1991

M. Li, P.M.B. Vitányi

Combinatorics and Kolmogorov complexity

Computer Science/Department of Algorithmics and Architecture Report CS-R9125 April

CWI, nationaal instituut voor onderzoek op het gebied van wiskunde en informatica

CWI is the research institute of the Stichting Mathematisch Centrum, which was founded on February 11, 1946, as a non-profit institution aiming at the promotion of mathematics, computer science, and their applications. It is sponsored by the Dutch Government through the Netherlands organization for scientific research (NWO).

Combinatorics and Kolmogorov Complexity

Ming Li[*] Paul M.B. Vitányi[†]

Computer Science Department
University of Waterloo
Waterloo, Ontario
Canada N2L 3G1

CWI
Kruislaan 413
1098 SJ Amsterdam
Netherlands

May 17, 1991

Abstract

On the one hand, we investigate combinatorial properties of finite sequences with high Kolmogorov complexity (like all blocks of equal length occur about equally frequent); on the other hand we demonstrate the utility of a Kolmogorov complexity method in combinatorial theory by several examples (like the 'coin-weighing' problem).

1980 Mathematics Subject Clasification: 94A15, 68C05, 60F10, 60A05, 60G50, 05C20, 05Axx

CR Categories: E.2, G.2, F.2, F.4

Keywords and Phrases: Kolmogorov complexity, algorithmic information theory, combinatorics, finite random sequences, normality, tournament graph, coin-weighing problem.

Note: This paper will appear in the *Proc. IEEE 6th Structures in Complexity Theory Conference*, Chicago, June 1991.

1 Introduction

On randomness related combinatorial properties of high Kolmogorov complexity finite binary sequences.

Infinite sequences generated by a $(\frac{1}{2}, \frac{1}{2})$ Bernoulli process (flipping a fair coin) have the property that the relative frequency of zeros in an initial n-length segment goes to $\frac{1}{2}$ for n goes to infinity. A related statement can be made for finite sequences, in the sense that one can say that the majority of all sequences will have about one half zeros. However, whereas the earlier statement is a property about individual infinite random sequences, the classical theory of probability has no machinery to define or deal with individual finite random sequences.

In [11], Kolmogorov established a notion of complexity (self-information) of finite objects which is essentially finitary and combinatorial. Martin-Löf has established, [17], that the proper definition of finite random sequences corresponds to finite sequences having about maximal Kolmogorov complexity.

It is useful for many applications (we give an application in combinatorial theory below), and of interest for its own sake, to determine the exact form in which certain combinatorial properties hold for high-complexity finite sequences (like equality of number of occurences of zeros and ones).

On a Kolmogorov complexity method in combinatorial theory.

Probabilistic arguments in combinatorial theory, as used by Erdös and Spencer [5], are usually aimed at establishing the existence of an object, in a nonconstructive sense. It is ascertained that a certain member of a class has a certain property, without actually exhibiting that object. Usually, the method proceeds by exhibiting a random process which produces the object with positive probability. Alternatively, a quantitative property is determined from a bound on its average in a probabilistic situation. The way to prove such 'existential' propositions often uses averages. We may call this 'first-moment' methods. 'Second-moment' methods, using means and variance of random variables

[*] Supported by the NSERC operating grants OGP-0036747 and OGP-046506. Email: mli@math.waterloo.edu

[†] Partially supported by the NSERC International Scientific Exchange Award ISE0046203.

to establish combinatorial results have been used by Moser [18]. Pippenger [19], has used related notions like 'entropy', 'self-information', and 'mutual information', from information theory, [21]. He gives two examples of 'universal propositions', such as a lower bound on the minimum of a quantity, or an upper bound on the maximum of a quantity.

Says Kolmogorov [12]: "The real substance of the entropy formula [based on probabilistic assumptions about independent random variables] ... holds under incomparably weaker and purely combinatorial assumptions... Information theory must precede probability theory, and not be based on it. By the very essence of this discipline, the foundations of information theory must have a finite combinatorial character."

It turns out to be quite natural to do combinatorial proofs by Kolmogorov complexity arguments, which are of themselves combinatorial in nature. We demonstrate the utility of a Kolmogorov complexity method in combinatorial theory by proving several combinatorial lower bounds (like the 'coin-weighing' problem).

2 Kolmogorov Complexity

To make this paper self-contained we briefly review notions and properties needed in the sequel. We identify the natural numbers $\mathcal{N}$ and the finite binary sequences as

$$(0,\epsilon),(1,0),(2,1),(3,00),(4,01),\ldots,$$

where ϵ is the empty sequence. The *length* $l(x)$ of a natural number x is the number of bits in the corresponding binary sequence. For instance, $l(\epsilon)=0$. If A is a set, then $|A|$ denotes the *cardinality* of A. Let $<\cdot,\cdot>: \mathcal{N}\times\mathcal{N}\rightarrow\mathcal{N}$ denote a standard computable bijective 'pairing' function. Throughout this paper, we will assume that $<x,y>=1^{l(x)}0xy$.

Define $<x,y,z>$ by $<x,<y,z>>$.

We need some notions from the theory of algorithms, see [20]. Let $\phi_1,\phi_2,\ldots$ be a standard enumeration of the partial recursive functions. The (Kolmogorov) *complexity* of $x\in\mathcal{N}$, given y, is defined as

$$C(x|y)=\min\{l(<n,z>):\phi_n(<y,z>)=x\}.$$

This means that $C(x|y)$ is the *minimal* number of bits in a description from which x can be effectively reconstructed, given y. The unconditional complexity is defined as $C(x)=C(x|\epsilon)$.

An alternative definition is as follows. Let

$$C_\psi(x|y)=\min\{l(z):\psi(<y,z>)=x\} \qquad (1)$$

be the conditional complexity of x given y with reference to decoding function ψ. Then $C(x|y)=C_\psi(x|y)$ for a universal partial recursive function ψ that satisfies $\psi(<y,n,z>)=\phi_n(<y,z>)$.

We will also make use of the *prefix* complexity $K(x)$, which denotes the shortest *self-delimiting* description. To this end, we consider so called *prefix* Turing machines, which have only 0's and 1's on their input tape, and thus cannot detect the end of the input. Instead we define an input as that part of the input tape which the machine has read when it halts. When $x\neq y$ are two such input, we clearly have that x cannot be a prefix of y, and hence the set of inputs forms what is called a *prefix code*. We define $K(x)$ similarly as above, with reference to a universal prefix machine that first reads 1^n0 from the input tape and then simulates prefix machine n on the rest of the input.

A survey is [15]. We need the following properties. Throughout 'log' denotes the binary logarithm. We often use $O(f(n))=-O(f(n))$, so that $O(f(n))$ may denote a negative quantity. For each $x,y\in\mathcal{N}$ we have

$$C(x|y)\leq l(x)+O(1). \qquad (2)$$

For each $y\in\mathcal{N}$ there is an $x\in\mathcal{N}$ of length n such that $C(x|y)\geq n$. In particular, we can set $y=\epsilon$. Such x's may be called *random*, since they are without regularities that can be used to compress the description. Intuitively, the shortest effective description of x is x itself. In general, for each n and y, there are at least $2^n-2^{n-c}+1$ distinct x's of length n with

$$C(x|y)\geq n-c. \qquad (3)$$

In some cases we want to encode x in *self-delimiting* form x', in order to be able to decompose $x'y$ into x and y. Good upper bounds on the prefix complexity of x are obtained by iterating the simple rule that a self-delimiting (s.d.) description of the length of x followed by x itself is a s.d. description of x. For example, $x'=1^{l(x)}0x$ and $x''=1^{l(l(x))}0l(x)x$ are both s.d. descriptions for x, and this shows that $K(x)\leq 2l(x)+O(1)$ and $K(x)\leq l(x)+2l(l(x))+O(1)$.

Similarly, we can encode x in a self-delimiting form of its shortest program $p(x)$ ($l(p(x))=C(x)$) in $2C(x)+1$ bits. Iterating this scheme, we can encode x as a selfdelimiting program of $C(x)+2\log C(x)+1$ bits, which shows that $K(x)\leq C(x)+2\log C(x)+1$, and so on.

Denote $C(<x,y>)$ by $C(x,y)$. (We also denote $C(x|<y,z>)$ by $C(x|y,z)$.) It can be proved, [12, 15], that, up to a an additive term $O(\log\min\{C(x),C(y)\})$,

$$C(x,y)=C(x)+C(y|x)=C(y)+C(x|y). \qquad (4)$$

This identity is sometimes referred to as *symmetry of information.* The logarithmic error term is caused by the fact that we need to encode a delimitor to separate two concatenated binary sequences (description of x and description of y given x) in the original pair.

3 Combinatorial Properties of High Kolmogorov Complexity Finite Sequences

E. Borel (1909) has called an infinite sequence of zeros and ones 'normal' in the scale of two if, for each k, the frequency of occurrences of each block y of length k in the initial segment of length n goes to limit 2^{-k} for n grows unbounded, [10]. It is known that normality is not sufficient for randomness, since Champernowne's sequence

$$123456789101112\ldots$$

is normal in the scale of ten. On the other hand, it is universally agreed that a random infinite sequence must be normal. (If not, then some blocks occur more frequent than others, which can be used to obtain better than fair odds for prediction.)

While in the infinite case one considers limiting values of quantitive properties which hold for each individual sequence of a set of probability 1, in the finite case one considers the *expected* value of quantities over a set of all sequences of a given length.

We would like to obtain statements that *individual* random finite sequences have such-and-such quantitative properties in terms of their length. But as the result of a sequence of n fair coin flips, *any* sequence of length n can turn up. This raises the question which subset of finite sequences can be regarded as genuinely random. In [17] the viewpoint is taken that finite sequences which satisfy all *effective* tests for randomness (known and unknown alike), are as random as we will ever be able to verify. This form of randomness of individual sequences turns out to be equivalent to such sequences having maximal Kolmogorov complexity. In the sequel we use 'complexity' in the sense of 'Kolmogorov complexity'.

We prove that each high complexity finite binary sequence is 'normal' in the sense that each binary block of length k occurs about equally frequent for k relatively small. In particular, this holds for $k = 1$. We quantify the 'about' and the 'relatively small' in this statement.

To distinguish individual random sequences obtained by flipping a physical coin from random sequences written down by human subjects, psychological tests [the correct reference is unknown to the authors] have shown that a consistent high classification score is reached by using the criterion that a real random sequence of length, say 40, contains a run of zeros or ones of length 6. In contrast, human subjects feel that short random sequences should not contain such long uniform runs.

We determine the maximal length of runs of zeros or ones which are *with certainty* contained in each high complexity finite sequence. We prove that each such sequence must contain a relatively long run of zeros.

The properties must be related to length of the sequence. In a sequence of length 1, or odd length, the number of zeros and ones cannot be equal. To apply such properties in mathematical arguments, it is often of importance that the precise extent to which such properties hold is known.

3.1 Expectation versus Complexity

To derive our results, we often use a common pattern of argument. Following a suggestion of John Tromp, we can formulate it in the form of a general 'Tail Law'.

Consider the sample space $S = \{0,1\}^*$ with uniform probability $\Pr(x) = 2^{-2l(x)}$. Put $S^n = \{0,1\}^n$. Then, $\Pr(x|x \in S^n) = 2^{-n}$. Let $R : S \rightarrow \mathcal{Z}$, total recursive, be a function that (in our case) measures the deviation between some function g of x and a reference value $r(l(x))$ for all strings of the same length. We are interested in the relation between the complexity of a string x and this deviation. A natural choice of r would be the average $g(x)$ over S^n. Fix a class D of *deficiency* functions $\delta : \mathcal{N} \rightarrow \mathcal{N}$ for which $K(n|n - \delta(n)) = O(1)$. This is satisfied by every monotone sublinear recursive function that we are interested in. The complexity of R can be identified with the complexity of its index in the effective enumeration of recursive functions, which we can assume equals some constant plus (optionally) the complexity of its parameters.

Define the tail probability

$$p(R; n, m) = \Pr\{x \in S^n : |R(x)| \geq m\}.$$

Lemma 1 (Tail Lemma) *Let f be a function from $\mathcal{N} \times \mathcal{N}$ to $\mathcal{N}$ satisfying*

$$-\log p(R; n, f(n,k)) \geq K(R|n) + k + O(1).$$

Then for any $\delta \in D$, we have that all x with $C(x) > n - \delta(n)$ $(n = l(x))$, satisfy

$$|R(x)| < f(n, \delta(n)).$$

Proof. By contradiction. Assume that for some $\delta \in D$, there exists an n such that $A = \{x \in S^n : |R(x)| \geq f(n, \delta(n))\}$ is non-empty. We can describe such an $x \in A$ in the following way:

1. let s be a s.d. program for n given $n - \delta(n)$, of length $l(s) = K(n|n - \delta(n)) = O(1)$.
2. let q be a s.d. program for R given n, of length $l(q) = K(R|n)$.
3. let i be the index of x in an effective enumeration of A, from the x's with the highest $|R(x)|$'s down. From $|A| = 2^n \Pr(A) = 2^n p(R; n, f(n, \delta(n)))$ it follows that the length of the (not necessarily s.d.) description of i satisfies:

$$\begin{aligned} l(i) \leq \log |A| &= n + \log p(R; n, f(n, \delta(n))) \\ &\leq n - K(R|n) - \delta(n) - O(1). \end{aligned}$$

The string sqi has length at most $n - \delta(n) - O(1)$ and can be padded to a string z of length exactly $n - \delta(n) - c$, where c is a constant determined below. From z we can reconstruct x by first using its length plus c to compute n (and $\delta(n)$) from s, then use n to obtain R from q, and finally enumerate A to obtain the ith element. Note that we can enumerate A up to the ith element without using f at all, since we enumerate from the x's with the highest $|R(x)|$ down. So, if recursive function ψ embodies above procedure for reconstructing x, we have, by Equation 1,

$$C(x) \leq C_\psi(x) + c_\psi \leq n - \delta(n) - c + c_\psi.$$

Choosing $c = c_\psi$ finishes the proof. □

Corollary 1 (Tail Lemma Dual) *The exact same argument shows that for sufficiently random x, the deviation $|R(x)|$ is not too small. We thus obtain a Tail Lemma Dual starting from $q(R; n, m) = \Pr\{x \in S^n : |R(x)| \leq m\}$.*

3.2 Number of Zeros and Ones

Let x have length n. It is known that if $C(x|n) = n + O(1)$, then the number of zeros it contains is, [17],

$$\frac{n}{2} + O(\sqrt{n}).$$

3.2.1 Fixed Complexity

We analyse what complexity can say about the number of zeros and ones. Let $x = x_1 x_2 \ldots x_n$ and $\delta \in D$ a deficiency function. Suppose,

$$C(x) \geq n - \delta(n).$$

Let $R(x) = \sum x_i - \frac{n}{2}$ be the deviation in the number of ones in x. With $x \in \{0,1\}^n$ uniformly distributed, $\#ones(x) = \sum x_i$ is distributed according to the binomial distribution.

A general estimate of the tail probability of the binomial distribution, with s_n the number of successful outcomes in n experiments with probability of success $0 < p < 1$ and $q = 1 - p$, is given by Chernoff's bounds, [5, 2],

$$\Pr(|s_n - np| \geq m) \leq 2e^{-m^2/4npq}. \qquad (5)$$

The tail probability $p(R; n, m)$ bounded by Equation 5 (with $R(x) = s_n - \frac{n}{2}$ and $p = q = 1/2$) yields:

$$-\log p(R; n, m) \geq \frac{m^2 \log e}{n} - 1$$

Clearly, R is a recursive function with $K(R|n) = O(1)$. Thus, choosing $f(n, k) = \sqrt{(k + O(1)) n \ln 2}$, Lemma 1 gives us: all x with $C(x) > n - \delta(n)$ ($n = l(x)$), satisfy

$$|\#ones(x) - \frac{n}{2}| < \sqrt{(\delta(n) + O(1)) n \ln 2}. \qquad (6)$$

If the complexity of x satisfies that the conditional complexity $C(x|n) \geq n - \delta(n)$, clearly Equation 6 holds a fortiori.

3.2.2 Fixed Number of Zeros

It may be surprising at first glance, but there are no maximally complex sequences with about equal number of zeros and ones. Equal numbers of zeros and ones is a form of regularity, and therefore lack of complexity. Using the same notation as before, if $R(x) = O(1)$ then the randomness deficiency $\delta(n) = n - C(x)$ is relatively large. For instance,

$$\begin{aligned} q(R; n, m) &= \Pr\{x \in S^n : |R(x)| \leq m\} \\ &\leq (2m+1) 2^{-n} \binom{n}{n/2} = O(\frac{m}{\sqrt{n}}). \end{aligned}$$

Thus, setting $f(n, k) = 2^{-k - O(1)} \sqrt{n}$, the Tail Law Dual (Corollary 1) gives us: all x with $C(x) > n - \delta(n)$ ($n = l(x)$), satisfy

$$|\#ones(x) - \frac{n}{2}| > 2^{-\delta(n)-O(1)}\sqrt{n}.$$

Perhaps more interestingly, we can define

$$R'(x) = \#ones(x) - (\frac{n}{2} + j),$$

so that $K(R'|n)$ is about $K(j)$. Applying the Tail Law Dual with

$$f(n,k) = 2^{-k-K(j)-O(1)}\sqrt{n},$$

we then find that all x with $C(x) > n - \delta(n)$ satisfy

$$|\#ones(x) - (\frac{n}{2} + j)| > 2^{-\delta(n)-K(j)-O(1)}\sqrt{n}.$$

This means that for a random x having exactly $j + n/2$ ones, $K(j)$ must be at least about $\log n$.

3.3 Number of Blocks

An infinite binary sequence is called *normal* if each block of length k occurs with limiting frequency of 2^{-k}. This justifies our intuition, that a random infinite binary sequence contains about as many zeros as ones. But also, blocks 00, 01, 10, and 11 should appear about equally often. In general we expect that each block of length k occurs with about the same frequency. Can we find an analogue for finite binary sequences? We analyse these properties for high complexity finite binary sequences to obtain a quantification of a similar statement in terms of the length of the sequence and its complexity.

3.3.1 Fixed Complexity

Let $x = x_1 \ldots x_n$ be a binary sequence of length n, and y a much smaller string of length l. Let $p = 2^{-l}$ and $\#y(x)$ be the number of (possibly overlapping) distinct occurrences of y in x. Put $R_y(x) = \#y(x) - np$. (So $R_1(x) = \sum x_i - n/2$.) For convenience, we assume that x 'wraps around' so that an occurrence of y starting at the end of x and continuing at the start also counts.

Theorem 1 *All x with $C(x) > n - \delta(n)$ satisfy*

$$|\#y(x) - np| < \sqrt{\alpha np},$$

with $\alpha = [K(y|n) + \log l + \delta(n) + O(1)](1-p)l4\ln 2$.

Proof. We prove by contradiction. Assume that n is divisible by l. (If it is not we can put x on a Procrustus bed to make its length divisible by l at the cost of having the above frequency estimate up to a $l/2$ additive term.) There are l ways of dividing (the ring) x into $N = n/l$ contiguous equal sized blocks, each of length l. For each such division $i \in \{0, 1, \ldots, l-1\}$, let $R_{y,i}(x)$ be the number of (now nonoverlapping) occurrences of block y minus Np. Notice that $R_{y,i}(x)$ is the deviation from the expectation of a Bernoulli sequence of length N with probability of succes (a block matching y) p, for which we can use the Chernoff bound 5.

$$p(R_{y,i}; n, m) \le 2e^{-m^2/4Np(1-p)}.$$

Taking the negative logarithm on both sides:

$$-\log p(R_{y,i}; n, m) \ge \frac{m^2 \log e}{4Np(1-p)} - 1. \qquad (7)$$

Choose $m = f(n,k)$, such that

$$\frac{f(n,k)^2 \log e}{4Np(1-p)} = K(R_{y,i}|n) + k + O(1). \qquad (8)$$

Equations 7, 8 enable us to apply the Tail Lemma 1. Application of the Tail Lemma yields that all x with $C(x) > n - \delta(n)$ satisfy $|R_{y,i}(x)| < f(n, \delta(n))$. Substitution of f according to Equation 8, with $K(R_{y,i}|n) = K(y,i|n) + O(1)$, gives:

$$|R_{y,i}(x)| < \sqrt{\frac{K(y,i|n) + \delta(n) + O(1)}{\log e} 4Np(1-p)}.$$

The theorem now follows by noting that $R_y(x) = \sum_{i=0}^{l-1} R_{y,i}(x)$, and $K(i|l) \le \log l$ □

With $C(x|n, R_y) \ge n - \delta(n)$, Theorem 1 holds a fortiori.

3.3.2 Fixed Number of Blocks

Similar to the analysis of blocks of length 1, the complexity drops below its maximum in case some block y of length l occurs in one of the l block divisions, say i, with frequency exactly pN ($p = 1/2^l$). Then we can point out x by giving n, y, i and its index in a set of cardinality

$$\binom{N}{pN}(2^l - 1)^{N-pN} = O(\frac{2^{Nl}}{\sqrt{Np(1-p)}}).$$

Therefore,

$$C(x|n,y) \le n - \frac{1}{2}\log n + \frac{1}{2}(l + 3\log l) + O(1).$$

3.4 Length of Runs of Zeros

It is known from probability theory, that in a randomly generated finite sequence the *expectancy* of the length of the longest run of zeros or ones is pretty high. For each individual finite sequence with high Kolmogorov complexity we are *certain* that it contains each block up to a certain length (like a run of zeros).

Theorem 2 *Let x of length n satisfy $C(x) \geq n - \delta(n)$. Then x contains all blocks y of length*

$$l = \log n - \log\log n - \log(\delta(n) + \log n) - O(1).$$

Proof. We are sure that y occurs at least once in x, if $\sqrt{\alpha np}$ in Theorem 1 is at most np. This is the case if $\alpha \leq np$, that is:

$$\frac{K(y|n) + \log l + \delta(n) + O(1)}{\log e} 4l \leq np.$$

Substitute $K(y|n) \leq l + 2\log l + O(1)$ (since $K(y|n) \leq K(y)$), and $p = 2^{-l}$ with l set at

$$l = \log n - \log(3\delta(n)\log n + 3\log^2 n),$$

(which equals l in the statement of the theorem up to an additive constant). The result is

$$\frac{l + 3\log l + \delta(n) + O(1)}{\log e} 4l \leq 3(\delta(n)\log n + \log^2 n),$$

and it is easy to see that this holds for sufficiently large n. □

Corollary 2 *If $\delta(n) = O(\log n)$ then each block of length $\log n - 2\log\log n - O(1)$ is contained in x.*

Corollary 3 *Analysing the proof of Theorem 2 we can improve this in case $K(y|n)$ is low. If $\delta(n) = O(\log\log n)$, then for each $\epsilon > 0$ and n large enough, x contains an all-zero run y (for which $K(y|n) = O(\log l)$) of length $l = \log n - (1+\epsilon)\log\log n + O(1)$.*

Corollary 4 (improving [2]) *Since there are $2^n(1 - O(1/\log n))$ strings x of length n with $C(x) \geq n - \log\log n + O(1)$, the* expected length of the longest run of consecutive zeros *if we flip a fair coin n times, is at least l as in Corollary 3. This improves the lower bound of $\log n - 2\log\log n$ cited in [2] by a $\log\log n$ additive term.*

We show in what sense Theorem 2 is sharp. Let $x = uvw$, $l(x) = n$ and $C(x) \geq n - \delta(n)$. We can describe x by giving

1. A description of v in $K(v)$ bits.
2. The literal representation of uw.
3. A description of $l(u)$ in $\log n + \log\log n + 2\log\log\log n + O(1)$

Then, since we can find n by $n = l(v) + l(uw)$,

$$\begin{aligned} C(x) &\leq n - l(v) + K(v) + \log n \\ &+ (1 + o(1))\log\log n + O(1). \end{aligned} \qquad (9)$$

Substitute $C(x) = n - \delta(n)$ and $K(v) = o(\log\log n)$ (choose v to be very regular) in Equation 9 to obtain:

$$l(v) \leq \delta(n) + \log n + (1 + o(1))\log\log n.$$

This means that, for instance, for each $\epsilon > 0$, no maximally complex string x with $C(x) = n + O(1)$ contains a run of zeros (or the initial binary digits of π) of length $\log n + (1+\epsilon)\log\log n$ for n large enough and regular enough. By Corollary 3, on the other hand, such a string x *must* contain a run of zeros of length $\log n - (1+\epsilon)\log\log n + O(1)$.

4 A Kolmogorov Complexity Method in Combinatorial Theory

One can often convert Kolmogorov arguments (or probabilistic arguments for that matter) into counting arguments. Our intention is pragmatic: we aim for arguments which are easy to use in the sense that they supply rigorous analogs for our intuitive reasoning why something should be the case, rather than have to resort to nonintuitive meanderings along seemingly unrelated mathematical byways. It is always a matter of using regularity in an object, imposed by a property under investigation and quantified in an assumption to be contradicted, to compress the object's description to below its minimal value.

We treat two examples from Erdös and Spencer's book, and the two examples in Pippenger's article. It is only important to us to show that the application of Kolmogorov complexity in combinatorics is not restricted to trivialities.

4.1 Tournaments

The first example proved by Erdös and Spencer in [5] by the probabilistic method, Theorem 3, is originally due to Erdös and Moser [4]. (Rather, a version with $\lfloor 2\log n \rfloor$ instead of $2\lceil \log n \rceil$.) A *tournament* T is a complete directed graph. That is, for each pair of nodes i and j in T, exactly one of edges (i,j), (j,i) is in the graph. The nodes of a tournament can be viewed as *players* in a game tournament. If (i,j) is in T we say player j *dominates* player i. We call T *transitive* if (i,j), (j,k) in T implies (i,k) in T.

Let Γ be the set of all tournaments on $N = \{1,\ldots,n\}$. Given a tournament $T \in \Gamma$, fix a standard coding $E : \Gamma \rightarrow \mathcal{N}$, such that $l(E(T)) = n(n-1)/2$ bits, one bit for each edge. The bit for edge (i,j) is set to 1 if $i<j$ and 0 otherwise.

Theorem 3 *If $v(n)$ is the largest integer such that every tournament on N contains a transitive subtournament on $v(n)$ nodes, then $v(n) \leq 1 + 2\lceil \log n \rceil$ from some n onwards.*

Proof. By Equation 3, fix $T \in \Gamma$ such that

$$C(E(T)|n) \geq l(E(T)). \tag{10}$$

Let S be the transitive subtournament of T on $v(n)$ nodes. We compress $E(T)$, to an encoding $E'(T)$, as follows.

1. Prefix the list of nodes in S in lexicographical order of dominance to $E(T)$, each node using $\lceil \log n \rceil$ bits, adding $v(n)\lceil \log n \rceil$ bits.

2. Delete all redundant bits from the $E(T)$ part, representing the edges between nodes in S, saving $v(n)(v(n)-1)/2$ bits.

Then,

$$l(E(T)) = l(E'(T)) + \frac{v(n)}{2}(v(n) - 2\lceil \log n \rceil - 1). \tag{11}$$

Given n, an $O(1)$ bit description of this discussion and $E'(T)$ suffice to reconstruct $E(T)$. (We can find $v(n)$ by exhaustive search.) Therefore,

$$C(E(T)|n) \leq l(E'(T)) + O(1). \tag{12}$$

For large enough n, Equations 10, 11, and 12 can only be satisfied with $v(n) \leq 1 + 2\lceil \log n \rceil$. □

The general idea used is the following. [1] If each tournament contains a large transitive subtournament, then also a T of maximal complexity contains one. But the regularity induced by the transitive subtournament can be used to compress the description of T to below its complexity, yielding the required contradiction. Use the method on the following.

Exercise. Let $w(n)$ be the largest integer so that for each tournament T on N there exist disjoint sets A and B in N of cardinality $w(n)$ such that $A \times B \subseteq T$. Prove $w(n) \leq 2\lceil \log n \rceil$. Hint: add $2w(n)\lceil \log n \rceil$ (describe nodes), and save $w(n)^2$ (on edges).

The second example is Theorem 9.1 in [5], originally due to Erdös [3]. A tournament T on N has property $S(k)$ if for every set A of k nodes (players) there is a node (player) in $N - A$ which dominates (beats) all nodes in A. Let $s(k)$ be the minimum number of nodes (players) in a tournament with property $S(k)$. An upper bound on $s(k)$ has applications in constructing time stamp systems in distributed computing, [14].

Theorem 4 $s(k) \leq 2^k k^2(\log_e 2 + o(1))$.

Proof. Assume the notation of the previous theorem. By Equation 3, choose T such that

$$C(E(T)|n,k) \geq l(E(T)) = n(n-1)/2. \tag{13}$$

Assume that $S(k)$ is false for

$$n = 2^k k^2(\log_e 2 + o(1)). \tag{14}$$

Fix a set A of k nodes with no common dominator in $N - A$. Describe T as follows by a compressed effective encoding $E'(T)$.

1. List the nodes in A first, using $\lceil \log n \rceil$ bits each;

2. Secondly, list $E(T)$ with the bits representing edges between $N-A$ and A deleted (saving $(n-k)k$ bits).

3. Thirdly, code the edges between $N - A$ and A. From each $i \in N-A$, there are 2^k-1 possible ways of directing edges to A, in total $t = (2^k-1)^{n-k}$ possibilities. To encode the list of edges $\lceil \log t \rceil$ bits suffice.

Given n, one can reconstruct $E(T)$ from this discussion ($O(1)$ bits), and $E'(T)$. Hence,

$$C(E(T)|n,k) \leq l(E'(T)) + O(1). \tag{15}$$

Calculation shows that, for large enough n, Equation 14 is consistent with:

$$l(E(T)) > l(E'(T)) + k^{2-\epsilon}, 0 < \epsilon < 1. \tag{16}$$

Equations 13, 14, 15, 16, yield the desired contradiction. Therefore, $s(k) \leq n$. □

[1] For each n, define T_n as the Turing machine that on input $E'(T)$ outputs $E(T)$. Define complexity C_{T_n} relative to T_n and repeat the given argument, dispensing with the $O(1)$ error term in Equation 12. This proves Theorem 3 for each n.

4.2 The Coin-Weighing Problem

A family $\mathcal{D} = \{D_1, \ldots, D_j\}$ of subsets of $N = \{1, \ldots, n\}$ is called a *distinguishing family* for N if for any two distinct subsets M and M' of N there exists an i $(1 \leq i \leq j)$ such that $|D_i \cap M|$ is different from $|D_i \cap M'|$. Let $f(n)$ denote the minimum of $|\mathcal{D}|$ over all distinguishing families for N. To determine $f(n)$ is commonly known as the *coin-weighing problem*. It is known, that

$$f(n) = \frac{2n}{\log n}(1 + O(\frac{\log\log n}{\log n})).$$

Erdös and Rényi, [6], Moser, [18], and Pippenger, [19], have used various methods in combinatorics to show the lower bound in the theorem below. Pippenger used an information theoretic argument. We will supply a proof using Kolmogorov complexity. Fix a standard encoding $E : 2^N \rightarrow \mathcal{N}$, such that $E(A)$, $A \subseteq N$, is n bits, one bit for each node in N. The bit for node i is set to 1 if node i is in A, and 0 otherwise. Define $E(\mathcal{D}) = (E(D_1), \ldots, E(D_j))$. To simplify notation, in the proof below we identify A with $E(A)$, where $A \subseteq N$ or $A = \mathcal{D}$.

Theorem 5

$$f(n) \geq \frac{2n}{\log n}[1 + O(\frac{\log\log n}{\log n})].$$

Proof. Use the notation above. By Equations 2, 3, choose M such that

$$C(M|\mathcal{D}) \geq n. \qquad (17)$$

Let $m_i = |D_i \cap M|$. Since $\mathcal{D}$ is a distinguishing family for N: given $\mathcal{D}$, the values $m_1, \ldots, m_j$ determine M. Hence,

$$C(M|\mathcal{D}) \leq C(m_1, \ldots, m_j|\mathcal{D}) + O(1). \qquad (18)$$

Let $d_i = |D_i|$, and assume $d_i > \sqrt{n}$. By a standard argument (detailed after the proof), Equation 17 implies that the *randomness deficiency* $k = d_i - C(M \cap D_i|D_i)$ is $O(\log d_i)$. Therefore, by Equation 6 or the general Theorem 1, m_i is within range $\frac{d_i}{2} + O(\sqrt{d_i \log d_i})$. Since m_i can be described by its discrepancy with $d_i/2$, and $d_i \leq n$,

$$C(m_i|D_i) \leq \frac{1}{2}\log n + O(\log\log n), 1 \leq i \leq j.$$

For $d_i \leq \sqrt{n}$ this holds trivially. Pad each description of an m_i to a block of length $\frac{1}{2}\log n + O(\log\log n)$. Then,

$$C(m_1, \ldots, m_j|\mathcal{D}) \leq \sum_{i=1}^{j}(\frac{1}{2}\log n + O(\log\log n)). \qquad (19)$$

By Equations 17, 18, and 19, $j \geq n/(\frac{1}{2}\log n + O(\log\log n))$, which is equivalent to the theorem. □

Standard Argument. A useful property states that if an object has maximal complexity, then the complexity of a part, which is easily described as part of the whole, cannot be too far below maximal. In the particular case involved in the proof above, the standard argument runs as follows. The randomness deficiency k as defined in the proof cannot be large, since we can reconstruct M from:

1. A description of this discussion, and delimitors between the separate description items, in $O(\log n)$ bits.

2. The literal description of $E(M)$ leaving out the bits corresponding to elements in D_i, saving d_i bits.

3. The assumed short program to reconstruct the bits in $E(M)$ corresponding to elements in D_i, adding $d_i - k$ bits.

4. A description of $\mathcal{D}$ and i.

Then, $C(M|\mathcal{D}, i) \leq n - k + O(\log n)$, which by Equation 17 implies that $k \leq C(i) + O(\log n)$. Since $i \leq j$, and $j \leq n$ (the set of singleton sets in N is a distinguishing family), we find $k = O(\log n)$.

4.3 Covering Families

Let n and N be as before, and let $K(N)$ denote the set of all unordered pairs of elements from N (the complete n-graph). If A and B are disjoint subsets of N, then $K(A, B)$ denotes the set of all unordered pairs $\{u, v\}$, $u \in A$ and $v \in B$ (complete bipartite graph on A and B). A family $\mathcal{C} = (K(A_1, B_1), \ldots, K(A_j, B_j))$ is called a *covering family* of $K(N)$, if for any pair $\{u, v\} \in K(N)$, there exists an i $(1 \leq i \leq j)$ such that $\{u, v\} \in K(A_i, B_i)$. For each i $(1 \leq i \leq j)$, set $C_i = A_i \cup B_i$, and $c_i = |C_i|$. Let $g(n)$ denote the minimum of

$$\sum_{1 \leq i \leq j} c_i,$$

over all covering families for $K(N)$. The problem of determining $g(n)$ arises in the study of networks of contacts realizing a certain symmetric Boolean function, and the following is known, [9]:

$$n \log n \leq g(n) < n \log n + (1 - \log e + \log \log e)n.$$

The lower bound on $g(n)$ was also proven by Pippenger, [19], using an information theoretic argument. There the reader can find additional references to the source of the problem and its solutions. We shall give a short Kolmogorov complexity proof for the following.

Theorem 6

$$\frac{g(n)}{n} \geq \log n + O(\log \log n).$$

Proof. Use the notation above. For each $x \in N$, there is a $y = y_1 \cdots y_j$, and a binary sequence z of an exactly sufficient number of bits for the construction below, with $C(z|n,x) \geq l(z)$.

1. If $x \in A_i$, then $y_i = 0$.
2. If $x \in B_i$, then $y_i = 1$.
3. If $x \in N - C_i$, then $y_i =$ next unused bit of z.

Denote y and z associated with x by y^x and z^x. Given n, we can reconstruct $\mathcal{C}$ as the lexicographically least minimal covering family. Therefore, we can reconstruct x from y^x and n, by exhaustive matching of all elements in N with y^x under $\mathcal{C}$. Namely, suppose distinct x and x' match. By the covering property, $\{x, x'\} \in K(A_i, B_i)$ for some i. But then $y_i^x \neq y_i^{x'}$. Hence, $C(x|n, y^x) = O(1)$. Then, by Equation 4, we have:

$$\begin{aligned} R(x) &\overset{\text{def}}{=} C(y^x|n) - C(y^x|n,x) - C(x|n) \\ &= O(\log C(x|n)). \end{aligned} \qquad (20)$$

Given n and x, we can reconstruct y^x from z^x and $\mathcal{C}$, first reconstructing the latter item from n as above. Thus, up to an $O(n)$ additive term, $\sum_{x \in N} C(y^x|n,x)$ can be evaluated, from the number of bits in the z^x's, as follows.

$$\begin{aligned} \sum_{x \in N} |\{i : x \in N - C_i\}| &= \sum_{1 \leq i \leq j} |\{x : x \in N - C_i\}| \\ &= nj - \sum_{1 \leq i \leq j} c_i. \end{aligned} \qquad (21)$$

For each x, by Equation 2,

$$C(y^x|n) \leq l(y^x) + O(1) = j + O(1), \qquad (22)$$

and $C(x|n) \leq \log n + O(1)$. Estimating the lower bound on $\sum C(x|n)$ by Equation 3,

$$\sum_{x \in N} C(x|n) = n \log n + O(n). \qquad (23)$$

By Equations 20, 2, 21, 22, and 23 we have

$$\begin{aligned} \sum_{1 \leq i \leq j} c_i - n \log n + O(n) &\geq \sum_{x \in N} R(x) \\ &= O(n \log \log n), \end{aligned}$$

from which the theorem follows. □

One may wonder whether we can remove the $O(\log \log n)$ error term. Recall that the prefix variant of complexity $K(x|y)$, [13, 7, 1] or [15] and Section 2, is the length of the shortest self-delimiting description from which x can be reconstructed, given the shortest self-delimiting description for y (rather than y literally). A description is 'self-delimiting' if the interpreter can determine the end of it without looking at additional bits. This K complexity is more precise for some applications. In its K version, Equation 4 holds to within an $O(1)$ additive term, rather than the $O(\log \log n)$ one, [7]. Then, in Equation 20, the K version of $R(x) = O(1)$. A straightforward, somewhat tedious, analysis shows that estimates of the quantities in Equations 21, 23, and 22, still hold in K-version. Together, it follows that $g(n)/n \geq \log n + O(1)$.

Acknowledgement

Prabhakar Ragde drew our attention to Pippenger's paper. John Tromp gave valuable comments on the manuscript.

References

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

[2] Corman, C. Leiserson, R. Rivest, *Introduction to Algorithms*, 1990.

[3] P. Erdös, On a problem in graph theory, *Math. Gazette*, **47**(1963), 220-223.

[4] P. Erdös and L. Moser, A problem on tournaments, *Canad. Math. Bull.*, **8**(1964), 351-356.

[5] P. Erdös and J. Spencer, *Probabilistic Methods in Combinatorics*, Academic Press, New York, 1974.

[6] P. Erdös and A. Rényi, On two problems of information theory, *Publ. Hungar. Ac. Sci.*, **8**(1963), 241-254.

[7] P. Gács, On the symmetry of algorithmic information, *Soviet Math. Dokl.*, **15**(1974), 1477-1480.

[8] R.L. Graham, D.E. Knuth, O. Patashnik, *Concrete Mathematics*, Addison-Wesley, 1989.

[9] G. Hansel, Nombre minimal de contacts de fermature necessaire pour realiser une function booleenne symetrique de n variables, *C.R. Acad. Sci. Paris*, **258**(1964), 6037-6040.

[10] D.E. Knuth, *Seminumerical Algorithms*, Addison-Wesley, 1981.

[11] A.N. Kolmogorov, Three approaches to the definition of the concept 'quantity of information', *Problems in Information Transmission*, **1:1**(1965), 1-7.

[12] A.N. Kolmogorov, Combinatorial foundation of information theory and the calculus of probabilities, *Russian Math. Surveys*, **38:4**(1983), 29-40.

[13] L.A. Levin, Laws of Information conservation (non-growth) and aspects of the foundation of probability theory, *Problems in Information Transmission*, **10**(1974), 206-210.

[14] A. Israeli and M. Li, Bounded Time Stamps, *Proc. 27th IEEE Symp. Found. Comp. Sci.*, 1987, 371-382.

[15] M. Li and P.M.B. Vitányi, Kolmogorov complexity and its applications, pp. 187-254 in: *Handbook of Theoretical Computer Science, Vol. A*, J. van Leeuwen, Ed., Elsevier/MIT Press, 1990.

[16] M. Li and P.M.B. Vitányi, Combinatorial properties of finite sequences with high Kolmogorov complexity, manuscript.

[17] P. Martin-Löf, The definition of random sequences, *Information and Control*, **9**(1966), 602-619.

[18] L. Moser, The second moment method in combinatorial analysis, pp. 283-384 in: *Combinatorial Structures and Their Applications*, Gordon and Breach, New York, 1970.

[19] N. Pippenger, An information-theoretic method in combinatorial theory, *J. Combinat. Th. (A)*, **23**(1977), 99-104.

[20] H.J. Rogers, Jr., *Theory of Recursive Functions and Effective Computability*, McGraw-Hill, 1967.

[21] C.E. Shannon, A mathematical theory of communication, *Bell System Tech. J.*, **27**(1948), 379-423, 623-656.