Kolmogorov Complexity Arguments in Combinatorics

Ming Li*
University of Waterloo

Paul M.B. Vitányi†
CWI and Universiteit van Amsterdam

Abstract

The utility of a Kolmogorov complexity method in combinatorial theory is demonstrated by several examples.

1 Introduction

Probabilistic arguments in combinatorial theory, as used by Erdös and Spencer [5], are usually aimed at establishing the existence of an object, in a non-constructive 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

*Supported by the NSERC operating grants OGP-0036747 and OGP-046506. Address: Computer Science Department, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1. Email: mli@math.waterloo.edu

†Partially supported by the NSERC International Scientific Exchange Award ISE0046203 and by NWO through NFI Project ALADDIN under Contract number NF 62-376. Address: CWI, Kruislaan 413, 1098 SJ Amsterdam, The Netherlands. Email: paulv@cwi.nl

'first-moment' methods. 'Second-moment' methods, using means and variance of random variables 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.

In [10], Kolmogorov established a notion of complexity (self-information) of finite objects which is essentially finitary and combinatorial. Says Kolmogorov [11]: "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 is the aim of this paper to demonstrate how to replace probability based arguments in combinatorics by complexity based arguments, which of themselves are essentially combinatorial in nature without probabilistic assumptions at all.

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 introduced this method, and gave a comparison of proofs of the first example in this paper by counting, by probabilistic argument, and by Kolmogorov complexity argument, in [14]. Here 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. To make this paper self-contained we briefly review notions and properties needed in the sequel.

2 Kolmogorov Complexity

We identify the natural numbers $\mathcal{N}$ and the finite binary sequences as

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

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, $l(\epsilon) = 0$. If A is a set, then $|A|$ denotes the *cardinality* of A. Let $< . >: \mathcal{N} \times \mathcal{N} \rightarrow \mathcal{N}$ denote a standard computable bijective pairing function of which the inverse is computable too. Define $< x, y, z >$ inductively by $< x, < y, z >>$.

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

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

This means that $K(x|y)$ is the *minimal* number of bits in a description from which x can be effectively reconstructed, given y. The unconditional complexity is defined as $K(x) = K(x|\epsilon)$. Alternatively, fix a universal partial recursive function ϕ_0, such that $\phi_0(< y, < n, x >>) = \phi_n(< y, x >)$. An equivalent definition, often used, is:

$$K(x|y) = \min\{l(z) : \phi_0(< y, z >) = x\}.$$

A survey is [14].

Throughout 'log' denotes the binary logarithm. We use $f(n) = O(g(n))$ (as $n \rightarrow \infty$) as meaning "there exist two constants C, n_0 such that $|f(n)| \leq C|g(n)|$ for all $n \geq n_0$." When $O(g(n))$ occurs in the middle of a formula it represents a function f satisfying this meaning. We use $f(n) = o(g(n))$ as meaning that $\lim_{n \rightarrow \infty} f(n)/g(n) = 0$.

We need the following properties. For each $x, y \in \mathcal{N}$ we have

$$K(x|y) \leq l(x) + O(1). \tag{1}$$

For each $y \in \mathcal{N}$ there is an x such that $K(x|y) \geq l(x)$. In particular, we can set $y = \epsilon$. Such x's may be called *random*, since they are without regularities which can be used to compress their description: 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

$$K(x|y) \geq n - c. \tag{2}$$

It is not too difficult to show that, if $K(x) \geq n - f(n)$ ($n = l(x)$), then the number of zeros $\#zeros(x)$ it contains is, [15] (or [17] for $f(n) = O(1)$),

$$|\#zeros(x) - n/2| = O(\sqrt{f(n)n}). \tag{3}$$

(If x contains less or more zeros, then it can be described as an element of an ensemble which is significantly smaller than 2^n.)

Denote $K(< x, y >)$ by $K(x, y)$. It can be proved, [11, 14], that, up to a an additive term $O(\log \min\{K(x), K(y)\})$,

$$K(x, y) = K(x) + K(y|x) = K(y) + K(x|y). \tag{4}$$

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. We also denote $K(x| < y, z >)$ by $K(x|y, z)$.

3 Tournaments

The first example proved by Erdös and Spencer in [5] by the probabilistic method, Theorem 1, 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 1 *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 2, fix $T \in \Gamma$ such that

$$K(E(T)|n) \geq l(E(T)). \tag{5}$$

Let S be the transitive subtournament of T on $v(n)$ nodes. We try to 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[1]

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,

$$l(E'(T)) = l(E(T)) - \frac{v(n)}{2}(v(n) - 1 - 2\lceil \log n \rceil). \tag{6}$$

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,

$$K(E(T)|n) \leq l(E'(T)) + O(1). \tag{7}$$

For large enough n, Equations 5, 6, and 7 can only be satisfied with $v(n) \leq 1 + 2\lceil \log n \rceil$. $\square$

The general idea used is the following. [2] 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. It now takes only a few lines to prove the following result with the new method.

[1]Better encoding of S yields a sharper result. But our goal here is to present the simple outlines of an example proof by Kolmogorov complexity.

[2]For each n, define T_n as the Turing machine that on input $E'(T)$ outputs $E(T)$. Define complexity K_{T_n} relative to T_n and repeat the given argument, dispensing with the $O(1)$ error term in Equation 7. This proves Theorem 1 for each n.

Claim 1 *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 $A \times B \subseteq T$. Then $w(n) \leq 2\lceil \log n \rceil$.*

Proof We can save $x = w(n)^2$ bits for the edges between A and B by adding the code of the nodes in A and B in $y = 2w(n)\lceil \log n \rceil$ bits. Since for a tournament satisfying Equation 5 we have $y - x \geq 0$, the claim follows. □

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, [13].

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

Proof. Choose

$$n = 2^k k^2 (\log_e 2 + o(1)). \tag{8}$$

Assume the notation of the previous theorem. By Equation 2, choose T such that

$$K(E(T)|n, k) \geq l(E(T)) = n(n-1)/2. \tag{9}$$

By way of contradiction, assume that $S(k)$ is false for T. Fix a set A of k nodes of T 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,

$$K(E(T)|n, k) \leq l(E'(T)) + O(1). \tag{10}$$

Calculation shows that, for large enough n, Equation 8 is consistent with:

$$l(E'(T)) < l(E(T)) - k^{2-\epsilon}, 0 < \epsilon < 1. \tag{11}$$

Equations 9, 8, 10, 11, yield the desired contradiction. Therefore, $s(k) \leq n$. $\square$

4 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

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

The upper bound was independently established in [16, 1]. Erdös and Rényi, [6], Moser, [18], and Pippenger, [19], have used various methods in combinatorics to show the lower bound. Pippenger used an information theoretic argument. We will supply a proof using Kolmogorov complexity. Fix a standard encoding $E : 2^N \to \mathcal{N}$, 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 3

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

Proof. Use the notation above. By Equations 1, 2, choose M such that

$$K(M|\mathcal{D}) \geq n. \tag{12}$$

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,

$$K(M|\mathcal{D}) \leq K(m_1, \ldots, m_j|\mathcal{D}) + O(1). \tag{13}$$

Let $d_i = |D_i|$, and assume $d_i > \sqrt{n}$. By a standard argument (detailed after the proof), Equation 12 implies that the *randomness deficiency* $k = d_i - K(M \cap D_i|D_i)$ is $O(\log d_i)$. Therefore, by Equation 3, m_i is within range $|m_i - d_i/2| = O(\sqrt{d_i \log d_i})$. Since m_i can be described by its discrepancy with $d_i/2$, and $d_i \leq n$,

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

Pad each description of an m_i to a block of length $\frac{1}{2} \log n + O(\log \log n)$. Then,

$$K(m_1, \ldots, m_j|\mathcal{D}) \leq \sum_{i=1}^{j} (\frac{1}{2} \log n + O(\log \log n)). \tag{14}$$

By Equations 12, 13, and 14, $j \geq n/(\frac{1}{2} \log n + O(\log \log n))$, which is equivalent to the theorem. $\square$

Standard Argument. A useful property states that if an object has maximal complexity, then the complexity of an easily describable part 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, $K(M|\mathcal{D}, i) \leq n - k + O(\log n)$, which by Equation 12 implies that $k \leq K(i) + O(\log n)$. Since $i \leq j$, and $j \leq n$ (the set of singleton sets in N is a distinguishing family), we find $k = O(\log n)$.

5 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, [8]:

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

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 4

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

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

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, $K(x|n, y^x) = O(1)$. Then, by Equation 4, we have:

$$R(x) \stackrel{\text{def}}{=} K(y^x|n) - K(y^x|n,x) - K(x|n) = O(\log K(x|n)). \tag{15}$$

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, $\sum_{x \in N} K(y^x|n,x)$ can be evaluated, from the number of bits in the z^x's, as follows.

$$\sum_{x \in N} |\{i : x \in N - C_i\}| = \sum_{1 \leq i \leq j} |\{x : x \in N - C_i\}| = nj - \sum_{1 \leq i \leq j} c_i. \tag{16}$$

For each x, by Equation 1,

$$K(y^x|n) \leq l(y^x) + O(1) = j + O(1), \tag{17}$$

and $K(x|n) \leq \log n + O(1)$. Estimating the lower bound on $\sum K(x|n)$ by Equation 2,

$$\sum_{x \in N} K(x|n) = n \log n + O(n). \tag{18}$$

By Equations 15, 1, 16, 17, and 18 we have

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

from which the theorem follows. □

One may wonder whether we can remove the $O(\log\log n)$ error term. The prefix variant of complexity $KP(x|y)$, [12, 7, 2] or [14], 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 KP complexity is more precise for some applications. In its KP version, Equation 4 holds to within an $O(1)$ additive term, rather than the $O(\log\log n)$ one, [7]. Then, in Equation 15, the KP version of $R(x) = O(1)$. A straightforward, somewhat tedious, analysis shows that estimates of the quantities in Equations 16, 18, and 17, still hold in KP-version. Together, it follows that $g(n)/n \geq \log n + O(1)$.

6 Expected Properties

By Equation 2, almost all strings have high (Kolmogorov) complexity. Hence, almost all tournaments and, as another example, almost all undirected graphs, have high complexity. Any combinatorial property proven about an arbitrary complex object in such a class will hold for almost all objects in the class. For example, the proof of Theorem 1 does not only show *for large enough n there exists a tournament on n nodes in which all transitive subtournaments have at most $1 + 2\lceil \log n \rceil$ nodes,* but can trivially be strengthened as follows.

By Equation 2 there are at least $2^{n(n-1)/2}(1 - 1/n)$ tournaments T on n nodes with

$$K(E(T)|n) \geq n(n-1)/2 - \log n. \tag{19}$$

This is a $(1 - 1/n)$th fraction of all tournaments on n nodes. Using Equation 19 instead of Equation 5 in the proof of Theorem 1 yields the stronger statement that:

Theorem 5 *For almost all tournaments on n nodes (at least a $(1 - 1/n)$th fraction), the largest transitive subtournament has at most $1+2\lceil \log n \rceil$ nodes, from some n onwards.*

Similarly, choosing $K(E(T)|n, k) \geq n(n-1)/2 - \log n$, instead of Equation 9 in the proof of Theorem 2 yields the stronger result:

Theorem 6 *For all large enough k, there is some n with $n \leq 2^k k^2(\log_e 2 + o(1))$, such that almost all tournaments on n nodes (at least a $(1 - 1/n)$th fraction), have property $S(k)$.*

The Kolmogorov complexity argument generally yields results on *expected* properties rather than worst-case properties, and is especially suited to obtain results on random structures. Since the submission of this paper other such applications (like expected maximum vertex degree of randomly generated trees and a related result on random mappings) have been exhibited, [9]. Bill Gasarch has recently informed us that the method also yields the lower bound in Ramsey's Theorem, although the bound he obtained is not as good as the usual one.

Acknowledgement

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

References

[1] D.G. Cantor and W.H. Mills, Determination of a subset from certain combinatorial properties, *Can. J. Math.*, **18**(1966), 42-48.

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

[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] 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.

[9] W.W. Kirchherr, Kolmogorov complexity and random graphs, *Information Processing Letters*, **41**(1992), 125-130.

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

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

[12] 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.

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

[14] 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.

[15] M. Li and P.M.B. Vitányi, Combinatorial properties of finite sequences with high Kolmogorov complexity, submitted to *Math. Systems Theory.*

[16] B. Lindström, On a combinatorial problem in number theory, *Can. Math. Bull.*, **8**(1965), 477-490.

[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.