbf{C}$) by $\leq$: if there is at most 2^n objects y such that $f(y) = n$ then the number of objects y such that $f(y) \leq n$ does not exceed $1 + 2 + \cdots + 2^n < 2 \cdot 2^n$.

To prove the coincidence of entropies we should prove that

- a simple Kolmogorov entropy function $KS(x)$ corresponding to $\mathbb{NN}$-entropy (description modes are all computable functions) satisfies the condition ($\mathbf{C}\mathbb{B}$);
- for any enumerable from above function f satisfying ($\mathbf{C}\mathbb{B}$) one can construct a description mode E such that the complexity function corresponding to E exceeds f not more than by 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 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 can not 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 description mode. Evidently, the corresponding complexity function does not exceed $f + 1$.

A byproduct of this argument is the existence of minimal (up to an additive constant) enumerable from above function satisfying the ($\mathbf{C}\mathbb{B}$) condition.

3.2 Decision Entropy

To get the decision entropy we should use the condition ($\mathbf{C}\mathbb{T}$) (see fig. 2 in [5]). It may be reformulated as follows (for a function $f : \Xi \to \mathbb{N} \cup \{\infty\}$):

- if M is a finite set of incomparable words (there is no word in M which is a prefix of another word in M) then the cardinality of M does not exceed 2^n

As in the previous section, to prove the coincidence of entropies we should prove that

- a decision entropy function $KD(x)$ corresponding to $\mathbb{N}\Xi$-entropy satisfies the condition ($\mathbf{C}\mathbb{T}$);

- for any enumerable from above function f satisfying ($\mathbf{C\mathbb{T}}$) one can construct a description mode E such that the complexity function corresponding to E exceeds f not more than by a constant.

Let us start with the first claim. Assume that M is a set of incomparable words (no one is a prefix of another one) having complexity n. That means that all these words have descriptions of length n. All these descriptions must be different (otherwise the conditions of sect. 1.2 of [5] 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 having complexities less than n contains not 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 object having e as a description. (All other object in this set will be prefixes of the longest one.) If any of these objects $z(e)$ is comparable with y than 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 enumerable and satisfies the conditions of sect. 1.2 of [5]. Evidently, the corresponding complexity function does not exceed $f + 1$.

3.3 Monotonic Entropy and a priori *probability*

Here we get two different entropies: a monotonic entropy ($\Xi\Xi$-entropy, sect. 1.2 of [5]) and *a priori* entropy ($\Sigma\mathbb{T}$-entropy, sect. 1.5 of [5]). They differ. The second entropy appeared in [8] as a logarithm of a so-called *a priori* probability; this original definition is discussed in details in [6]. We do not reproduce this discussion here; the only thing we want to explain is why the original definition of [8] coincides with that of sect. 1.5 of [5].

A *semimeasure* is a function m defined on Ξ with non-negative 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 exists a maximal (up to a constant factor) enumerable from below semimeasure $M(x)$ called *a priori probability* (see [6]). Its logarithm $-\log_2 M(x)$ coincides (up to an additive constant) with the $\Sigma\mathbb{T}$-entropy of sect. 1.5 of [5]. Let us explain shortly why this coincidence takes place. The main role is played by the following two facts:

- if $m(x)$ is a semimeasure then $\lceil -\log_2 m(x)\rceil$ 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 sets D of incomparable words such that x is a prefix of all words from 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 a correspondence between semimeasures and functions satisfying the condition $\Sigma\mathbb{T}$ which more or less preserves enumerability (we omit some details) and allows us to prove the coincidence mentioned above.

There is one more assertion concerning $\Sigma\mathbb{T}$-entropy called Muchnik's theorem on p. 93 of [5]. 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 sum of labels exceeding C. Then there exists an infinite set of pairwise incomparable words with the infinite sum of labels.

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. This supremum (finite or infinite) depends on x. Let us call a vertex *bad* if it is infinite. Now our task is as follows: the root of a tree is bad; find an infinite set of pairwise incomparable words with infinite sum of labels. Bad vertices form a subtree in the full binary tree; this subtree has no leaves (if x is bad, ar least one of the words $x0$ and $x1$ is bad). Now we shall 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 easy to find the required infinite set of vertices with infinite sum of labels.

3.4 Prefix Entropy

The prefix entropy ($\Xi\mathbb{N}$-entropy) with its encoding-free definition (using the ($\Sigma\mathbb{B}$)-condition is probably the most technically interesting among all the four entropies. It is discussed in details in [7]; however, this paper is not translated into English, so we shall try to give a self-contained description of what happened in this case.

Let us start with the definition of a semimeasure on $\mathbb{B}$. It differs from the definition of a semimeasure on $\mathbb{T}$ used in the previous section. However, from now on we shall consider only semimeasures on $\mathbb{B}$ and call them just "semimeasures" omitting the words "on $\mathbb{B}$".

A *semimeasure* is a (total) function m defined on the set Ξ of all binary words with non-negative 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 described in Remark made on p. 95, sect. 1.7 of [5]. Namely,

- if M is a probabilistic machine that can stop, 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 one can construct a probabilistic machine M such that $m(x) = P_M^s(x)$ for all x.

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

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 kth step. We may assume that for each k the value $m_k(x)$ differs from zero 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.

There exists 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 $m(x) \leq c \cdot M(x)$ for all words x.

This fact can be proved as follows: enumerate all probabilistic machines and construct an "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 ith machine and M is a semimeasure corresponding to the universal machine, then $M(x) \geq p_i \cdot m_i(x)$. Therefore, M is maximal.

The following connections between the definition of a semimeasure and the condition ($\Sigma\mathbb{B}$) are valid:

- if f is a function satisfying the 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 the condition ($\Sigma\mathbb{B}$).

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

Now it remains to show that this minimal function (or logarithm of the maximal semimeasure) coincides with $\Xi\mathbb{N}$-entropy. Here, as usual, we should prove two assertions:

- for any description mode E satisfying the conditions of sect. 1.2 of [5] for the case of $\Xi\mathbb{N}$-entropy, the corresponding complexity function Compl_E satisfies the condition ($\Sigma\mathbb{B}$) and is recursively enumerable from above;
- if a recursively enumerable from above function f satisfies the conditions ($\Sigma\mathbb{B}$) then there exists a description mode E satisfying the conditions of sect. 1.2 of [5] such that the corresponding complexity function Compl_E exceeds f not more than by a constant.

The first assertion is almost trivial. If $M = \{m_1, \ldots, m_k\}$ is a finite set of words and $e_1, \ldots, e_k$ 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, the 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 the level k we have 2^k parts of length 2^{-k} each. Assume the we get a sequence of natural numbers n_1, n_2,... 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 task is possible only if $\sum_i 2^{-n_i} \leq 1$. 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 union of disjoint 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 +... 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 the 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

$$2^{-f(x)-1} + 2^{-f(x)-2} + 2^{-f(x)-3} + \cdots \leq 2^{-f(x)}.$$

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). A 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 requirements of sect. 1.2 of [5] are 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 not more than by 1.

This argument implies also 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 of sect. 1.2 and 1.3 of [5] form a parallelogram (see Fig. 1, [5], p.91). It is easy to see that the restrictions for description modes become weaker when we go down along the sides of this parallelogram: each $\Xi\mathbb{N}$ description mode is $\Xi\Xi$ description mode and at the same time $\mathbb{N}\mathbb{N}$ description mode etc. Therefore, the corresponding entropies decrease when we go down.

The same is true for the other parallelogram (see p.92 of [5]) It is because the condition **C** is a consequence of the condition Σ and because $\mathbb{T}$-incomparable words are $\mathbb{B}$-incomparable.

When we try to combine two parallelograms into a pentagon (Fig. 3, p. 94 of [5]) we should prove that *a priori* entropy does not exceed $\Xi\Xi$-entropy. This is explained in details in [6]; here we give only a short comment. Assume that we have an optimal $\Xi\Xi$ 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.

4.2 Entropies and lengths

These bounds (see [5], p. 99) use the somewhat strange functions $Q_k(n,\varepsilon)$. The reason why these functions appear is the following one. The series

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

diverge; 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. The functions Q_k are used to mark the boundary between convergence and divergence: $Q_k(z,0)$ is "on the divergent side" and $Q_k(z,\varepsilon)$ is "on the convergent side". (To be exact, we should use qlog instead of log everywhere, as in [5], p. 99, to avoid division by zero. We ignore this problem.)

To show how these convergency consideration work let us consider the upper and lower bounds for the prefix entropy KP (inequalities (2) and (3) on p. 99, see [5]). Let us 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

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

converge, therefore it satisfies the condition $\Sigma'\mathbb{B}$ of sect. 1.5 of [5]. Therefore, $\Sigma\mathbb{B}$-entropy of n does not exceed $(1+\varepsilon)\log n$. Recalling that n is a number of some binary word x we see that $KP(x)$ does not exceed $(1+\varepsilon)|x|$. The convergent series

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

gives the upper bound

$$KP(x) \leq |x| + (1+\varepsilon)\log|x|$$

etc. Now the lower bounds. Assume, for instance, that the (weak) lower bound

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

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

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

and, therefore,

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

and the series in the right-hand side of this inequality should converge. 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.

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 $\mathbb{NN}$-entropy (or $\Xi\Xi$-entropy) but is not valid for $\Xi\mathbb{N}$-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

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

and we can repeat the trick: the encoding

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

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

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

This trick can be iterated.

Remark. The above arguments can be rephrased in terms of the prefix-free encodings π_i given in the Remark of Section 2.4: For any i, the length of $\pi_i(x)$ is, up to an additive constant, an upper bound on the prefix complexity $KP(x)$ of x.

We proved the statements about KP made in [5], sect. 2.1; other entropies are much simpler. For example, let us prove that the smallest entropy, $KD(y)$, is greater than y for infinitely many y's. More precisely, we prove that

$$KD(y) \ge |y|$$

for infinitely many y's. 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.

4.3 Differences between entropies
The similar (though a little more subtle) considerations allows us to establish bounds for differences of entropies (see sect. 2.2 of [5]).

$KP - KD$: upper bound Let us start with the bound $KP(y) - KD(y)$. Assume that q_n in one of the convergent series mentioned above. We should prove that

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

or, in other words, that the series

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

converges. Let us 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 shall sum over n and k; summing over k we consider only k not exceeding $n + O(1)$ (because $KD(y) \le |y| + O(1)$), 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 $\sum q_n < \infty$.

$KP - KD$: lower bound To prove the corresponding lower bound we should prove that if the series $\sum q_n$ diverges, then the series

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

diverges also. Consider the $\mathbb{N}\Xi$-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 $\sum q_n = +\infty$.

$KS - KA$, $KS - KM$: upper bounds Now let us consider another difference (see [5], p. 100, paragraph (2)) and prove that

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

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

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

According to the CB-definition, it is enough to show that the set

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

contains $O(2^n)$ elements. 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. Let us 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

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

(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.

$KS-KA$, $KS-KM$: lower bounds To obtain the matching (weak) lower bound, consider the sequences 0^n (n zeros). We have

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

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

$KA-KS$, $KM-KS$, $KP-KS$: upper bounds Next differences (see [5], p. 100, paragraph (3)) are $KM(y) - KS(y)$ and $KA(y) - KS(y)$. The upper bounds follow from the upper bound for $KP(y) - KS(y)$ mentioned in [5], p. 101. Let us prove the latter upper bound. Assume that $\sum q_n$ is any of the convergent series considered above; it is enough to prove that

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

According to the $\Sigma\mathbb{B}$-definition of KP, we should prove that

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

Let us 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.

$KA-KS$, $KM-KS$, $KP-KS$: 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 $KA \le KP$). Assume that $\sum q_i$ is any of the divergent series mentioned above. We prove that

$$KP(y) - KS(y) \ge -\log_2 q_{|y|}$$

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

$$KP(y) \ge |y| + \log_2 q_{|y|}$$

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". It is easy to see also that $KS(t(x)) = KS(x)$. Now the lower bound for $KP - KS$ can be rewritten as

$$KA(t(y)) - KS(t(y)) \ge -\log_2 q_{|t(y)|}$$

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

$KM-KD$, $KA-KD$: upper bounds Now let us prove the upper bound for $KM(y)-KD(y)$ (and therefore for $KA(y) - KD(y)$). When defining $KD(y)$ we use an optimal description mode G which is a "computable mapping" of type $\mathbb{N} \to \Xi$. Consider an optimal $(\Xi \to \mathbb{N})$-description mode F (corresponding to the prefix entropy KP) and a diagram

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

with two description modes. Their "composition" H will be a description mode of type $\Xi \to \Xi$. Therefore, the $\Xi\Xi$-entropy of some $y \in \Xi$ does not exceed the $\Xi\mathbb{N}$-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

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

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

(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 proved already.)

$KM - KD$, $KA - KD$: 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).

$KS - KD$: upper bound Assume that a $\mathbb{N}\Xi$-description mode F is used to define KD. Construct a $\mathbb{N}\mathbb{N}$-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,

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

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

Remark. The family of prefix-free encodings given in the Remark of Section 2.4 provide such stronger inequalities. We have for any i that $KS(y) \le KD(y) + |\pi_i(|y|) + O(1)$.

$KS - KD$: lower bound Let us prove that

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

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

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

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

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

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

$KP-KA$, $KP-KM$: upper bounds Assume that q_n is one of the convergent series considered above. Let us prove that

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

According to the $\Sigma\mathbb{B}$-definition of KP, 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.

$KP-KA$, $KP-KM$: 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 sect. 4.2.

5. Acknowledgements

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, authors should be blamed for all errors and omissions. The authors are also indebted to J. Keller for his generous help in preparing this document.

References

1. A. N. Kolmogorov. *Three approaches to the quantitative definition of information.* Problems Inform. Transmission, 1:1–7, 1965. (Translated from the Russian version.)
2. L. A. Levin. *Various measures of complexity for finite objects (axiomatic description).* Soviet Math. Dokl. 17:522–526, 1976. (Translated from the Russian version.)
3. D. W. Loveland. *A variant of the Kolmogorov concept of complexity.* Information and Control 9:602–619, 1966.
4. A. Kh. Shen. *Algorithmic variants of the notion of entropy.* Soviet Math. Dokl. 29:569–573, 1984. (Translated from the Russian version.)
5. V. A. Uspensky. *Complexity and Entropy: An Introduction to the Theory of Kolmogorov Complexity.* In: *Kolmogorov Complexity and Computational Complexity.* O. Watanabe, ed. Springer-Verlag, 1992. P. 85–102.

6. V. A. Uspensky, A. L. Semenov, 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.)

7. V. V. Vjugin. *Algorithmic entropy (complexity) of finite objects and its application to defining randomness and quantity of information.* Semiotika i Informatika 16:14–43, 1981; in Russian.

8. 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, 83–124, 1970. (Translated from the Russian version.)

Complexity and Entropy: An Introduction to the Theory of Kolmogorov Complexity *

Vladimir A. Uspensky
Department of Mathematical Logic
Faculty of Mechanics and Mathematics
Moscow Lomonosov University
V-234 Moscow, GSP-3, 119899 Russia
uspensky@int.glas.apc.org

Contents

1 Complexity, Entropy, and Randomness

Things can be large or small, and their size (the length or the volume or the weight or so on) can be measured by a number. Besides, things can be simple or complex, and their complexity can also be measured by a number. I do not know to whom we are indebted for measuring sizes by numbers. It was Andrei Kolmogorov [Kol65] who proposed to measure the complexity of a thing by a natural number (i.e., a non-negative integer), and he developed the rudiments of the theory.

Complexity of things (as opposed to the complexity of processes, e.g., of computational processes) took the name *descriptional complexity*, or *Kolmogorov complexity*. As will be seen here, in appropriate cases one may say "entropy" instead of "complexity".

Thus we assume that there is a set Y of things, or objects, y's, and a total function "complexity of y" defined on Y. That function will be denoted by Compl and its possible values are $0, 1, 2, 3, ..., n, ..., \infty$. So the function Compl is a total function from Y to $\mathbb{N} \cup \{\infty\}$. We do not put any further restrictions on Compl, but take it on an intuitive level as a measure of complexity, or a complexity function, or, shorter, a complexity.

Let Compl_1 and Compl_2 be two measures of complexity. Let us say that Compl_1 is *not worse* than Compl_2 if

$$\mathrm{Compl}_1(y) \underset{+}{\leq} \mathrm{Compl}_2(y).$$

Explanation 1. The notation $A(y) \underset{+}{\leq} B(y)$ means that for some constant c not depending on y and for all y, $A(y) \leq B(y) + c$ holds.

Let $\mathcal{Z}$ be some class of complexities, or (that is the same) of measures of complexity. Let Compl_0, belonging to $\mathcal{Z}$, be not worse than any complexity belonging to $\mathcal{Z}$. Then Compl_0 is called *optimal* in the class $\mathcal{Z}$. So a way of measuring complexity is called optimal if it gives, roughly speaking, the shortest complexities of things. Of course, a class of complexities may have no optimal one.

Any optimal complexity is called an *entropy*. It is possible that a class $\mathcal{Z}$ has several entropies, but any two entropies, Ent_1 and Ent_2, fulfill the following condition:

$$\mathrm{Ent}_1(y) \underset{\oplus}{=} \mathrm{Ent}_2(y)$$

Explanation 2. The notation $A(y) \underset{\oplus}{=} B(y)$ means that $|A(y) - B(y)| \underset{\oplus}{<} 0$, or $A(y) \underset{\oplus}{<} B(y)$ *and* $B(y) \underset{\oplus}{<} A(y)$.

Important Remark 1. There is no semantic problem when one speaks about *an* entropy related to some class of complexities. But in the theory of Kolmogorov complexity it is usual to speak about *the* entropy and even to denote it by a special notation. What does it mean? Here we have an *abus de langage* (after N.Bourbaki). Speaking about *the* entropy related to some class, one speaks in fact about *an arbitrary* entropy of that class. And the notation denotes any of such entropies. Of course, our statements must be invariant and do not change their truth value when a particular entropy changes to another one but still belonging to the same class. But we must be cautious. Let $\mathcal{V}$ and $\mathcal{W}$ be two classes of complexity functions, and let K be *the* entropy related to $\mathcal{V}$ and L be *the* entropy related to $\mathcal{W}$. In fact, K and L denote two families of entropies, or, it is better to say, any entropies of those two families. When we write $K(y) \underset{\oplus}{<} L(y)$, we suppose that this relation $\underset{\oplus}{<}$ holds for any particular entropy denoted by K and any particular entropy denoted by L (so there is an additive constant hidden in this relation depending on the choices of particular representatives of K and L). But when we declare that K and L coincide (are *the same* entropy), we do not want to express the opinion that any entropy denoted by K coincides with any entropy denoted by L. That is, we understand the coincidence statement in the following way: for any of the entropies K and any of the entropies L, there exists a constant c such that $|K(y) - L(y)| < c$ for all y.

Terminological Remark 1. In literature on Kolmogorov complexity, the term "complexity" (synonymous with "complexity function" and "measure of complexity") is most often used in the sense of the term "entropy". But we make the distinction between those two terms: entropy is an optimal complexity.

As has already been said, there may be no entropy among complexity functions belonging to a class $\mathcal{Z}$. An important property of a class $\mathcal{Z}$ is that of having an entropy. In such case, we say that the *Solomonoff–Kolmogorov Theorem* holds for $\mathcal{Z}$.

There exist several important classes of complexities that contain entropies. And among those entropies, there are ones of special interest — namely, those entropies that can be used for a definition of randomness. Kolmogorov has proposed the following definition of randomness

for an infinite binary sequence $a_1a_2a_3\cdots a_n\cdots$: the sequence is called *random*, or, more exactly, *Kolmogorov random*, if

$$\mathrm{Ent}(a_1\cdots a_n) \gtrapprox n,$$

where Ent is an entropy. Of course, the choice of Ent is to be specified. Not every sort of entropy goes to a "good" definition of randomness; a definition by Kolmogorov scheme is regarded as "good" if the class of Kolmogorov random sequences sprung up by that definition coincides with the class of sequences that are random in the sense of Martin-Löf (or are *typical sequences*; see [KU87a] and [KU87b] and [Mar66]).

To sum up: in order to define an entropy, one must define an appropriate class of complexities and show that the Solomonoff–Kolmogorov Theorem holds for that class.

1.1 Generation of Complexities by Means of an Encoding Procedure

The idea (due to Kolmogorov) is very simple. There are objects and there are descriptions (encodings) of objects, and the complexity of an object is the minimal size of its description.

In more detail, there is a set Y of objects y, and a set X of descriptions (names, encodings) x. There is a volume function ℓ defined on X; that ℓ is a function from X to $\mathbb{N}$. A *mode of description*, or a *description mode*, is an arbitrary set $E \subseteq X \times Y$. If $\langle x, y\rangle \in E$, then x is called a *description* (a *name*, an *encoding*) *of* y *with respect to* E. Thus an object y may have many descriptions and a description may serve as a description for many objects.

The *complexity of* y *with respect to* a description mode E is defined as follows:

$$\mathrm{Compl}_E(y) = \min\{\ell(x) : \langle x, y\rangle \in E\}.$$

We make the convention that $\mathrm{Compl}_E(y) = \infty$ if there is no x such that $\langle x, y\rangle \in E$.

Let $\mathcal{E}$ be a class of modes of description. Each mode $E \in \mathcal{E}$ gives the corresponding complexity function Compl_E. Then there arises the class $\mathcal{Z} = \mathcal{Z}(\mathcal{E})$ of all complexity functions related to the modes of $\mathcal{E}$, and one may ask whether the class $\mathcal{Z}$ contains an optimal function, or an entropy. If such an optimal function exists, then it corresponds to some description mode which is also called *optimal*.

Until now we have not imposed any restrictions on X, Y, E. It is reasonable to assume that X and Y consist of *constructive objects*, and E is a *generable* set (in the sense of Post) and, consequently, is a recursively *enumerable* set.

In the following exposition we shall restrict ourselves with the following simple case: Both X and Y are Ξ, where Ξ is the set of all binary words, or finite binary sequences. The volume function ℓ is defined to be $\ell(\xi) = |\xi|$ for every $\xi \in \Xi$, where $|\xi|$ is the length of ξ.

1.2 Two Symmetric Relations and Four Entropies

Our task is to define—in a reasonable way—a class $\mathcal{E}$ of modes of description as a class of subsets of $\Xi \times \Xi$. Having this goal in mind, we define a binary relation on Ξ which we shall call *the*

concordance relation: u_1 and u_2 are called *concordant* if they have a mutual continuation, i.e., if there are $t_1, t_2 \in \Xi$ such that $u_1 t_1 = u_2 t_2$. This concordance relation will be denoted by γ.

Thus we have two natural binary relations on Ξ: the equality $=$ and the concordance γ. Both are symmetric and decidable (see Explanation in Sect.1.3).

Let α and β be two binary relation on Ξ. We say that a mode of description $E \subseteq \Xi \times \Xi$ *fulfills the* (α, β)*-property* if for every $x_1, x_2, y_1, y_2 \in \Xi$,

$$\langle x_1, y_1 \rangle \in E \wedge \langle x_2, y_2 \rangle \in E \wedge x_1 \alpha x_2 \Rightarrow y_1 \beta y_2.$$

Let us consider the class $\mathcal{E} = \mathcal{E}(\alpha, \beta)$ of all recursively enumerable modes of description that fulfill the (α, β)-property and the related class $\mathcal{Z}^{\alpha,\beta} = \mathcal{Z}(\mathcal{E})$ of complexity functions. If the class $\mathcal{Z}^{\alpha,\beta}$ contains an optimal complexity function, that complexity function will be called (α, β)*-entropy.*

Now move from variable α and β to constants $=$ and γ. Taking $=$ or γ as α or β, we obtain four classes of complexity functions: $\mathcal{Z}^{=,=}$, $\mathcal{Z}^{=,\gamma}$, $\mathcal{Z}^{\gamma,=}$, $\mathcal{Z}^{\gamma,\gamma}$. For each of these four classes, the Solomonoff–Kolmogorov theorem is valid, so we have four entropies:

1. $(=,=)$-entropy, or $\mathbb{N}\mathbb{N}$-entropy,
2. $(=,\gamma)$-entropy, or $\mathbb{N}\Xi$-entropy,
3. $(\gamma,=)$-entropy, or $\Xi\mathbb{N}$-entropy, and
4. (γ,γ)-entropy, or $\Xi\Xi$-entropy.

Note 1. The notations $\mathbb{N}\mathbb{N}$, etc., have the following origin. In [US81] and [US87], the notation "Ξ" had the following meaning: the set of all binary words being considered together with relation γ. In place of the set of all binary words with the relation $=$ on that set, the set $\mathbb{N}$ of all natural numbers with the relation $=$ and the volume function $l(x) = \lfloor \log_2(x+1) \rfloor$ was considered. This volume function is induced by the following 1-1 correspondence between $\mathbb{N}$ and Ξ: zero $\sim \Lambda$, one ~ 0, two ~ 1, three ~ 00, four ~ 01, five ~ 10, and so on.

1.3 Two Approximation Spaces and Four Entropies

There is another way to come to the four basic entropies of Sect.1.2.

Any set of constructive objects with a decidable partial ordering defined on that set will be called an *approximation space.*

Explanation 3. The term "decidable" means that there is an algorithm to decide for any x' and x'', whether $x' \leq x''$ or not.

On an intuitive level, the elements of an approximation space can be taken as informations, and $x' \leq x''$ means that the information x'' is a refinement of the information x' (and hence x'' is closer than x' to some limit value to which both x' and x'' serve as approximations).

To develop a more attractive theory of approximation spaces, especially with the intention to apply this theory to an advanced theory of Kolmogorov complexity, one needs to include some additional requirements into the definition of an approximation space. For our goals, however, it suffices to have a decidable partial ordering. Moreover, only two approximation spaces will be considered: the bunch $\mathbb{B}$ and the tree $\mathbb{T}$. Their definitions follow immediately.

- The *bunch* $\mathbb{B}$: The set of objects is Ξ, and the partial ordering $\leq$ is $=$, i.e., $u \leq w$ iff $u = w$.

- The *tree* $\mathbb{T}$: The set of objects is Ξ. The partial ordering $\leq$ is defined as follows: $u \leq w$ iff u is a prefix of w (and w is a continuation of u), i.e., $\exists v[\, uv = w\,]$.

Let X and Y be two approximation spaces. The spaces X and Y will be treated, respectively, as the space of descriptions (names, encodings) and as the space of the objects described (named, encoded). Our near goal is to define the class $\mathcal{E}$ of acceptable description modes $E \subseteq X \times Y$.

We impose on E the following three requirements:

1. if $\langle x, y\rangle \in E$ and $x' \geq x$, then $\langle x', y\rangle \in E$,

2. if $\langle x, y\rangle \in E$ and $y' \leq y$, then $\langle x, y'\rangle \in E$, and

3. if $\langle x, y_1\rangle \in E$ and $\langle x, y_2\rangle \in E$, then there exists a y that $\langle x, y\rangle \in E$, $y_1 \leq y$, and $y_2 \leq y$.

Hence, the only cause of the existence of two different objects having the same description is the execution of the second requirement.

A description mode is called *acceptable* if it is (recursively) enumerable and fulfills all the above requirements.

If one wishes to relate a complexity function with any description mode, one needs to introduce a volume function ℓ defined on X. Here we are interested in the cases $X = \mathbb{B}$ and $X = \mathbb{T}$ only. In both these cases, we put $\ell(x) = |x|$ where $|x|$ is the length of x.

Now let us fix approximation spaces X and Y, and let us consider the class of all acceptable description modes and the corresponding class of complexities. Let us ask whether the Solomonoff–Kolmogorov theorem holds for that class, and, if it does hold, then the related entropy will be called XY *entropy.*

It turns out that the Solomonoff–Kolmogorov theorem is valid for four cases when X and Y is respectively $\mathbb{B}$ or $\mathbb{T}$:

1. For $X = \mathbb{B}$ and $Y = \mathbb{B}$, we have $\mathbb{B}\mathbb{B}$-entropy,

2. For $X = \mathbb{B}$ and $Y = \mathbb{T}$, we have $\mathbb{B}\mathbb{T}$-entropy,

3. For $X = \mathbb{T}$ and $Y = \mathbb{B}$, we have $\mathbb{T}\mathbb{B}$-entropy, and

4. For $X = \mathbb{T}$ and $Y = \mathbb{T}$, we have $\mathbb{T}\mathbb{T}$-entropy.

It is easy to see that $\mathbb{B}\mathbb{B}$-, $\mathbb{B}\mathbb{T}$-, $\mathbb{T}\mathbb{B}$-, $\mathbb{T}\mathbb{T}$-entropy respectively coincides with $(=,=)$-, $(=,\gamma)$-, $(\gamma,=)$-, (γ,γ)-entropy of Sect.1.2. Speaking on the coincidence, take into account the Important Remark of Sect.1.

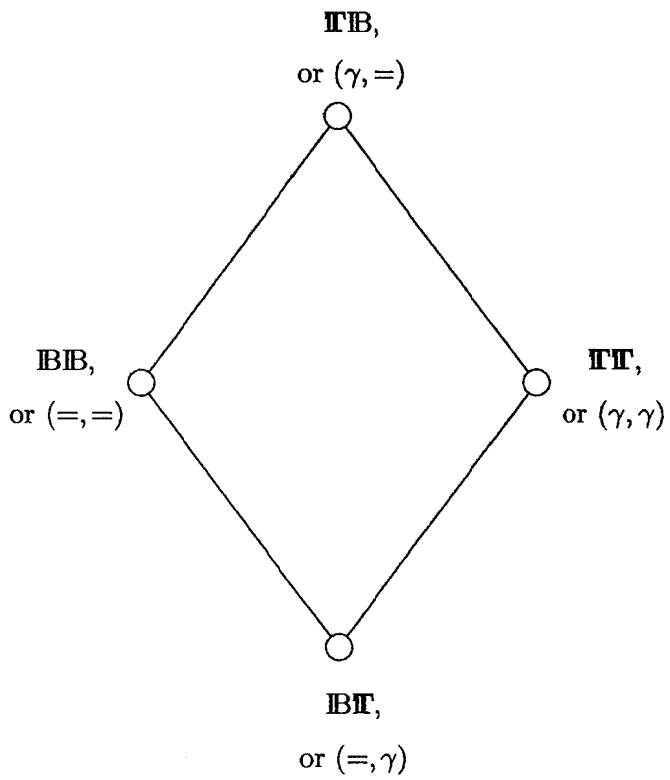

Figure 1: The ordering of the four entropies: $\mathbb{ITB}$, $\mathbb{IBB}$, $\mathbb{ITT}$, $\mathbb{IBT}$

1.4 The Ordering of the Four Entropies

Now we have four entropies, and any two of them, A and B, do not coincide; which means that the assertion $A(y) \underset{\text{fin}}{=} B(y)$ is not valid. Let us write $A < B$ if $A(y) \underset{\text{fin}}{\leq} B(y)$ but not vice versa. Then there is a partial ordering on the set of four entropies. That ordering can be shown by the following picture; Fig.1.

The picture is directed from bottom to top. That is, it shows that

$$\mathbb{IBT} < \mathbb{IBB},\ \mathbb{IBT} < \mathbb{ITT},\ \mathbb{IBB} < \mathbb{ITB},\ \mathbb{ITT} < \mathbb{ITB},$$

and, of course,

$$\mathbb{IBT} < \mathbb{ITB}.$$

On the other hand,

$$\text{neither } \mathbb{IBB} < \mathbb{ITT} \text{ nor } \mathbb{ITT} < \mathbb{IBB}.$$

1.5 Encoding-Free Generation of Complexities and Entropies

It turns out that the four entropies of Sect.1.4 admit an encoding-free definition with no use of such terms as "descriptions", "names", or "encodings".

As before and always, an entropy is defined as an optimal complexity function for some class $\mathcal{Z}$ of complexity functions; and all members of $\mathcal{Z}$ are functions from Y to $\mathbb{N} \cup \{\infty\}$. So our goal is to describe appropriate classes $\mathcal{Z}$.

Having this goal in mind, let us introduce two conditions, **C** and $\mathbf{\Sigma}$, which could be imposed on a function f: $Y \to \mathbb{N} \cup \{\infty\}$.

Condition 1. (**C** (of Cardinality of a set)) Let $n \in \mathbb{N}$, and let $M \subseteq Y$ be an arbitrary set such that

1. any two elements of M are non-comparable, and
2. $M \subseteq f^{-1}(n)$.

Then the cardinality of M is less than or equal to 2^n.

Condition 2. ($\mathbf{\Sigma}$ (of summation of a series)) Let $M \subseteq Y$ be an arbitrary set such that any two elements of M are non-comparable. Then

$$\sum_{y \in M} 2^{-f(y)} \leq 1.$$

Explanation 4. Elements y_1 and y_2 are *non-comparable* if neither $y_1 \leq y_2$ nor $y_2 \leq y_1$.

Thus, an arbitrary function f may or may not satisfy Condition **C** or Condition $\mathbf{\Sigma}$. And it is easy to see that Condition $\mathbf{\Sigma}$ implies Condition **C**.

Further, a definition of "enumerability from above" is to appear. A function $f : Y \to \mathbb{N} \cup \{\infty\}$ is called *enumerable from above* if the set $\{\langle y, n\rangle : y \in Y, n \in \mathbb{N}, f(y) \leq n\}$ is enumerable, that is, recursively enumerable.

Let us denote by $\mathcal{Z}(\square, Y)$ the class of all functions from Y to $\mathbb{N} \cup \{\infty\}$ that are enumerable from above and satisfy the condition $\square$, where $\square$ is either **C** or $\mathbf{\Sigma}$. Any element of $\mathcal{Z}(\square, Y)$ may be called a $\square$-*acceptable complexity*. Hence, we have four classes of acceptable complexities: $\mathcal{Z}(\mathbf{C}, \mathbb{B})$, $\mathcal{Z}(\mathbf{C}, \mathbb{\Pi})$, $\mathcal{Z}(\mathbf{\Sigma}, \mathbb{B})$, and $\mathcal{Z}(\mathbf{\Sigma}, \mathbb{\Pi})$. For each of these four classes, there holds the Solomonoff–Kolmogorov theorem.

Thus, there are four entropies: **C**$\mathbb{B}$-entropy, **C**$\mathbb{\Pi}$-entropy, $\mathbf{\Sigma}\mathbb{B}$-entropy, and $\mathbf{\Sigma}\mathbb{\Pi}$-entropy.

If one imposes the ordering on these four entropies, as in Sect.1.4, then one obtains the following picture; Fig.2.

The four entropies of this section also admit definitions with slightly modified versions of conditions **C** and $\mathbf{\Sigma}$.

Condition 3. (**C**′) There exists a constant b such that the cardinality of M is less than or equal to $b \cdot 2^n$ for every $M \subseteq Y$ satisfying the requirements (i) and (ii) of Condition **C**.

Condition 4. ($\mathbf{\Sigma}'$) There exists a constant b such that

$$\sum_{y \in M} 2^{-f(y)} \leq b$$

for every $M \subseteq Y$ of mutually non-comparable elements.

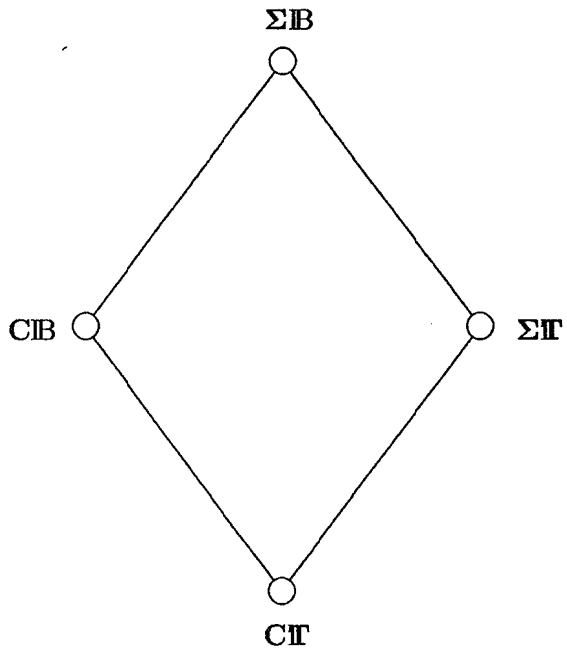

Figure 2: The ordering of four entropies: $\Sigma\mathbb{I}\mathbb{B}$, $\mathbf{C}\mathbb{I}\mathbb{B}$, $\Sigma\mathbb{I}\mathbb{T}$, $\mathbf{C}\mathbb{I}\mathbb{T}$

Classes $\mathcal{Z}(\mathbf{C}', \mathbb{I}\mathbb{B})$, $\mathcal{Z}(\mathbf{C}', \mathbb{I}\mathbb{T})$, $\mathcal{Z}(\Sigma', \mathbb{I}\mathbb{B})$, and $\mathcal{Z}(\Sigma', \mathbb{I}\mathbb{T})$ differ from the corresponding classes $\mathcal{Z}(\mathbf{C}, \mathbb{I}\mathbb{B})$, $\mathcal{Z}(\mathbf{C}, \mathbb{I}\mathbb{T})$, $\mathcal{Z}(\Sigma, \mathbb{I}\mathbb{B})$, and $\mathcal{Z}(\Sigma, \mathbb{I}\mathbb{T})$; nevertheless, the related entropies coincide in the sense of Important Remark of Sect.1. That is,

$$\mathbf{C}'\mathbb{I}\mathbb{B} = \mathbf{C}\mathbb{I}\mathbb{B}, \ \mathbf{C}'\mathbb{I}\mathbb{T} = \mathbf{C}\mathbb{I}\mathbb{T}, \ \Sigma'\mathbb{I}\mathbb{B} = \Sigma\mathbb{I}\mathbb{B}, \ \text{and} \ \Sigma'\mathbb{I}\mathbb{T} = \Sigma\mathbb{I}\mathbb{T}.$$

Condition 5. (Σ^∞) For an arbitrary set $M \subseteq Y$ of mutually non-comparable elements,

$$\sum_{y \in M} 2^{-f(y)} < +\infty.$$

It is obvious that Condition Σ^∞ is equivalent to Condition Σ' for $Y = \mathbb{I}\mathbb{B}$. Hence, the $\Sigma^\infty\mathbb{I}\mathbb{B}$-entropy coincide with the $\Sigma'\mathbb{I}\mathbb{B}$-entropy and consequently with the $\Sigma\mathbb{I}\mathbb{B}$-entropy.

Theorem 1. (Andrei Muchnik) The conditions Σ' and Σ^∞ are equivalent in the case $Y = \mathbb{I}\mathbb{T}$.

Hence the $\Sigma^\infty\mathbb{I}\mathbb{T}$-entropy coincides with the $\Sigma'\mathbb{I}\mathbb{T}$-entropy and with the $\Sigma\mathbb{I}\mathbb{T}$-entropy.

1.6 Relations between Two Quadruplets of Entropies

Now we have two quadruplets of entropies: the quadruplet $\mathbb{I}\mathbb{B}\mathbb{I}\mathbb{B}$, $\mathbb{I}\mathbb{B}\mathbb{I}\mathbb{T}$, $\mathbb{I}\mathbb{T}\mathbb{I}\mathbb{B}$, and $\mathbb{I}\mathbb{T}\mathbb{I}\mathbb{T}$, which respectively generated by means of encoding, and the quadruplet $\mathbf{C}\mathbb{I}\mathbb{B}$, $\mathbf{C}\mathbb{I}\mathbb{T}$, $\Sigma\mathbb{I}\mathbb{B}$, and $\Sigma\mathbb{I}\mathbb{T}$,

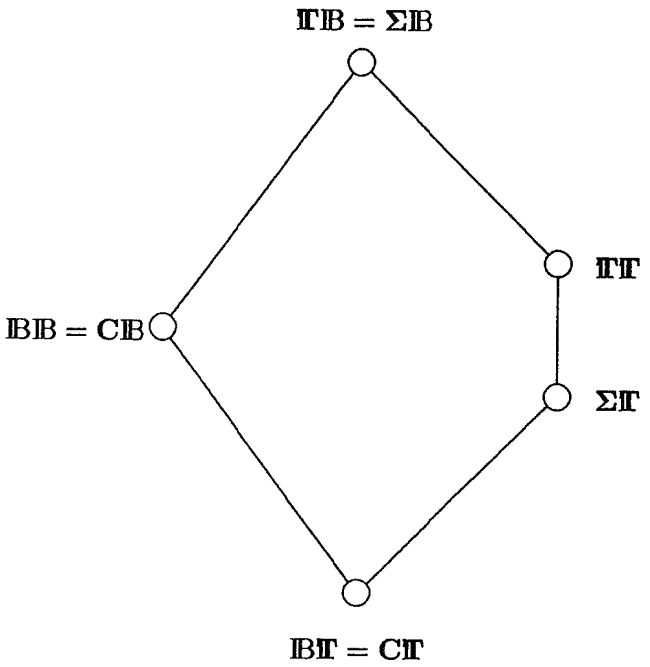

Figure 3: The relation between entropies

which respectively generated by using some quantitative approach represented by conditions $\mathbf{C}$ and $\mathbf{\Sigma}$.

It turns out that (in the sense of the equality explained in Sect.1, Important Remark) the following relations hold:

$$\mathbf{C}\mathbb{B} = \mathbb{B}\mathbb{B}, \ \mathbf{C}\mathbb{\Pi} = \mathbb{B}\mathbb{\Pi}, \ \text{ and } \ \Sigma\mathbb{B} = \mathbb{\Pi}\mathbb{B}.$$

As to $\Sigma\mathbb{\Pi}$, we have the following non-trivial fact, which will be discussed in Sect.2.2(5):

$$\Sigma\mathbb{\Pi} < \mathbb{\Pi}\mathbb{\Pi}.$$

Summarizing them, we obtain the following picture; Fig.3.

1.7 A Semantic for $\Sigma\mathbb{\Pi}$-Entropy

Four entropies of Fig.3 have an encoding semantic, but the fifth entropy, $\Sigma\mathbb{\Pi}$, has not yet obtained an appropriate semantic. Now a semantic for $\Sigma\mathbb{\Pi}$ will be set forth. That semantic is based upon probabilistic machines.

To this end let us consider a probabilistic Turing machine with one-way infinite output tape whose head moves in only one direction. "Probabilistic" means that one must flip a symmetric coin before performing any command, and the result of flipping determines which command is

to be performed. Another version: at the input tape, there comes a random infinite binary sequence with equal probabilities of digits. We suppose that our machine has binary output alphabet and never stops, so a finite or infinite binary sequence appears on the output tape.

Let us fix a machine M. For any $y \in \Xi$, let us denote by $\mathrm{P}^{\mathrm{n}}_M(y)$ the probability of the event 'y is the beginning of the output sequence'; in this notation, "n" stands for "non-stop". Consider a preorder relation $\leq$ on the set of machines: $M \leq N$ means that $\mathrm{P}^{\mathrm{n}}_M(y) \underset{\cap}{\leq} \mathrm{P}^{\mathrm{n}}_N(y)$.

Explanation 5. $A(y) \underset{\cap}{\leq} B(y)$ means that for some constant c not depending on y, and for every y, $A(y) \leq c \cdot B(y)$.

It turns out that there exists a *maximal* machine W such that $M \leq W$ for every machine M. In fact, there are many such machines, but any two of them, U and V, satisfy the condition

$$\mathrm{P}^{\mathrm{n}}_U(y) \underset{\cap}{=} \mathrm{P}^{\mathrm{n}}_V(y).$$

Explanation 6. $A(y) \underset{\cap}{=} B(y)$ iff $A(y) \underset{\cap}{\leq} B(y)$ and $B(y) \underset{\cap}{\leq} A(y)$.

Hence for any two maximal machines U and V, we have

$$|\log_2 \mathrm{P}^{\mathrm{n}}_U(y)| \underset{\cap}{=} |\log_2 \mathrm{P}^{\mathrm{n}}_V(y)|.$$

So we have moved from the probability $\mathrm{P}^{\mathrm{n}}_W$ to its logarithm. For any maximal machine W one can verify that

$$|\log_2 \mathrm{P}^{\mathrm{n}}_W(y)| \underset{\cap}{=} \mathbf{\Sigma\Pi}(y)$$

This fact enables us to identify $|\log_2 \mathrm{P}^{\mathrm{n}}_W|$ (or, if you prefer, the integer $\lfloor |\log_2 \mathrm{P}^{\mathrm{n}}_W| \rfloor$) with $\mathbf{\Sigma\Pi}$. (Recall again Important Remark of Sect.1). Then the probabilistic definition of $\mathrm{P}^{\mathrm{n}}_W$ just given can be taken as a semantic for $\mathbf{\Sigma\Pi}$.

The probability $\mathrm{P}^{\mathrm{n}}_W(y)$, related to an arbitrary maximal machine W, can be called *a priori probability* of y as an element of the tree $\mathbf{\Pi}$.

Remark 1. There exists also the a priori probability of y as an element of the bunch $\mathbf{B}$. To obtain the a priori probability of that second sort, one should consider probabilistic machines of a slightly different type. The change is: instead of machines that never stop, one should take now probabilistic machines that can stop. Then $\mathrm{P}^{\mathrm{s}}_M(y)$ is, by definition, the probability of the event 'the word printed on output tape after machine M stops coincides with y'; here "s" stands for "stop". A preorder on machines and the notion of a maximal machine are defined as above, and maximal machines do exist. Then $\mathrm{P}^{\mathrm{s}}_W(y)$ calculated for an arbitrary maximal machine W is the a priori probability of y as an element of $\mathbf{B}$. Here it occurs that

$$|\log_2 \mathrm{P}^{\mathrm{s}}_W(y)| \underset{\cap}{=} \mathbf{\Sigma B}(y).$$

Hence $\mathbf{\Sigma B}$ has a probabilistic semantic too. But, since $\mathbf{\Sigma B} = \mathbf{\Pi B}$, the entropy $\mathbf{\Sigma B}$ has also an encoding semantic.

1.8 Historical, Bibliographical, and Terminological Remarks; Acknowledgments

We begin the history of the theory of Kolmogorov complexity with Kolmogorov's paper [Kol65]. The purpose of that paper was to bring the notion of complexity (now we should say "of entropy") to the foundations of information theory. In his paper Kolmogorov expounded some results of his studies of 1963–1964. In those years he knew nothing about the paper [Sol64] in which Ray Solomonoff presented some similar ideas — but in vague and rather non-mathematical manner. We place the paper [Sol64] in the prehistory of the theory of Kolmogorov complexity. At the early stage of the theory's development, an important role belonged to the paper [ZL70].

In the papers of pioneers of the theory, there were introduced all five basic entropies of our Sect.1.6. The authors gave them various names and various notations. What was common in all those notations was the use of the letter "K" or the letter "k" as a part of the notation; one should believe the cause of this usage is a homage to Kolmogorov. Here we try to set some system of names and notations with the observance of the historical tradition. (In such a way the author makes his own contribution to the existing chaos of names and notations. This contribution is not too great because some names and notations are already in use. Simultaneously the author expresses the hope to introduce a standard system.)

We would like to fix the following names and notations for the five basic entropies.

1. For $\mathbb{IB}\mathbb{IB}$-entropy. Name: **simple entropy**; notation: KS.
2. For $\mathbb{IB}\mathbb{I}\Gamma$-entropy. Name: **decision entropy**; notation: KD.
3. For $\mathbb{I}\Gamma\mathbb{IB}$-entropy. Name: **prefix entropy**; notation: KP.
4. For $\mathbb{I}\Gamma\mathbb{I}\Gamma$-entropy. Name: **monotonic entropy**; notation: KM.
5. For $\Sigma\mathbb{I}\Gamma$-entropy. Name: **a priori entropy**; notation: KA.

The entropies should be attributed to the following authors:

- simple entropy KS to Kolmogorov [Kol65](§3) and also (though in some nebulous form) to Solomonoff [Sol64],
- decision entropy KD to Loveland [Lov69],
- a priori entropy KA to Levin [ZL70](n° 3.3) and [Lev73],
- monotonic entropy KM to Levin [Lev73], and
- prefix entropy KP to Levin [Lev76].

Remark 2. Strictly speaking, we denote by the symbols KS, KD, KA, KM and KP exactly those versions of entropies as they were formulated by Kolmogorov, Loveland, and Levin. Let us recall the Important Remark of Sect.1. The coincidence KS with $\mathbb{IB}\mathbb{IB}$ has the following meaning:

for any particular entropy KS and for any particular entropy $\mathbb{IB}$, there holds $\mathrm{KS}(y) \underset{\mathrm{A}}{=} \mathbb{IB}(y)$. The other coincidences, KD with $\mathbb{IT}$, etc., are to be understood in the same way.

Attributing the entropies to their inventors, we make no claim about the usage of these notations by the inventors. None of them made any essential use of the term "entropy"; usually the term "complexity" was used. Kolmogorov used simply the word "complexity" with no adjective. Loveland used the term "uniform complexity", and it was renamed as "decision complexity" by Zvonkin and Levin [ZL70](Definition 2.2). Levin used the words "monotonic complexity" and "complexity related to a prefix algorithm". He had not introduced any name for KA, but used terms "universal semicomputable measure" (in [ZL70](n° 3.3)) and "a priori probability" (in [Lev73]) for related quantities of which the logarithm is to be taken.

As to notations, Kolmogorov in [Kol65](§3) employed the notation $\mathrm{K}_A(y)$ for the simple entropy. Loveland in [Lov69](p.513) employed the notation $\mathrm{K}_A(x^n; n)$ for the decision entropy; and Zvonkin and Levin used for it the notation KR [ZL70](Definition 2.2). Zvonkin and Levin in [ZL70](n° 3.3) employed the notation $-\log_2 \mathrm{R}\{\Gamma_x\}$ for the a priori entropy; later, in [Lev73] that entropy was denoted by Levin as kM. In the same paper [Lev73] the notation km was used for the monotonic entropy. The notation KP (for the prefix entropy) appeared in [Lev76].

The general idea of an approximation space as a space of informations refining, or exactifying, one another is, without doubt, due to D. Scott. This idea was embodied into the notion of f_0-space in the sense of Yu. Ershov. A classification of entropies on the basis of that notion is given in [She84](Theorem 8); the classification of our Sect.1.3 is very close to that of [She84]. The general idea of the encoding-free approach to entropies (see Sect.1.5 above) was laid down in [Lev76].

A very useful exposition of various entropies and their interrelation is given in [Vyu81]. A survey of the use of entropies in a definition of randomness is presented in [KU87a] and [KU87b].

In the process of preparing this paper, the author had many discussions with Andrei Muchnik, Alexander Shen', and Nikolai Vereshchagin. The author enjoyed their advice and help. Many final formulations emerged from those thankworthy discussions. The bounds of Sect.2.1 and of Sect.2.2 probably belongs to what is called "mathematical folk-lore", but the final formulae are also due to discussions with Muchnik, Shen', and Vereshchagin.

To conclude this section let us redraw Fig.3 in terms and notations that we accept as standard; Fig.4.

The pentagon of Fig.4 shows, in particular, that neither KA $<$ KS nor KS $<$ KA. The exclamation note attached to an entropy means that the entropy can be used in the Kolmogorov definition of randomness.

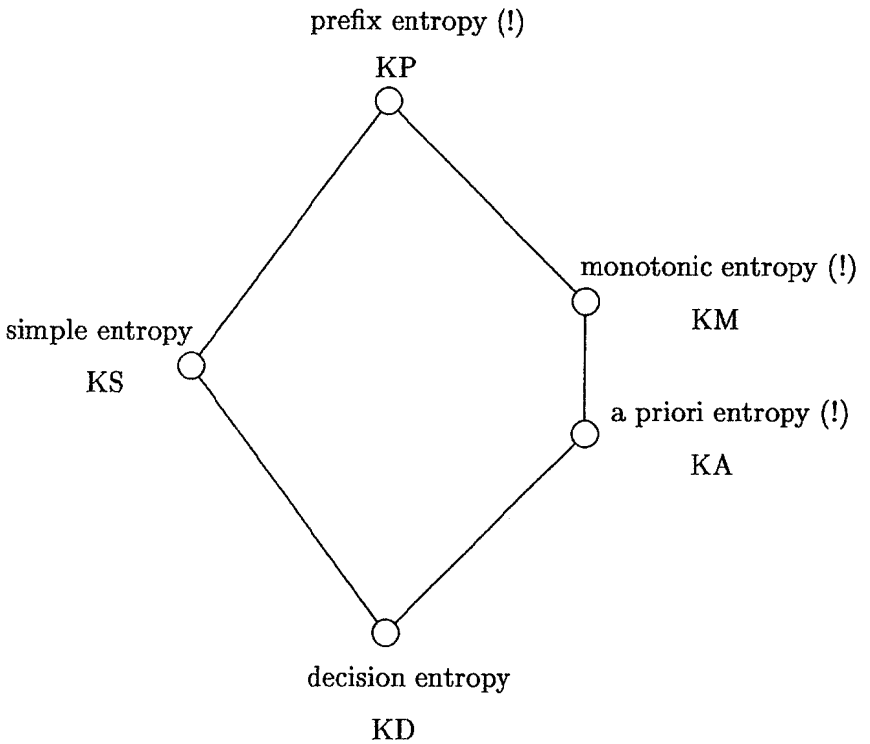

Figure 4: Five basic entropies

2 Quantitative Analysis on Entropies

2.1 Bounds for Entropies

Some upper and lower bounds for entropies will be written down in this section. But first of all the reader must be warned that in this exposition, the sense of an upper bound and the sense of a lower bound are rather different. An upper bound for an entropy shows that the entropy cannot be too large. A lower bound for an entropy does not show that the entropy cannot be too small but does show that in infinitely many instances, the entropy can be large enough. So upper bounds are absolute, or strong, upper bounds. Lower bounds are not absolute; we shall call them *weak* lower bounds. A weak lower bound has the purpose of supporting the corresponding upper bound and demonstrating, in a favorable case, that the upper bound cannot be improved.

After this warning let us consider the five basic entropies of Sect.1.8, Fig.4.

(1) Entropies KS, KM, KA, and KD.

Let *Ent* denotes one of the entropies KS, KM, KA, KD. Then (an upper bound) for all y (in Ξ),

$$Ent(y) \underset{+}{<} |y|,$$

and (a weak lower bound) for infinitely many y (in Ξ),

$$Ent(y) \underset{+}{>} |y|. \tag{1}$$

Let us formulate the lower bound of (1) more exactly. We have four cases: Ent = KS, Ent = KM, Ent = KA, Ent = KD. For each of these cases, the symbol *Ent* (as well as KS, KM, KA or KD) denotes an arbitrary function belonging to some collection, i.e., the collection of *Ent*-entropies. In each case the meaning of (1) is as follows: for any particular function *Ent* of that collection, there exist a constant c, perhaps negative, and an infinite set $M \subseteq \Xi$ such that

$$\forall y \in M \; [\; Ent(y) \geq |y| + c \;].$$

(2) Entropy KP.

It is helpful to introduce some notation. A function qlog, quasilogarithm, is introduced by the following definition:

$$\mathrm{qlog}z = \begin{cases} \log_2 z, & z \geq 1, \\ 0, & z \leq 1. \end{cases}$$

The iterations of that function are defined as follows:

$$\mathrm{qlog}^{(1)}z = \mathrm{qlog}z, \quad \text{and} \quad \mathrm{qlog}^{(k+1)}z = \mathrm{qlog}(\mathrm{qlog}^{(k)}z).$$

Then we have for any k, any $\epsilon > 0$, and all y,

$$\mathrm{KP}(y) \underset{\oplus}{\leq} |y| + \mathrm{qlog}^{(1)}|y| + \mathrm{qlog}^{(2)}|y| + \cdots + \mathrm{qlog}^{(k-1)}|y| + (1+\epsilon)\mathrm{qlog}^{(k)}|y|. \qquad (2)$$

It is an upper bound for KP.

For a weak lower bound, let k be an arbitrary positive integer. Then, for infinitely many y,

$$[\ \mathrm{KP}(y) \geq |y| + \mathrm{qlog}^{(1)}|y| + \mathrm{qlog}^{(2)}|y| + \cdots + \mathrm{qlog}^{(k)}|y|\ ]. \qquad (3)$$

That means that inequality (3) holds for any function denoted by KP (i.e. for any prefix entropy) and for an appropriate infinite set of y's, depending on the choice of that function.

Now it is reasonable to introduce an abbreviation for the sum $\mathrm{qlog}^{(1)}z + \cdots + (1+\epsilon)\mathrm{qlog}^{(k)}z$. Let us take, e.g., $Q_k(z,\epsilon)$ as such an abbreviation. That is,

$$Q_k(z,\epsilon) = \mathrm{qlog}^{(1)}z + \mathrm{qlog}^{(2)}z + \cdots + \mathrm{qlog}^{(k-1)}z + (1+\epsilon)\mathrm{qlog}^{(k)}z.$$

Then (2) and (3) can be rewritten as follows:

$$\forall y\ [\ \mathrm{KP}(y) \underset{\oplus}{\leq} |y| + Q_k(|y|,\epsilon)\ ],$$

and

$$\text{for infinitely many } y\ \ [\ \mathrm{KP}(y) \geq |y| + Q_k(|y|,0)\ ].$$

2.2 Bounds for Differences of Entropies

By how much can two entropies of different sorts, e.g., KP and KD, differ from one another? Perhaps it is better to ask, how much can one entropy exceed the other? Upper and lower bounds are to give the answer. The warning of the beginning of Sect.2.1 about the different meanings of upper and lower bounds is valid here too.

When $A(y) \underset{\oplus}{\leq} B(y)$, an upper bound for the difference $A(y) - B(y)$ is trivial; namely, a constant. So in this section, only differences $A(y) - B(y)$ for which the assertion $A(y) \underset{\oplus}{\leq} B(y)$ is false will be studied.

Now let us proceed to the differences.

(1) Difference KP – KD.

For any k, any $\epsilon > 0$, and all y,

$$\mathrm{KP}(y) - \mathrm{KD}(y) \underset{\oplus}{\leq} \mathrm{qlog}|y| + Q_k(|y|,\epsilon). \qquad (4)$$

For any k and infinitely many y,

$$\mathrm{KP}(y) - \mathrm{KD}(y) \geq \mathrm{qlog}|y| + Q_k(|y|,0). \qquad (5)$$

Note 2. Let us not forget that an additive constant implied in (4) depends not only on k and ϵ but also on the particular versions of KP and KD. In (5) the set of y's depends not only on k but also on the versions of KP and KD. This point is valid for further inequalities related to lower and upper bounds.

(2) Differences KS − KM and KS − KA.
Let $|y| \neq 0$. Then, for all y,

$$\mathrm{KS}(y) - \mathrm{KM}(y) \underset{+}{\leq} \mathrm{KS}(y) - \mathrm{KA}(y) \underset{+}{\leq} \log_2 |y|.$$

For some c and for infinitely many y,

$$\mathrm{KS}(y) - \mathrm{KM}(y) \geq \log_2 |y| + c\ ,$$

and for some c and for infinitely many y,

$$\mathrm{KS}(y) - \mathrm{KA}(y) \geq \log_2 |y| + c.$$

(3) Differences KM − KS and KA − KS.
For any k, any $\epsilon > 0$, and all y,

$$\mathrm{KM}(y) - \mathrm{KS}(y) \underset{+}{\leq} Q_k(|y|, \epsilon),$$
$$\mathrm{KA}(y) - \mathrm{KS}(y) \underset{+}{\leq} Q_k(|y|, \epsilon).$$

For any k and infinitely many y,

$$\mathrm{KM}(y) - \mathrm{KS}(y) \geq Q_k(|y|, 0),$$
$$\mathrm{KA}(y) - \mathrm{KS}(y) \geq Q_k(|y|, 0).$$

(4) Differences KP − KS, KS − KD, KP − KM, KP − KA, KM − KD, and KA − KD.

Let $B - A$ be any of the six entropy differences mentioned above. For any k, any $\epsilon > 0$, and all y,

$$B(y) - A(y) \mathop{\leq}_{\oplus} Q_k(|y|, \epsilon).$$

And for any k and infinitely many y,

$$B(y) - A(y) \geq Q_k(|y|, 0).$$

(5) The Difference KM − KA.

This difference is of special interest. The very fact that the conjecture $\mathrm{KM}(y) \mathop{=}_{\oplus} \mathrm{KA}(y)$ is false is disappointing. The refutation of that conjecture is due to Petér Gács [Gac83]. (The Hungarian surname "Gács" is to be pronounced as English "garch".)

Both KM and KA are defined on the binary tree $\mathbb{T}$. Gács studied two entropies K and H of similar sorts; but his K and H are defined not on $\mathbb{T}$ but on the tree consisting of all words in a countable alphabet (say, in $\mathbb{N}$ if one takes $\mathbb{N}$ as an alphabet). Some bound for the difference K−H is stated in Theorem 1.1 of [Gac83]; there the author 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" [Gac83](p.75). As it is known, Ackermann's function is a function from $\mathbb{N}$ to $\mathbb{N}$ which exceeds in its growth any primitive recursive function. The inverse f^{-1} for a function f is defined as follows:

$$f^{-1}(a) = \min\{z : f(z) \geq a\}.$$

Thus, for infinitely many y,

$$\mathrm{KM}(y) - \mathrm{KA}(y) \geq f^{-1}(|y|), \tag{6}$$

where f is the version of Ackermann's function mentioned by Gács.

Let $\mathrm{Z}(y)$ denote the number of zeros in the word y. Then, as a corollary of Theorem 1.1 of [Gac83], we have the following: for any k and for any m, there exists a $y \in \Xi$ such that $\mathrm{Z}(y) > m$ and

$$\mathrm{KM}(y) - \mathrm{KA}(y) \geq Q_k(Z(y), 0). \tag{7}$$

Therefore, we have two weak lower bounds. As to upper bounds, there is known no one except the following trivial one: for any k, any $\epsilon > 0$, and all y,

$$\mathrm{KM}(y) - \mathrm{KA}(y) \underset{+}{<} Q_k(|y|, \epsilon). \tag{8}$$

Since the weak lower bounds (6) and (7) do not support the upper bound (8), the task of improving all those bounds is open.

References

[Gac83] P. Gács. On the relation between descriptional complexity and algorithmic probability. *Theoretical Computer Science* 22:71–93, 1983.

[Kol65] A. N. Kolmogorov. Three approaches to the quantitative definition of information. *Problems Inform. Transmission* 1:1–7, 1965. (Translated from the Russian version.)

[Kol68] A. N. Kolmogorov. Logical basis for information theory and probability theory. *IEEE Trans. on Information Theory* IT-14.5:662–664, 1968. (The Russian version exists.)

[KU87a] A. N. Kolmogorov and V. A. Uspensky. Algorithms and randomness. In *Proc. 1st World Congress of the Bernoulli Society*, vol. 1, Yu. V. Prohorov and V. V. Sazonov, (eds.), VNU Science Press, Utrecht 3–53, 1987.

[KU87b] A. N. Kolmogorov and V. A. Uspensky. Algorithms and randomness. *Theory Probab. Appl.* 32:389–412, 1987. (Translated from the Russian version.) (*Comment:* There are two regrettable errors in the English version: p.394, line 2 from the bottom, and p.395, lines 1 and 3 from the top, the word "countable" must be replaced by "enumerable (i.e. recursively enumerable)"; p.395, line 1 from the top, the word "in" must be removed.)

[Lev73] L. A. Levin. On the notion of a random sequence. *Soviet Math. Dokl.* 14:1413–1416, 1973. (Translated from the Russian version.)

[Lev76] L. A. Levin. Various measures of complexity for finite objects (axiomatic description). *Soviet Math. Dokl.* 17:522–526, 1976. (Translated from the Russian version.)

[Lov69] D. W. Loveland. A variant of the Kolmogorov concept of complexity. *Information and Control* 15:510–526, 1969.

[Mar66] P. Martin-Lof. On the definition of random sequences. *Information and Control* 9:602–619, 1966.

[She84] A. Kh. Shen'. Algorithmic variants of the notion of entropy. *Soviet Math. Dokl.* 29:569–573, 1984. (Translated from the Russian version.) (*Comment:* There are many misprints in the English version.)

[Sol64] R. Solomonoff. A formal theory of inductive inference, Part I. *Information and Control* 7:1–22, 1964.

[US81] V. A. Uspensky and A. L. Semenov. What are the gains of the theory of algorithms: basic developments connected with the concept of algorithm and with its applications in mathematics (Part I,§17). In Springer-Verlag, Lecture Notes in Computer Science 122:100-234, 1981.

[US87] V. A. Uspensky and A. L. Semenov. *Teoria algoritmov: osnovnye otkrytiya i prilozheniya* (Theory of algorithms: main discoveries and applications) (§1.17). Nauka, Moscow, 1987; in Russian.

[Vyu81] V. V. V'yugin. Algorithmic entropy (complexity) of finite objects and its application to defining randomness and quantity of information. *Semiotika i Informatika* (Semiotics and Informatics) 16:14–43, 1981; in Russian.

[ZL70] A. K. Zvonkin and L. A. Levin. The complexity of finite objects and the development of the concepts of information and randomness by means of the theory of algorithms. *Russian Math. Surveys* 25(6):83–124, 1970. (Translated from the Russian version.)