Institute for Language, Logic and Information

KOLMOGOROV COMPLEXITY ARGUMENTS IN COMBINATORICS

Ming Li
Paul M.B. Vitanyi

ITLI Prepublication Series
for Computation and Complexity Theory CT-91-01

University of Amsterdam

The ITLI Prepublication Series

1986
86-01 The Institute of Language, Logic and Information
86-02 Peter van Emde Boas — A Semantical Model for Integration and Modularization of Rules
86-03 Johan van Benthem — Categorial Grammar and Lambda Calculus
86-04 Reinhard Muskens — A Relational Formulation of the Theory of Types
86-05 Kenneth A. Bowen, Dick de Jongh — Some Complete Logics for Branched Time, Part I Well-founded Time, Forward looking Operators
86-06 Johan van Benthem — Logical Syntax

1987
87-01 Jeroen Groenendijk, Martin Stokhof — Type shifting Rules and the Semantics of Interrogatives
87-02 Renate Bartsch — Frame Representations and Discourse Representations
87-03 Jan Willem Klop, Roel de Vrijer — Unique Normal Forms for Lambda Calculus with Surjective Pairing
87-04 Johan van Benthem — Polyadic quantifiers
87-05 Víctor Sánchez Valencia — Traditional Logicians and de Morgan's Example
87-06 Eleonore Oversteegen — Temporal Adverbials in the Two Track Theory of Time
87-07 Johan van Benthem — Categorial Grammar and Type Theory
87-08 Renate Bartsch — The Construction of Properties under Perspectives
87-09 Herman Hendriks — Type Change in Semantics: The Scope of Quantification and Coordination

1988

Logic, Semantics and Philosophy of Language:
LP-88-01 Michiel van Lambalgen — Algorithmic Information Theory
LP-88-02 Yde Venema — Expressiveness and Completeness of an Interval Tense Logic
LP-88-03 Year Report 1987
LP-88-04 Reinhard Muskens — Going partial in Montague Grammar
LP-88-05 Johan van Benthem — Logical Constants across Varying Types
LP-88-06 Johan van Benthem — Semantic Parallels in Natural Language and Computation
LP-88-07 Renate Bartsch — Tenses, Aspects, and their Scopes in Discourse
LP-88-08 Jeroen Groenendijk, Martin Stokhof — Context and Information in Dynamic Semantics
LP-88-09 Theo M.V. Janssen — A mathematical model for the CAT framework of Eurotra
LP-88-10 Anneke Kleppe — A Blissymbolics Translation Program

Mathematical Logic and Foundations:
ML-88-01 Jaap van Oosten — Lifschitz' Realizabiility
ML-88-02 M.D.G. Swaen — The Arithmetical Fragment of Martin Löf's Type Theories with weak Σ-elimination
ML-88-03 Dick de Jongh, Frank Veltman — Provability Logics for Relative Interpretability
ML-88-04 A.S. Troelstra — On the Early History of Intuitionistic Logic
ML-88-05 A.S. Troelstra — Remarks on Intuitionism and the Philosophy of Mathematics

Computation and Complexity Theory:
CT-88-01 Ming Li, Paul M.B.Vitanyi — Two Decades of Applied Kolmogorov Complexity
CT-88-02 Michiel H.M. Smid — General Lower Bounds for the Partitioning of Range Trees
CT-88-03 Michiel H.M. Smid, Mark H. Overmars, Leen Torenvliet, Peter van Emde Boas — Maintaining Multiple Representations of Dynamic Data Structures
CT-88-04 Dick de Jongh, Lex Hendriks, Gerard R. Renardel de Lavalette — Computations in Fragments of Intuitionistic Propositional Logic
CT-88-05 Peter van Emde Boas — Machine Models and Simulations (revised version)
CT-88-06 Michiel H.M. Smid — A Data Structure for the Union-find Problem having good Single-Operation Complexity
CT-88-07 Johan van Benthem — Time, Logic and Computation
CT-88-08 Michiel H.M. Smid, Mark H. Overmars, Leen Torenvliet, Peter van Emde Boas — Multiple Representations of Dynamic Data Structures
CT-88-09 Theo M.V. Janssen — Towards a Universal Parsing Algorithm for Functional Grammar
CT-88-10 Edith Spaan, Leen Torenvliet, Peter van Emde Boas — Nondeterminism, Fairness and a Fundamental Analogy
CT-88-11 Sieger van Denneheuvel, Peter van Emde Boas — Towards implementing RL

Other prepublications:
X-88-01 Marc Jumelet — On Solovay's Completeness Theorem

1989

Logic, Semantics and Philosophy of Language:
LP-89-01 Johan van Benthem — The Fine-Structure of Categorial Semantics
LP-89-02 Jeroen Groenendijk, Martin Stokhof — Dynamic Predicate Logic, towards a compositional, non-representational semantics of discourse
LP-89-03 Yde Venema — Two-dimensional Modal Logics for Relation Algebras and Temporal Logic of Intervals
LP-89-04 Johan van Benthem — Language in Action
LP-89-05 Johan van Benthem — Modal Logic as a Theory of Information
LP-89-06 Andreja Prijatelj — Intensional Lambek Calculi: Theory and Application
LP-89-07 Heinrich Wansing — The Adequacy Problem for Sequential Propositional Logic
LP-89-08 Víctor Sánchez Valencia — Peirce's Propositional Logic: From Algebra to Graphs
LP-89-09 Zhisheng Huang — Dependency of Belief in Distributed Systems

Mathematical Logic and Foundations:
ML-89-01 Dick de Jongh, Albert Visser — Explicit Fixed Points for Interpretability Logic
ML-89-02 Roel de Vrijer — Extending the Lambda Calculus with Surjective Pairing is conservative
ML-89-03 Dick de Jongh, Franco Montagna — Rosser Orderings and Free Variables
ML-89-04 Dick de Jongh, Marc Jumelet, Franco Montagna — On the Proof of Solovay's Theorem
ML-89-05 Rineke Verbrugge — Σ-completeness and Bounded Arithmetic
ML-89-06 Michiel van Lambalgen — The Axiomatization of Randomness
ML-89-07 Dirk Roorda — Elementary Inductive Definitions in HA: from Strictly Positive towards Monotone
ML-89-08 Dirk Roorda — Investigations into Classical Linear Logic
ML-89-09 Alessandra Carbone — Provable Fixed points in $I\Delta_0+\Omega_1$

Computation and Complexity Theory:
CT-89-01 Michiel H.M. Smid — Dynamic Deferred Data Structures
CT-89-02 Peter van Emde Boas — Machine Models and Simulations
CT-89-03 Ming Li, Herman Neuféglise, Leen Torenvliet, Peter van Emde Boas — On Space Efficient Simulations
CT-89-04 Harry Buhrman, Leen Torenvliet — A Comparison of Reductions on Nondeterministic Space
CT-89-05 Pieter H. Hartel, Michiel H.M. Smid, Leen Torenvliet, Willem G. Vree — A Parallel Functional Implementation of Range Queries
CT-89-06 H.W. Lenstra, Jr. — Finding Isomorphisms between Finite Fields
CT-89-07 Ming Li, Paul M.B. Vitanyi — A Theory of Learning Simple Concepts under Simple Distributions and Average Case Complexity for the Universal Distribution (Prel. Version)
CT-89-08 Harry Buhrman, Steven Homer, Leen Torenvliet — Honest Reductions, Completeness and Nondeterminstic Complexity Classes
CT-89-09 Harry Buhrman, Edith Spaan, Leen Torenvliet — On Adaptive Resource Bounded Computations
CT-89-10 Sieger van Denneheuvel — The Rule Language RL/1
CT-89-11 Zhisheng Huang, Sieger van Denneheuvel, Peter van Emde Boas — Towards Functional Classification of Recursive Query Processing

Other Prepublications:
X-89-01 Marianne Kalsbeek — An Orey Sentence for Predicative Arithmetic
X-89-02 G. Wagemakers — New Foundations: a Survey of Quine's Set Theory
X-89-03 A.S. Troelstra — Index of the Heyting Nachlass
X-89-04 Jeroen Groenendijk, Martin Stokhof — Dynamic Montague Grammar, a first sketch
X-89-05 Maarten de Rijke — The Modal Theory of Inequality
X-89-06 Peter van Emde Boas — Een Relationele Semantiek voor Conceptueel Modelleren: Het RL-project

1990 SEE INSIDE BACK COVER

Instituut voor Taal, Logica en Informatie
Institute for Language, Logic and Information

Faculteit der Wiskunde en Informatica
(Department of Mathematics and Computer Science)
Plantage Muidergracht 24
1018TV Amsterdam

Faculteit der Wijsbegeerte
(Department of Philosophy)
Nieuwe Doelenstraat 15
1012CP Amsterdam

KOLMOGOROV COMPLEXITY ARGUMENTS IN COMBINATORICS

Ming Li
Computer Science Department
University of Waterloo
Paul M.B. Vitanyi
Department of Mathematics and Computer Science
University of Amsterdam
& CWI

ITLI Prepublication Series
for Computation and Complexity Theory
ISSN 0924-8374

Received March 1991

Kolmogorov Complexity Arguments in Combinatorics

Ming Li*
University of Waterloo

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

February 27, 1991

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 [4], are usually aimed at establishing the existence of an object, in a non-constructive sense. It is assertained 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

*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. Address: Centrum voor Wiskunde en Informatica, Kruislaan 413, 1098 SJ Amsterdam, The Netherlands. Email: paulv@cwi.nl

prove such 'existential' propositions often uses averages. We may call this 'first-moment' methods. 'Second-moment' methods, using means and variance of random variables to establish combinatorial results have been used by Moser [14]. Pippenger [15], has used related notions like 'entropy', 'self-information', and 'mutual information', from information theory, [17]. 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 [8], Kolmogorov established a notion of complexity (self-information) of finite objects which is essentially finitary and combinatorial. Says Kolmogorov [9]: "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 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 [16]. 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 [12]. 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

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

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 $K(x) \geq n + O(\log n)$ ($n = l(x)$), then the number of zeros it contains is, [13],

$$n/2 + O(\sqrt{n}). \tag{3}$$

(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, [9, 12], 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 [4] by the probabilistic method, Theorem 1, is originally due to Erdös and Moser [3]. (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 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,

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

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. [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 $w(n) \leq 2\lceil \log n \rceil$.

The second example is Theorem 9.1 in [4], originally due to Erdös [2]. 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, [11].

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

[1] 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.

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

$$K(E(T)|n,k) \geq l(E(T)) = n(n-1)/2. \tag{8}$$

Assume that $S(k)$ is false for

$$n = 2^k k^2(\log_e 2 + o(1)). \tag{9}$$

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,

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

Calculation shows that, for large enough n, Equation 9 is consistent with:

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

Equations 8, 9, 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

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

Erdös and Rényi, [5], Moser, [14], and Pippenger, [15], 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 $E : 2^N \to \mathcal{N}$, 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

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

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 $\frac{d_i}{2} + O(\sqrt{d_i})$. Since m_i can be described by its discrepancy with $d_i/2$, and $d_i \leq n$,

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

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

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

By Equations 12, 13, and 14, $j \geq n/(\frac{1}{2}\log n + O(1))$, 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 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, [7]:

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

The lower bound on $g(n)$ was also proven by Pippenger, [15], 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:

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

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.

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

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 $KP(x|y)$, [10, 6, 1] or [12], 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, [6]. 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 $g(n)/n \geq \log n + O(1)$.

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

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

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

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

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

[7] 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.

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

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

[10] 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.

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

[12] 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.

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

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

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

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

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

The ITLI Prepublication Series

1990

Logic, Semantics and Philosophy of Language

LP-90-01 Jaap van der Does	A Generalized Quantifier Logic for Naked Infinitives
LP-90-02 Jeroen Groenendijk, Martin Stokhof	Dynamic Montague Grammar
LP-90-03 Renate Bartsch	Concept Formation and Concept Composition
LP-90-04 Aarne Ranta	Intuitionistic Categorial Grammar
LP-90-05 Patrick Blackburn	Nominal Tense Logic
LP-90-06 Gennaro Chierchia	The Variablity of Impersonal Subjects
LP-90-07 Gennaro Chierchia	Anaphora and Dynamic Logic
LP-90-08 Herman Hendriks	Flexible Montague Grammar
LP-90-09 Paul Dekker	The Scope of Negation in Discourse, towards a flexible dynamic Montague grammar
LP-90-10 Theo M.V. Janssen	Models for Discourse Markers
LP-90-11 Johan van Benthem	General Dynamics
LP-90-12 Serge Lapierre	A Functional Partial Semantics for Intensional Logic
LP-90-13 Zhisheng Huang	Logics for Belief Dependence
LP-90-14 Jeroen Groenendijk, Martin Stokhof	Two Theories of Dynamic Semantics
LP-90-15 Maarten de Rijke	The Modal Logic of Inequality
LP-90-16 Zhisheng Huang, Karen Kwast	Awareness, Negation and Logical Omniscience
LP-90-17 Paul Dekker	Existential Disclosure, Implicit Arguments in Dynamic Semantics

Mathematical Logic and Foundations

ML-90-01 Harold Schellinx	Isomorphisms and Non-Isomorphisms of Graph Models
ML-90-02 Jaap van Oosten	A Semantical Proof of De Jongh's Theorem
ML-90-03 Yde Venema	Relational Games
ML-90-04 Maarten de Rijke	Unary Interpretability Logic
ML-90-05 Domenico Zambella	Sequences with Simple Initial Segments
ML-90-06 Jaap van Oosten	Extension of Lifschitz' Realizability to Higher Order Arithmetic, and a Solution to a Problem of F. Richman
ML-90-07 Maarten de Rijke	A Note on the Interpretability Logic of Finitely Axiomatized Theories
ML-90-08 Harold Schellinx	Some Syntactical Observations on Linear Logic
ML-90-09 Dick de Jongh, Duccio Pianigiani	Solution of a Problem of David Guaspari
ML-90-10 Michiel van Lambalgen	Randomness in Set Theory
ML-90-11 Paul C. Gilmore	The Consistency of an Extended NaDSet

Computation and Complexity Theory

CT-90-01 John Tromp, Peter van Emde Boas	Associative Storage Modification Machines
CT-90-02 Sieger van Denneheuvel, Gerard R. Renardel de Lavalette	A Normal Form for PCSJ Expressions
CT-90-03 Ricard Gavaldà, Leen Torenvliet, Osamu Watanabe, José L. Balcázar	Generalized Kolmogorov Complexity in Relativized Separations
CT-90-04 Harry Buhrman, Edith Spaan, Leen Torenvliet	Bounded Reductions
CT-90-05 Sieger van Denneheuvel, Karen Kwast	Efficient Normalization of Database and Constraint Expressions
CT-90-06 Michiel Smid, Peter van Emde Boas	Dynamic Data Structures on Multiple Storage Media, a Tutorial
CT-90-07 Kees Doets	Greatest Fixed Points of Logic Programs
CT-90-08 Fred de Geus, Ernest Rotterdam, Sieger van Denneheuvel, Peter van Emde Boas	Physiological Modelling using RL
CT-90-09 Roel de Vrijer	Unique Normal Forms for Combinatory Logic with Parallel Conditional, a case study in conditional rewriting

Other Prepublications

X-90-01 A.S. Troelstra	Remarks on Intuitionism and the Philosophy of Mathematics, Revised Version
X-90-02 Maarten de Rijke	Some Chapters on Interpretability Logic
X-90-03 L.D. Beklemishev	On the Complexity of Arithmetical Interpretations of Modal Formulae
X-90-04	Annual Report 1989
X-90-05 Valentin Shehtman	Derived Sets in Euclidean Spaces and Modal Logic
X-90-06 Valentin Goranko, Solomon Passy	Using the Universal Modality: Gains and Questions
X-90-07 V.Yu. Shavrukov	The Lindenbaum Fixed Point Algebra is Undecidable
X-90-08 L.D. Beklemishev	Provability Logics for Natural Turing Progressions of Arithmetical Theories
X-90-09 V.Yu. Shavrukov	On Rosser's Provability Predicate
X-90-10 Sieger van Denneheuvel, Peter van Emde Boas	An Overview of the Rule Language RL/1
X-90-11 Alessandra Carbone	Provable Fixed points in $I\Delta_0+\Omega_1$, revised version
X-90-12 Maarten de Rijke	Bi-Unary Interpretability Logic
X-90-13 K.N. Ignatiev	Dzhaparidze's Polymodal Logic: Arithmetical Completeness, Fixed Point Property, Craig's Property
X-90-14 L.A. Chagrova	Undecidable Problems in Correspondence Theory
X-90-15 A.S. Troelstra	Lectures on Linear Logic

1991

Mathematical Logic and Foundations

ML-91-01 Yde Venema	Cylindric Modal Logic
ML-91-02 Alessandro Berarducci, Rineke Verbrugge	On the Metamathematics of Weak Theories
ML-91-03 Domenico Zambella	On the Proofs of Arithmetical Completeness for Interpretability Logic

Computation and Complexity Theory

CT-91-01 Min Li, Paul M.B. Vitanyi	Kolmogorov Complexity Arguments in Combinatorics
CT-91-02 Min Li, John Tromp, Paul M.B. Vitanyi	How to Share Concurrent Wait-Free Variables
CT-91-03 Min Li, Paul M.B. Vitanyi	Average Case Complexity under the Universal Distribution Equals Worst Case Complexity

Other Prepublications

X-91-01 Alexander Chagrov, Michael Zakharyaschev	The Disjunction Property of Intermediate Propositional Logics
X-91-02 Alexander Chagrov, Michael Zakharyaschev	On the Undecidability of the Disjunction Property of Intermediate Propositional Logics
X-91-03 V. Yu. Shavrukov	Subalgebras of Diagonizable Algebras of Theories containing Arithmetic
X-91-04 K.N. Ignatiev	Partial Conservativity and Modal Logics
X-91-05 Johan van Benthem	Temporal Logic