

Kolmogorov complexity and algorithmic randomness, by Alexander Shen, Vladimir A. Uspensky, and Nikolay K. Vereshchagin, Mathematical Surveys and Monographs, Vol. 220, American Mathematical Society, Providence, RI, 2017, xviii+511 pp., ISBN 9781470431822

1. INTRODUCTION

As young children we all quickly develop a rudimentary intuition for gravity. But it is not until we have played games of chance—and perhaps not until we have lost in a rigged game—that we have some inkling of what randomness is. To be fair, one eventually learns in school that gravity has at least as many mysteries as the definition of randomness. But the question remains: What is randomness?

A mathematically satisfying answer can be found in the book under review (henceforth [SUV], from the authors’ initials). [SUV] presents a rich account of numerous earlier definitions of randomness, their interconnection with algorithmic complexity, and their applications. Before exploring [SUV] further, let us first see some concrete examples of the boundary between randomness and nonrandomness.

2. ADVERSARIES AND PSEUDORANDOM GENERATORS

If one makes a yes/no decision via a coin-flip while standing in the middle of a gusting windstorm, then one may be inclined to believe that this decision method is random, and that either outcome is equally likely. On the other hand, if an experienced gambler picks your coin and tosses it for you in his private casino, then you would be less inclined to believe that the outcome will be random. The key difference between these two coin-tossing scenarios is *predictability*: One can argue that the necessary measurements and calculations to predict the coin-flip in the first scenario make prediction completely impractical. On the other hand, making predictable coin-tosses in controlled settings has been known long enough in the gambling world that mathematicians have finally written papers on it (see, e.g., [DHM07]). Our two toy scenarios (inspired by a similar discussion in [BM84]) thus suggest that a suitable definition of randomness should include some notion of unpredictability.

Our two toy scenarios also hint at the deep connection between randomness and physics. (See [Rue93] for a beautiful exposition on the connections between dynamics, physics, and randomness.) Furthermore, it becomes apparent that computational power is a factor behind predicting (or *derandomizing*) a sequence of events that is intended to be random. More to the point, we can use computational complexity to define *pseudorandomness*, and perhaps sidestep the question of how to reliably extract randomness from some physical process.

The algorithmic construction of unpredictable sequences of bits has numerous applications, especially if certain natural efficiency and equidistribution properties are also guaranteed. For instance, in cryptography, such sequences can be used¹ to

2010 *Mathematics Subject Classification*. Primary 68-XX, 03-XX, 60-XX, 97-XX.

The reviewer is partially supported by NSF grants CCF-1409020 and CCF-1900881, and by NSF REU grant DMS-1757872.

¹via the ancient idea of a *one-time pad*

safely send messages that need to be kept secret from an adversary. Also, in some statistical applications, the only way to practically estimate the mean of a quantity over a large population is to average over a smaller, well-distributed sample. Pseudorandomness attempts to capture the most useful aspects of randomness in an algorithmically efficient way, but its definition is best preceded with an explicit example.

2.1. Pseudorandom bit streams from the Discrete Logarithm Problem.

Consider the following famous construction, by Blum and Micali [BM84], of a family of (putatively) unpredictable sequences of bits. In what follows, n is a positive integer, $p \in \{2^{n-1} + 1, \dots, 2^n - 1\}$ is an odd prime (so p has exactly n bits in its binary expansion), g is a generator for the multiplicative group \mathbb{F}_p^* of nonzero integers mod p , and $x_0 \in \{0, \dots, p-2\}$ is called the *seed* for our bit stream. Letting $M_p : \{1, \dots, p-1\} \rightarrow \{0, 1\}$ denote the function satisfying $M_p(a) = 1$ if and only if $a \geq (p-1)/2$ (so $M_p(a)$ is akin to the most significant bit of a), we then define a sequence (x_0, x_1, \dots, x_N) with, say, $N := n^{10}$, via the recurrence $x_{j+1} := g^{x_j} \pmod p$, valid for all $j \geq 0$.² Our pseudorandom sequence of bits—an instance of the *Blum–Micali pseudorandom generator (PRG)*—is then

$$B(p, g, x_0) := (M_p(x_N), M_p(x_{N-1}), \dots, M_p(x_1)).$$

Letting $(b_1, \dots, b_N) := B(p, g, x_0)$, a remarkable property of the Blum–Micali PRG is that it can be *proved* that b_{i+1} is unpredictable from b_i (and even x_{N-i+1}), provided a well-known number-theoretic problem is hard to solve. To properly define unpredictability and hardness, let us first recall the following problem:

The Discrete Logarithm Problem (DLP). Given an n -bit prime p , a generator g for the multiplicative group \mathbb{F}_p^* , and an integer $h \in \{1, \dots, p-1\}$, find an $\ell \in \{0, \dots, p-2\}$ with $g^\ell = h \pmod p$. \diamond

We call the ℓ in the definition above the *mod p base- g discrete logarithm of h* , and we think of n as a rough measure of the size of an instance (p, g, h) of DLP. In what follows, we will use the usual number-theoretic/computer science notations $O(\cdot)$, $o(\cdot)$, $\Omega(\cdot)$, and $\omega(\cdot)$ for asymptotic growth comparisons (see, e.g., [AB09, pp. 3–4]). For instance, if we say that a function $f : \mathbb{N} \rightarrow \mathbb{N}$ satisfies $f(n) = n^{\omega(1)}$, then this means that for any real c , there is an $n_0(c) \in \mathbb{N}$ such that $f(n) \geq n^c$ for all $n \geq n_0(c)$, i.e., f ultimately grows faster than any polynomial.

It is easy to compute $h = g^\ell \pmod p$, given (p, g, ℓ) as above, in near-quadratic time: $n^{2+o(1)}$ bit operations suffice (see, e.g., [BS96, pp. 43, 102–104]). However, computing the discrete logarithm ℓ from (p, g, h) in time $n^{O(1)}$ remains an open problem: The best general complexity bound is $e^{O(n^{1/3}(\log n)^{2/3})}$, via refinements of the index calculus method (see, e.g., [HPS14, Sec. 3.8] and [A+19]), and the underlying algorithm uses randomization. In particular, DLP is conjectured to be hard in the following sense:

DLP Hardness Assumption. For any $k, n \in \mathbb{N}$, let $C_{n,k}$ be any boolean circuit that solves DLP for at least $\frac{1}{n^k}$ of the n -bit primes p . Then the size of $C_{n,k}$ is $n^{\omega(1)}$. \diamond

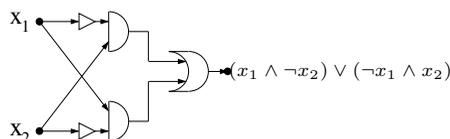
²The length $N = n^{10}$ is somewhat arbitrary: One can in fact use $N = n^k$ for any positive integer k [Blu19], although this was not explicitly stated in [BM84].

Basic complexity theory (see, e.g., [AB09, Ch. 6]) tells us that the existence of an algorithm for solving DLP in time polynomial in n would imply that there is a family of boolean circuits $\{C_n\}_{n \in \mathbb{N}}$ such that C_n solves DLP for any input (p, g, h) , with p an n -bit prime, and C_n has size $n^{O(1)}$. So the DLP Hardness Assumption rules out such a family of circuits, *as well as* polynomial-time algorithms for DLP in the classical Turing model.³

To better understand this hardness assumption, recall that a *boolean circuit* C is an acyclic directed graph with nodes (also called *gates*) consisting of one of five types:

- (0) *input* nodes (having in-degree 0 and finite out-degree);
- (1) *output* nodes (having in-degree 1 and out-degree 0);
- (2) a *not* gate \neg (having in-degree 1 and out-degree 1);
- (3) an *and* gate \wedge (having in-degree 2 and out-degree 1); and
- (4) an *or* gate \vee (having in-degree 2 and out-degree 1).

The *size* of C is simply the number of vertices in the underlying graph, and the *number of inputs* is the number of input nodes. Identifying 1 and 0 with “true” and “false”, respectively, and labelling the input nodes with variables, we then obtain a natural interpretation of boolean circuits as implementations of logical formulae. For instance, the circuit below (with two input nodes and size 8) computes the well-known XOR function $(x_1 \vee \neg x_2) \wedge (\neg x_1 \vee x_2)$.



Abusing notation slightly, we then say that a function

$$F : \{0, \dots, 2^n - 1\} \longrightarrow \{0, \dots, 2^m - 1\}$$

is *computed by a boolean circuit* if and only if there are boolean circuits computing each of the m bits of $F(x)$ from the binary expansion of $x \in \{0, \dots, 2^n - 1\}$.

The (conditional) unpredictability of the Blum–Micali PRG can then be phrased as follows:

Blum–Micali Theorem ([BM84]). *Assume the DLP Hardness Assumption is true, and let $\mathcal{C} := \{C_n\}$ be any family of boolean circuits, with $n \in \mathbb{N}$ and C_n having exactly $3n$ inputs and size $n^{O(1)}$. Considering the uniform distribution over the set of all (p, g, h) with p an n -bit prime, g a generator for \mathbb{F}_p^* , and $h \in \mathbb{F}_p^*$, let $\rho_n^{\mathcal{C}}$ be the probability that $C_n(p, g, h) = M_p(\ell)$, where $h = g^\ell \pmod p$. Then $\rho_n^{\mathcal{C}} < \frac{1}{2} + \frac{1}{n^{\omega(1)}}$.*

In other words, if DLP is hard, then no (classical) polynomial-time algorithm can do much better than coin-tossing to guess the $(i + 1)$ st bit of an instance of the Blum–Micali PRG, even if one knows p , g , and the element of \mathbb{F}_p^* whose discrete logarithm yielded the i th bit. Curiously, as of June 2019, there is still no known polynomial-time construction that attains unpredictability in the preceding sense *without* unproven hypotheses.

³One motivation for stating the hardness of DLP more stringently, in terms of families of circuits, comes from cryptography: One wants PRGs that evade prediction not just by polynomial-time algorithms but also by special purpose hardware.

2.2. A rigorous definition of pseudorandomness. The Blum–Micali PRG, under the DLP Hardness Assumption, is actually an instance of a more general construction by Blum and Micali [BM84]. Following [BM84]⁴ and a refinement from [Nis92a], one can define a more general class of PRGs as follows:

Definition 2.1. A *pseudorandom generator (PRG)* for class Γ is a family of functions $G := \{G_n : \{0, 1\}^n \rightarrow \{0, 1\}^{Q(n)}\}_{n \in \mathbb{N}}$ such that $Q : \mathbb{N} \rightarrow \mathbb{N}$ is an increasing function and, writing $(y_1, \dots, y_{Q(n)}) := G_n(x_1, \dots, x_n)$, satisfies the following property: For each $i \in \{0, \dots, Q(n) - 1\}$ and every algorithm A in Γ , we have

$$\left| \text{Prob}_{(x_1, \dots, x_n) \in \{0, 1\}^n} [A(y_1, \dots, y_i) = y_{i+1}] - \frac{1}{2} \right| = \frac{1}{n^{\omega(1)}}. \quad \diamond$$

Remark 2.1. The function $\frac{Q(n)}{n}$ (which is bounded from below by 1, thanks to our definition) is sometimes called the *stretch-factor* of the PRG and is a useful measure of how the PRG *amplifies* randomness, i.e., simulates a long stream of uniformly random bits from just a small number of “truly” uniformly random bits. Turning this around, we can also view a PRG as a means of *compressing* the sequence $(y_1, \dots, y_{Q(n)})$ into the shorter representation (x_1, \dots, x_n) . This leads to the notion of *Kolmogorov complexity*, which is a central theme in the book under review. \diamond

The *existence* of PRGs (for the complexity classes $\mathbf{P}_{/\text{poly}}$ or \mathbf{P}) with polynomial stretch-factor can be proved quite easily. Technically, this is a rephrasing of the classical containment of complexity classes $\mathbf{BPP} \subseteq \mathbf{P}_{/\text{poly}}$ (see, e.g., [Adl78, BG81] and [AB09, Sec. 7.5]). Much how probabilistic methods yield nonconstructive proofs in combinatorics, there are still no proven *explicit* construction of such PRGs, unless one uses unproved hypotheses (like our earlier Blum–Micali example). The subtlety of finding explicit PRGs is also revealed by how the famous $\mathbf{P} \stackrel{?}{=} \mathbf{BPP}$ question from complexity theory [IW97, IK04] can be reformulated as the construction of a PRG for \mathbf{P} with exponential stretch-factor (and the complexity of G_n , say, exponential in n) [Gol11]. One should also be aware of the fact that the existence of PRGs for \mathbf{P} , with stretch-factor $n^{\omega(1)}$ and G_n computable in time polynomial in n , implies an even more famous conjecture: $\mathbf{P} \neq \mathbf{NP}$ (see, e.g., [Vad11, Sec. 7.2]).

In closing our discussion of pseudorandomness, we point out that there has been important recent progress on explicit constructions of PRGs (for complexity classes *lower* than \mathbf{P}) with exponential stretch; see, e.g., [Nis92b, MZ13, GKM18]. Further background on pseudorandomness can be found in [Yao82, Kab04, Vad11].

3. THE BOOK OF SHEN, USPENSKY, AND VERESHCHAGIN

3.1. Two initial definitions for randomness. Since *pseudorandom* bit-streams can be defined by unpredictability against an adversary with computational power determined by a complexity class Γ , why not try to define randomness by letting $\Gamma = \mathbf{RE}$?⁵ Alternatively, why not define randomness in terms of equidistribution of the 0’s and 1’s within subsequences? These two approaches, which we will call

⁴To be more precise, the definition from [BM84] took $\Gamma = \mathbf{P}_{/\text{poly}}$ and assumed that each function G_n is computable in time polynomial in n . Here we make no restriction on the complexity of computing the G_n .

⁵ \mathbf{RE} is the class of decision problems where a “yes” answer can be verified by a Turing machine in finite time. (So the machine might never halt if the answer is “no”.) In other words, a PRG for class \mathbf{RE} has the property that even unlimited computational time makes the next bit no more predictable than a fair coin-toss.

try to *compress* our sequence $(b_i)_{i \in \{1, \dots, n\}}$ by replacing it with a program of *minimal length* that generates $(b_i)_{i \in \{1, \dots, n\}}$. One can then consider the finite sequence $(b_i)_{i \in \mathbb{N}}$ to be far from random if and only if it is not compressible in the preceding sense. The Linux command `gzip` is an example of this kind of compression: One can check on any recent version of Linux that the first two examples can be compressed by a factor of around one million, thus suggesting that our first two examples are far from random.⁸

Needless to say, our preceding description is far from rigorous. However, the key insight derived independently by Solomonoff, Kolmogorov, and Chaitin is that one *can* formalize the notion of compressibility and can do so in an *invariant* way. More precisely, consider any partial computable function $D : \{0, 1\}^* \rightarrow \{0, 1\}^*$, where $\{0, 1\}^*$ denotes the set of all finite binary strings, and *partial computable* means that D is computable by a Turing machine that may not terminate on certain inputs. The *Kolmogorov complexity* of an $x \in \{0, 1\}^*$, with respect to the *decompressor* D , is then defined to be

$$C_D(x) := \min\{L(y) \mid D(y) = x, y \in \{0, 1\}^*\},$$

where $L(y)$ denotes the length of the string $y \in \{0, 1\}^*$. In particular, Solomonoff and Kolmogorov proved independently that there are *optimal* decompressors D in the sense that for all $x \in \{0, 1\}^*$ and all decompressors D' , we have $C_D(x) \leq C_{D'}(x) + O(1)$. In other words, we can fix some optimal decompressor D once and for all and write Kolmogorov complexity without mentioning any decompressor by setting $C(x) := C_D(x)$.

A surprising aspect of Kolmogorov complexity is that it is simultaneously hard to compute but possesses numerous natural functorial properties. For example, while it is easy to show that $C(x) \leq n + O(1)$ for all $x \in \{0, 1\}^n$, the length of the *shortest* bit-stream x with $C(x) = n$ is a Turing-uncomputable function of n (see [SUV, Theorem 15]). On the other hand, one can naturally conjecture that Kolmogorov complexity is subadditive with respect to concatenation of strings. And, indeed, writing xy for the concatenation of two bit-streams x and y , one has a property close to subadditivity: $C(xy) \leq C(x) + C(y) + 2 \log_2(C(x)) + O(1)$ (see [SUV, Theorem 4]).

More importantly, Kolmogorov complexity leads us to new perspectives not just on randomness but also on information theory, algorithmic complexity, and probability theory. Indeed, Chapter 8 is a tasty dessert in the middle of [SUV] highlighting important applications of Kolmogorov complexity in several other areas of mathematics.

3.3. Final comments. The account of Kolmogorov complexity and randomness in [SUV] is masterful. Perhaps the first question that comes to mind, if deciding whether to purchase [SUV], is how it compares to the classic text by Li and Vitányi [LV08]. My answer would be to buy both books: While [LV08] has the advantage of additional polishing (having gone through three editions in 14 years), [SUV] maintains amazing clarity while getting quickly to the heart of almost everything one needs from Kolmogorov complexity and its variants. The exercises in [SUV] are also elegantly designed, well accented by hints, and nicely amplify the development.

⁸The same can be said for our third example, and we leave the precise compression factor as an exercise.

ACKNOWLEDGMENTS

I am indebted to Manuel Blum, Qi Cheng, Eviatar Procaccia, Salil Vadhan, and David Zuckerman for their patience with, and answers to, my many naive questions.

REFERENCES

- [Adl78] L. Adleman, *Two theorems on random polynomial time*, 19th Annual Symposium on Foundations of Computer Science (Ann Arbor, Mich., 1978), IEEE, Long Beach, Calif., 1978, pp. 75–83. MR539832
- [A+19] D. Adrian, K. Bhargavan, Z. Durumeric, P. Gaudry, M. Green, J. A. Halderman, N. Heninger, D. Springall, E. Thomé, L. Valenta, B. VanderSloot, E. Wustrow, S. Zanella-Béguelin, and P. Zimmermann, *Imperfect forward secrecy: how Diffie-Hellman fails in practice*, Comm. Amer. Comput. Mach., January 2019, Vol. 62, No. 1, pp. 106–114. doi 10.1145/3292035
- [AB09] S. Arora and B. Barak, *Computational complexity*, Cambridge University Press, Cambridge, 2009. A modern approach. MR2500087
- [BS96] E. Bach and J. Shallit, *Algorithmic number theory. Vol. 1*, Foundations of Computing Series, MIT Press, Cambridge, MA, 1996. Efficient algorithms. MR1406794
- [BG81] C. H. Bennett and J. Gill, *Relative to a random oracle A , $\mathbf{P}^A \neq \mathbf{NP}^A \neq \text{co} - \mathbf{NP}^A$ with probability 1*, SIAM J. Comput. **10** (1981), no. 1, 96–113, DOI 10.1137/0210008. MR605606
- [Blu19] M. Blum, *e-mail communication*, March 23, 2019.
- [BM84] M. Blum and S. Micali, *How to generate cryptographically strong sequences of pseudorandom bits*, SIAM J. Comput. **13** (1984), no. 4, 850–864, DOI 10.1137/0213053. MR764183
- [DHM07] P. Diaconis, S. Holmes, and R. Montgomery, *Dynamical bias in the coin toss*, SIAM Rev. **49** (2007), no. 2, 211–235, DOI 10.1137/S0036144504446436. MR2327054
- [Gol11] O. Goldreich, *In a world of $\mathbf{P} = \mathbf{BPP}$* , Studies in Complexity and Cryptography, Lecture Notes in Comput. Sci., vol. 6650, Springer, Heidelberg, 2011, pp. 191–232, DOI 10.1007/978-3-642-22670-0_20. MR2844264
- [GKM18] P. Gopalan, D. M. Kane, and R. Meka, *Pseudorandomness via the discrete Fourier transform*, SIAM J. Comput. **47** (2018), no. 6, 2451–2487, DOI 10.1137/16M1062132. MR3892445
- [HPS14] J. Hoffstein, J. Pipher, and J. H. Silverman, *An introduction to mathematical cryptography*, 2nd ed., Undergraduate Texts in Mathematics, Springer, New York, 2014. MR3289167
- [IK04] V. Kabanets and R. Impagliazzo, *Derandomizing polynomial identity tests means proving circuit lower bounds*, Comput. Complexity **13** (2004), no. 1-2, 1–46, DOI 10.1007/s00037-004-0182-6. MR2105971
- [IW97] R. Impagliazzo and A. Wigderson, *$\mathbf{P} = \mathbf{BPP}$ unless E has subexponential circuits: derandomizing the XOR Lemma*, Proceedings of the 29th STOC, pp. 220–229, 1997.
- [Kab04] V. Kabanets, *Derandomization: a brief overview*, in Current Trends in Theoretical Computer Science: The Challenge of the New Century, Vol 1: Algorithms and Complexity, edited by G. Paun, G. Rosenberg, and A. Salomaa, World Scientific Publishing Company, 2004.
- [LV08] M. Li and P. Vitányi, *An introduction to Kolmogorov complexity and its applications*, 3rd ed., Texts in Computer Science, Springer, New York, 2008. MR2494387
- [MZ13] R. Meka and D. Zuckerman, *Pseudorandom generators for polynomial threshold functions*, SIAM J. Comput. **42** (2013), no. 3, 1275–1301, DOI 10.1137/100811623. MR3504629
- [Nis92a] N. Nisan, *Using hard problems to create pseudorandom generators*, ACM Distinguished Dissertations, MIT Press, Cambridge, MA, 1992. MR1149128
- [Nis92b] N. Nisan, *Pseudorandom generators for space-bounded computation*, Combinatorica **12** (1992), no. 4, 449–461, DOI 10.1007/BF01305237. MR1194735
- [Rue93] D. Ruelle, *Chance and chaos*, Princeton University Press, Princeton, NJ, 1991. MR1145490

- [SUV] A. Shen, V. A. Uspensky, and N. Vereshchagin, *Kolmogorov complexity and algorithmic randomness*, Mathematical Surveys and Monographs, vol. 220, American Mathematical Society, Providence, RI, 2017. MR3702040
- [Vad11] S. P. Vadhan, *Pseudorandomness*, Found. Trends Theor. Comput. Sci. **7** (2011), no. 1-3, front matter, 1–336, DOI 10.1561/0400000010. MR3019182
- [Yao82] A. C. Yao, *Theory and applications of trapdoor functions*, 23rd Annual Symposium on Foundations of Computer Science (Chicago, Ill., 1982), IEEE, New York, 1982, pp. 80–91. MR780384

J. MAURICE ROJAS
TEXAS A&M UNIVERSITY
COLLEGE STATION, TEXAS
Email address: rojas@math.tamu.edu