MATHEMATICAL FOUNDATIONS OF RANDOMNESS

Abhijit Dasgupta

This article is dedicated to the centenary of the Borel Strong Law.

1 INTRODUCTION

1.1 A random blackbox?

Imagine a "blackbox" which supposedly produces its outcomes "randomly" according to some fixed finite probability distribution.

$$\boxed{\text{BLACKBOX}} \rightarrow \text{Outcome}$$

Thus there are a finite number of possible outcomes, say $\omega_1, \omega_2, \ldots, \omega_n$, and each outcome ω_i has a fixed non-trivial probability p_i, so that $P(\omega_i) = p_i$ with $0 < p_i < 1$ for $i = 1, 2, \ldots, n$, and $\sum_{i=1}^{n} p_i = 1$. The outcomes may be generated either automatically from a continuously running process, or on demand, say by pressing a button on the blackbox.

We think of this as an abstract model representing a process, a machine, or an experiment. Familiar examples are the flip of a coin, the turn of a casino gambling wheel, the roll of an "electronic die" on a handheld video game device, the snapshot of weather data, the time between two successive clicks of a Geiger counter detecting radioactive decay, the stock market index value, etc.

Here, we approach the problem in a *purely extensional way.* This means that except for the given knowledge of the probability values of for each possible outcome, we are only able to *observe* the outcomes, and do not have any access to, or information about, the internal workings of the machine. Hence the term *blackbox.*

Gambling houses and forecasters of weather and stock market as well as philosophers of probability and statistics have found the following question to be of considerable interest.

> **Question A.** Is the blackbox a *random* device? Does it produce its outcomes *randomly* (while obeying a fixed probability distribution)?

Handbook of the Philosophy of Science. Volume 7: Philosophy of Statistics.
Volume editors: Prasanta S. Bandyopadhyay and Malcolm R. Forster. General Editors: Dov M. Gabbay, Paul Thagard and John Woods.
© 2010 Elsevier BV. All rights reserved.

1.2 Sequences

We cannot hope to answer the last question by observing a single outcome of the blackbox. In fact, no finite amount of observation of outcomes can fully confirm that a process is random (or not). On the other hand, by repeating the process a sufficiently large number of times and observing the resulting *sequence of outcomes,* we may hope to gain enough information for answering the question with a desired level of confidence. (Such a sequence is denoted by $\langle x_1, x_2, \ldots, x_k, \ldots \rangle$, where each x_k, the k-th term of the sequence, equals one of the outcome values $\omega_1, \ldots, \omega_n$.)

Given a sequence of outcomes of the blackbox, we want to determine whether or not it was produced randomly. Our extensional approach means that other than a knowledge of the underlying probability distribution for the possible outcome values, we only have the completed sequence of outcomes available, with all information about its origination permanently removed, so the answer must depend solely on the sequence itself and not on how it was produced. This necessitates the consideration of a different but more precise question:

> **Question B.** When is a given sequence of outcomes random (relative to a fixed probability distribution for the outcome values)?

This article deals entirely with this last question (Question B, randomness of sequences), and not with the earlier question (Question A, randomness of processes). It is one of the fundamental questions in the philosophy of probability and statistics. The longer the sequence, the more information we have for determining if it is random or not. And in the ideal case, the sequences will be infinite sequences. We will consider both cases — finite and infinite sequences — of the question in detail.

It was von Mises who first treated this question rigorously (for infinite sequences), and considered its answer to be the very foundation of probability, known as the *frequentist* interpretation of probability.

1.3 Pseudo-randomness: A Galilean Dialogue

In his celebrated work *Gödel, Escher, Bach,* Douglas Hofstadter quotes the following "beautiful and memorable passage" from *Are Quanta Real? — a Galilean Dialogue* by J. M. Jauch [Jauch, 1990]:

> Salviati: Suppose I give you two sequences of numbers, such as
>
> $$7\,8\,5\,3\,9\,8\,1\,6\,3\,3\,9\,7\,4\,4\,8\,3\,0\,9\,6\,1\,5\,6\,6\,0\,8\,4\ldots$$
>
> and
>
> $$1,\ -1/3,\ +1/5,\ -1/7,\ +1/9,\ -1/11,\ +1/13,\ -1/15,\ \ldots$$
>
> If I asked you, Simplicio, what the next number of the first sequence is, what would you say?

SIMPLICIO: I could not tell you. I think it is a random sequence and that there is no law in it.

SALVIATI: And for the second sequence?

SIMPLICIO: That would be easy. It must be $+1/17$.

SALVIATI: Right. But what would you say if I told you that the first sequence is also constructed by a law and this law is in fact identical with the one you have just discovered for the second sequence?

SIMPLICIO: This does not seem probable to me.

SALVIATI: But it is indeed so, since the first sequence is simply the beginning of the decimal fraction [expansion] of the sum of the second. Its value is $\pi/4$.

SIMPLICIO: You are full of such mathematical tricks ...

The dialogue illustrates an aspect (among many) of the problem of defining randomness for sequences. An apparently random sequence of digits may really be *pseudo-random:* While it may appear to be "statistically random" and unpredictable, there may be a (hidden) rule or arithmetical method for generating the entire infinite sequence purely deterministically.

Our intuition tells us that if there is a deterministic and effective rule for computing every term of an infinite sequence, then, despite appearance, the sequence cannot be genuinely random, as von Neumann famously cautioned. But how do we precisely define the class of infinite sequences that are "genuinely random"?[1]

1.4 A Laplacian problem

Imagine our blackbox to be a computer program simulating the flip of a fair coin, with 1 denoting heads, and 0 denoting tails. Suppose that we run it to generate fifty flips and observe the resulting outcome sequence. Such an outcome sequence is a binary string of length 50, and there are 2^{50} possible outcome sequences. If we observe the outcome sequence to be

$$10010001001100111011001010100000110100001100011100,$$

we are not surprised. We regard it as a "random binary string", and consider the program to be running normally.

But if the outcome sequence is

$$01010101010101010101010101010101010101010101010101,$$

we consider it to be an extraordinary event, and justifiably suspect malfunction, perhaps a bug in the program. But according to simple probability, this binary

[1]Note that it would be an error to define random infinite sequences as those which cannot be generated deterministically by a specified rule or arithmetical method. This is because there are only countably many such methods, while there are uncountably many sequences which are not random in any sense of the word. For example, there are uncountably many infinite binary sequences for which the bits at even positions are all set to 0 but the bits at odd positions are allowed to be arbitrary, and none of these sequences is random.

string of perfectly alternating 0s and 1s is as likely to be produced as the first supposedly "random-looking" string, and so should not be regarded any more special than the first one. Yet, in a clear intuitive sense, the "regularity" of the second string makes it much less random compared to the first string.

P. S. Laplace (1749–1827) was aware of this problem and pointed out the following reason why, intuitively, a regular outcome of a random event is unlikely:

> "We arrange in our thought, all possible events in various classes; and we regard as *extraordinary* those classes which include a very small number. In the game of heads and tails, if heads comes up a hundred times in a row, then this appears to us extraordinary, because the almost infinite number of combinations that can arise in a hundred throws are divided in regular sequences, or those in which we observe a rule that is easy to grasp, and in irregular sequences, that are incomparably more numerous." [de Laplace, 18191952]

In addition to the rarity of regular patterns, Laplace points out that certain strings may have a "cause", making them unlikely to be random:

> "The regular combinations occur more rarely only because they are less numerous. If we seek a cause whenever we perceive symmetry, it is not that we regard the symmetrical event as less possible than the others, but, since this event ought to be the effect of a regular cause or that of chance, the first of these suppositions, is more probable than the second. On a table, we see letters arranged in this order: `C o n s t a n t i n o p l e`, and we judge that this arrangement is not the result of chance, not because it is less possible than others, for if this word were not employed in any language we would not suspect it came from any particular cause, but this word being in use among us, it is incomparably more probable that some person has thus arranged the aforesaid letters than this arrangement is due to chance." [de Laplace, 18191952]

We therefore observe (still using our vague and imprecise language) that among all finite strings of a fixed large size (say length 100), there are only a relatively small number of "non-random" strings — strings which possess "regularity" or more generally have some "cause" behind them. We may call the other strings to be random.

Thus, it appears that for finite strings of a fixed large size (say length 100), there is some attribute that corresponds to the intuitive notion of randomness. But how do we precisely define it?

This intuitive attribute of randomness can be *partially approximated* by formulating certain events described in the ordinary language of classical probability. E.g., by requiring that the standard deviation of the run lengths of digits be within certain limits one can avoid such regular sequences as the one above with alternating 0 and 1. In fact, such events are designed and formulated as statistical tests

for estimating the "randomness confidence" of a finite sequence of digits. Some examples are restrictions on the distribution of run-lengths, autocorrelation, serial correlation, comparison with standard test distributions such as χ^2-tests, etc.[2]

However, all such tests appear to be only partial approximations, and no event formulated in the usual language of classical probability seem to precisely capture this attribute of randomness in an intuitively satisfactory way.

1.5 The main problems. What this article is about

The considerations above lead us to two classic problems of mathematical and statistical philosophy:

PROBLEM 1 Randomness for Infinite Sequences. When is a given infinite sequence of digits random?

PROBLEM 2 Randomness for Finite Strings. When is a given finite string of digits random?

As it turned out, most of the early mathematical results concerning randomness were about infinite sequences.[3] Work on defining randomness for finite sequences started later.[4] As Ulam had noted [MacHale, 1993]: "The infinite we shall do right away. The finite may take a little longer."

It also turned out that the two concepts are closely connected in a remarkable way with "algorithm" or "effective computability" playing a central role in their definitions. Quite satisfactory answers to both problems emerged almost simultaneously in mid 1960s, together with the birth of the new fields of *Algorithmic Randomness* and *Algorithmic Complexity.*

The primary goal of this article is to present these celebrated solutions of the above two classic philosophical problems as a brief introduction to algorithmic randomness and complexity. It is, however, important to note that a vigorous effort to further sharpen, refine, and calibrate the definition of randomness is currently being pursued through a large and growing body of lively research activity in algorithmic randomness [Nies, 2009; Downey and Hirschfeldt, 2010; Li and Vitanyi, 2008; Shen *et al.*, 20??]. Therefore the topic is best viewed as a part of continuing research, and there is still ample scope for debate if the answers obtained in the 1960s are final.

---

[2] See Knuth [1998] for such statistical tests of randomness. Chris Wetzel of Rhodes College has interactive web pages illustrating such tests of randomness (`http://faculty.rhodes.edu/wetzel/random/intro.html`).

[3] Borel's 1909 work on normal numbers and the Weyl equidistribution theorem of 1916 are preliminary forms of randomness for infinite sequences, but it was von Mises who in 1919 directly focused on the concept and gave a fundamental definition. Church's introduction of "algorithm" in 1940 made von Mises' definition mathematically precise and also made the subject permanently "algorithmic". For a truly satisfactory answer, one had to wait another quarter century until Martin-Löf found it in 1965.

[4] Early work was done in the 1960s by Solomonoff [1960; 1964], Kolmogorov [1963; 1965], and then Chaitin (see [Chaitin, 1992] for references).

Throughout this article, we will consider only strings and sequences, and make the further simplifying assumption that the only digits (possible outcomes) are 0 and 1 with equal probability of $\frac{1}{2}$ each — the "fair coin model". In other words, the blackbox always represents the flip a perfectly fair coin. In particular, all strings and sequences will be *binary* strings and sequences. This restriction is not really as stringent as it may appear to be (see subsection 2.5).

1.6 The solutions. Algorithmic randomness

There are three key approaches for defining randomness in sequences: *unpredictability, typicality,* and *incompressibility* [Downey and Hirschfeldt, 2010]. We now briefly and informally introduce these approaches. They will be described in detail using mathematically precise language later in this article.

- *Unpredictability.* This can be understood in terms of impossibility of successful gambling strategies. According to *randomness as unpredictability,* an infinite binary sequence is random if, roughly speaking, it is impossible to effectively specify a gambling strategy which can make long run gains for the gambler when played against this sequence as the outcomes. A *frequentist* version of randomness as unpredictability was used by von Mises; see the comments of Feller quoted in subsection 5.1.

- *Typicality.* A property or attribute of infinite binary sequences is called *special* if the probability that the property holds is zero, and is called *typical* if the probability that the property holds is one. An attribute is special if its complement (negation) is typical, and vice versa. For example, consider the property of having no run of zeros of length seven. It can be shown that the probability is zero that an infinite binary sequence has no run of zeros of length seven, thus this property is special. The intuition here is that if a sequence has a special property or attribute, then it cannot be random, and randomness is equivalent to the complete lack of special attributes. According to *randomness as typicality,* an infinite binary sequence is random if, roughly speaking, it is impossible to effectively specify a special property that the sequence possesses. This is essentially the definition of Martin-Löf-randomness. What Martin-Löf did was to provide a mathematically precise notion of effective specifiability in this context.

- *Incompressibility.* Some finite strings can be effectively specified by descriptions much shorter than the string itself. For example, the string

  $$0101010101010101010101010101010101010101010101010101,$$

  can be specified as "01 repeated fifty times". The short description here exploits the regularity of pattern. The idea is also utilized by data compression programs: Strings which compress well are the ones having regularity of pattern or redundancy of information, and highly irregular or "random" strings

do not compress well. This approach, known as *Kolmogorov Complexity,* was introduced by Solomonoff, Kolmogorov, and Chaitin, and yields a definition of randomness for finite strings: A string is random if its shortest description has length equal to the length of the string itself. This can also be carried over to infinite strings: According to *randomness as incompressibility,* an infinite binary sequence is random if none of its initial segments compress "very much".

These three approaches to defining randomness may appear to be independent of each other, and we might expect them to lead to different definitions of randomness. It is therefore a remarkable fact that (with appropriate choices for the notion of effective specifiability) all three definitions turn out to be equivalent! Schnorr's Theorem, a celebrated result, establishes the equivalence of randomness as typicality (i.e. Martin-Löf randomness) with randomness as incompressibility.

This surprising equivalence between randomness as unpredictability, randomness as typicality, and randomness as incompressibility, is strong evidence that this common equivalent definition is satisfactory, and that algorithmic randomness provides an adequate mathematical foundation for randomness.

1.7 What this article is not about

Let us now clearly state a disclaimer. This article's scope is restricted to discussions of only Problems 1 and 2 above, dealing with randomness of sequences and strings from a purely extensional view. Our coverage of randomness is thus limited to what may perhaps be called "the randomness of bit patterns", while the word randomness as used in ordinary language is overloaded with meanings many of which do not necessarily involve bit patterns (binary strings and sequences).

Randomness may be discussed in a more general setting than just bit patterns. This has been done, e.g., by Eagle, who introduces the idea of randomness as *maximal unpredictability,* and has written a critique of algorithmic randomness. One may also regard *random and probabilistic processes* as the main framework for investigating randomness. See Eagle [2005], where further references can be found.

Also, there are popular books on randomness (e.g. Aczel [2004], Beltrami [1999], and Bennett [1998]) with accessible approaches to various aspects of randomness. The books by Taleb [2005] and by Mlodinow [2008] include various financial and social implications of randomness as well.

Even within the limited case of "bit-pattern randomness", there are other important issues, such as the following broad question of considerable philosophical and practical interest: *How can random sequences be physically generated?* Another example is the relation between quantum mechanics and (algorithmic) randomness. Several interesting discussions and further references are in [Calude, 2005; Yurtsever, 2000; Svozil, 1993; Longo, 2009; Bailly and Longo, 2007; Penrose, 1989; Stewart, 2002].

Because of our extensional approach, we have omitted all such issues and limited ourselves only to mathematical definitions of randomness for sequences, i.e., *precise criteria to distinguish random sequences from non-random ones.*

But even for mathematical definitions based on the extensional view, algorithmic randomness is not the only approach possible. For example, in [van Lambalgen, 1990], Van Lambalgen introduces an axiomatic approach to randomness. We will not discuss such approaches either.

Other than reasons of space limitation and cohesion, there is another point which has compelled us to select only the topic of algorithmic randomness for discussion,

This year marks the 100th anniversary of Borel's strong law (1909), which first showed the importance of the concept of limiting frequency, and is the precursor of many ideas based on it, including von Mises' idea of randomness. We believe that in the one hundred years since Borel's strong law, algorithmic randomness stands out as the crowning achievement in the study of random sequences. In addition to providing deep philosophical insight, it has unraveled fundamental connections of randomness with many diverse areas that has been truly spectacular.[5]

1.8 How this article is organized

Section 2 sets up notation for strings and sequences, the Cantor Space, which forms the basic framework for carrying out further discussions, and contains a brief and informal introduction to Lebesgue measure on the unit interval and the Cantor space, which we treat as equivalent.

Discussion of mathematical randomness begins in Section 3, where we define a series of classical stochastic or frequency properties in an effort to distill out randomness, but conclude by noting how this forces upon us notions like definability and effective computability.

In Section 4, an informal but precise definition of effective computability is given.

Sections 5, 6, 7, 8, and 9 form the core material on algorithmic randomness: Von Mises randomness, Martin-Löf randomness, Kolmogorov complexity and randomness of finite strings, application to Gödel incompleteness, and Schnorr's theorem.

Sections 10 and 11 form an overview of somewhat more advanced and recent topics, including definitions of various other forms of related randomness notions that have been studied, and indicating the current state of affairs.

[5] Algorithmic randomness is intimately related to the creation and development of several other fields such as Algorithmic Probability (originated by R. J. Solomonoff) and Universal Search (originated by L. A. Levin), all of which together comprise a broader (and vast) area now generally known as *Algorithmic Information Theory*. In addition to the creation of such fields, there has been wide and fundamental impact on many areas in mathematics, philosophy, statistics, and computer science: We mention recursion theory and Hausdorff dimensions; inductive and statistical inference and prediction; classical information theory; and complexity theory, machine learning, and artificial intelligence.

Finally, in Section 12 we express our view on the Martin-Löf-Chaitin thesis.

Proofs of mathematical results are provided or outlined whenever they are fairly straightforward. Involved proofs are omitted, almost all of which can be found in one of the books [Feller, 1968; Knuth, 1998; Li and Vitanyi, 2008; Downey and Hirschfeldt, 2010; Nies, 2009; Chaitin, 1992; Calude, 1994; Odifreddi, 1992].

2 STRINGS, SEQUENCES, CANTOR SPACE, AND LEBESGUE MEASURE

We use the term "string" (if not further qualified) to be a synonym for "finite sequence", and the term "sequence" (if not further qualified) to be a synonym for "infinite sequence".

To describe sequences and strings, we will use a finite *alphabet* Σ consisting of a finite number of letters or symbols. E.g., the *decimal* alphabet $\Sigma = \{0, 1, 2, \ldots, 9\}$ consists of ten *digits,* and the binary alphabet $\Sigma = \{0, 1\}$ consists of just two *bits.* A *string* from Σ is simply a finite sequence of members of Σ. The empty string is denoted by $\mathbf{\Lambda}$.

The set of all strings (finite sequences) from Σ is denoted by Σ^*. For a string $\sigma \in \Sigma^*$, $|\sigma|$ denotes the length of σ.

$\mathbf{N}$ denotes the set of all natural numbers. We deliberately leave it ambiguous if zero is considered a natural number or not. It can always be understood from context, and this ambiguity does not cause any problem. The set of all infinite sequences from Σ will be denoted by $\Sigma^{\mathbf{N}}$. For $x \in \Sigma^{\mathbf{N}}$, we will often write $x = \langle x_n \rangle_{n=1}^{\infty} = \langle x(n) \rangle_{n=1}^{\infty}$, where $x_n = x(n) \in \Sigma$ for each $n = 1, 2, \ldots$.

2.1 Binary strings and sequences. The Cantor space

To simplify the discussion, we will almost always limit ourselves to binary sequences and strings, i.e. the case where the alphabet is $\Sigma = \{0, 1\}$. The set $\{0, 1\}^*$ is the set of all *binary strings.* The set $\{0, 1\}^{\mathbf{N}}$ is the set of all infinite binary sequences, and is called the *Cantor space.*

If $\sigma \in \{0, 1\}^*$ is a finite binary string, we will let $N(\sigma)$ to denote the set of all infinite binary sequences beginning with σ. E.g., $N(101)$ consists of all infinite binary sequences $\langle x_n \rangle_{n=1}^{\infty}$ for which $x_1 = x_3 = 1$, $x_2 = 0$, and x_k is arbitrary for all $k \geq 4$.

The subsets of the Cantor space $\{0, 1\}^{\mathbf{N}}$ having the form $N(\sigma)$ (for some binary string σ) will be called the *basic intervals* of the Cantor space. (By defining open sets as finite or infinite unions of basic intervals, the Cantor space becomes a topological space which is compact and metrizable.)

2.2 Lebesgue measure over the unit interval

The problem of Lebesgue measure over the unit interval $[0, 1] = \{x \in \mathbf{R} \colon 0 \leq x \leq 1\}$ is essentially a geometric one: Given a subset $E \subseteq [0, 1]$, we want to assign it a number $\mu(E)$ which represents its "size" or "length". For very simple subsets

such as an interval, its measure is simply its length: If $J \subseteq [0,1]$ is an interval with endpoints $a \leq b$ (i.e., J is one of the intervals (a,b), $[a,b)$, $(a,b]$, or $[a,b]$), the measure of J, denoted by $\mu(J)$, is defined as the length of the interval, $\mu(J) = b-a$.

The next step is to define the length of any *open subset* of $[0,1]$. A subset G of $[0,1]$ is said to be *open* if it is a union of open intervals. A standard fact about this linear continuum is that any open set can be expressed *uniquely* as a disjoint union of (possibly infinitely many) intervals. This allows us to naturally and uniquely define the measure $\mu(G)$ of an open set $G \subseteq [0,1]$ to be the sum (possibly as an infinite series) of all the constituent disjoint intervals.

A key idea here is that a set of "small measure" can be covered by an open set of "small measure": A set E is said to be *measure-zero* if E can be covered by open sets of arbitrarily small measure, i.e., for any $\epsilon > 0$ there is an open set G containing A with $\mu(G) < \epsilon$. Slightly more constructively, E has measure-zero if there is an infinite sequence of open sets $G_1, G_2, G_3, \ldots$ with each G_n covering E and with $\mu(G_n) < 1/n$.

A subset $E \subseteq [0,1]$ is defined to be *(Lebesgue) measurable* if for any $\epsilon > 0$ there is an open set G containing E and an open set H containing the difference $G \setminus E$ with $\mu(H) < \epsilon$. Thus a measurable set is one which can be approximated from outside by open sets arbitrarily closely. If E is measurable, it can be shown that the measure of the open set G above approaches a unique limit as $\epsilon \to 0$, and we denote this limit by $\mu(E)$. This defines the Lebesgue measure for every measurable subset of $[0,1]$. If $E \subseteq F \subseteq [0,1]$ are measurable sets, then we have $0 \leq \mu(E) \leq \mu(F) \leq 1$.

The class of measurable sets form a vast collection. If $E \subseteq [0,1]$ is measurable, so is its complement $[0,1] \setminus E$ with $\mu([0,1] \setminus E) = 1 - \mu(E)$. If $\langle E_n \rangle$ is a sequence of measurable sets, then their union $\cup_n E_n$ and intersection $\cap_n E_n$ are also measurable. If the sequence $\langle E_n \rangle$ consists of disjoint measurable sets, then the measure of their union is the sum of the measures of the individual sets: $\mu(\cup_n E_n) = \sum_n \mu(E_n)$.

2.3 Lebesgue measure on $[0,1]$ *as probability*

Consider the experiment of choosing a member x of the unit interval $[0,1]$ in such a fashion that for any two subintervals $J_1, J_2 \subseteq [0,1]$ of equal length, it is equally likely for x to be in J_1 as to be in J_2. (This is referred to as the uniform distribution over $[0,1]$.) For a subset $E \subseteq [0,1]$, the problem of determining the "probability that x is in E", denoted by $P(E)$, is identical to Lebesgue's problem of finding the geometric measure (or length) of E.

We therefore identify, for this experiment, the notion of "events" with the class of measurable sets, and the probability of an event E with the Lebesgue measure of E: $P(E) = \mu(E)$.

2.4 The Cantor space: Infinite sequence of flips of a fair coin

An infinite sequence of flips of a coin, with 1 representing heads and 0 representing tails, can be naturally represented by an infinite binary sequence, i.e., by a member of the Cantor Space $\{0,1\}^{\mathbf{N}}$, the set of all possible outcomes in an infinite sequence of flips of a coin. Subsets of $\{0,1\}^{\mathbf{N}}$ correspond to events, e.g., the basic interval $N(101)$ represents the event that the first and third flips are heads and the second is tails.

We stipulate that the coin is fair and the flips are independent by requiring that for each $n \in \mathbf{N}$ the 2^n possible outcomes in the first n flips are equally likely, or equivalently that the probability of the event $N(\sigma)$ equals $1/2^{|\sigma|}$. Just as we extended Lebesgue measure on $[0,1]$ from intervals to open sets and then to arbitrary measurable sets, a similar process can be carried out to obtain define a "probability measure" for the "measurable subsets" of $\{0,1\}^{\mathbf{N}}$, a process which we now briefly describe.

The open sets in $\{0,1\}^{\mathbf{N}}$ are defined to be arbitrary unions of basic intervals of the form $N(\sigma)$. A basic interval $N(\sigma)$ is said to be maximally contained in a set A if $N(\sigma) \subseteq A$ but $N(\tau) \not\subseteq A$ for each proper initial segment τ of σ. Every open set G then decomposes *uniquely* into a disjoint union of component basic open intervals maximally contained in G. So we can now naturally and uniquely define the measure of G, denote by $\mu(G)$, to be the sum of the lengths of these components. A key idea is again that of "small sets": If $\epsilon > 0$ and if we can cover a set E by an open set of measure less than ϵ, then we can expect the measure of E to be less than ϵ as well. A set E is said to have *measure-zero* if there is an infinite sequence $G_1, G_2, G_3, \ldots$ of open sets each containing E with $\mu(G_n) < 1/n$ for all n.[6] Finally, as before in the case of the unit interval, define $E \subseteq \{0,1\}^{\mathbf{N}}$ to be *(Lebesgue) measurable* if E can be approximated from outside by open sets arbitrarily closely, i.e., if for any $\epsilon > 0$ there is an open set G containing E and and open set H containing the difference $G \setminus E$ with $\mu(H) < \epsilon$.

It can then be shown that the measurable sets form a comprehensive collection including the open sets (and so all the basic intervals) and each measurable set E gets naturally assigned a unique measure $\mu(E) \in [0,1]$. Also, the complement of any measurable set E is measurable with $\mu(\{0,1\}^{\mathbf{N}} \setminus E) = 1 - \mu(E)$, and for any sequence $\langle E_n \rangle$ of measurable sets, their union $\cup_n E_n$ and intersection $\cap_n E_n$ are also measurable, with $\mu(\cup_n E_n) = \sum_n \mu(E_n)$ whenever the sequence $\langle E_n \rangle$ consists of pairwise disjoint sets. Also, if $E \subseteq F \subseteq \{0,1\}^{\mathbf{N}}$ are measurable, then $0 \leq \mu(E) \leq \mu(F) \leq 1$.

Thus, starting with the simple method of assigning the probability (or measure) $1/2^{|\sigma|}$ to each basic interval $N(\sigma)$, we are then able to naturally extend this assignment procedure to assign probabilities (or measure) to vastly more general types of infinite coin-flip events (measurable subsets of $\{0,1\}^{\mathbf{N}}$). We call this as-

[6] It is easy to see that a singleton is measure-zero. Using convergent infinite geometric series, it now follows that a countable union of measure-zero sets is also measure-zero. Thus all countable sets are measure-zero. A less trivial fact is that there are uncountable measure-zero sets.

signment (μ) to be the *Lebesgue (or uniform) probability measure on the Cantor Space* $\{0,1\}^{\mathbf{N}}$.

In fact, by mapping an infinite binary string $x = \langle x(n)\rangle_{n=1}^{\infty}$ to the real number $\sum_n 1/2^{x(n)} \in [0,1]$, we can naturally identify the uniform probability measure on the Cantor space with the Lebesgue measure on $[0,1]$: The interval $[1/2,1]$, e.g., corresponds to the event "the first flip is a heads". This correspondence, as a mapping, is not quite one-to-one since the dyadic rationals of the form $m/2^n$ (where m,n are positive integers with $0 < m < 2^n$) have two different binary expansions, but these form only a countable set of exceptional points. All other reals in $[0,1]$ have a unique infinite binary expansion. So this mapping between the Cantor space $\{0,1\}^{\mathbf{N}}$ and the unit interval $[0,1]$ is an almost one-to-one correspondence satisfying $\mu(E) = P(E')$ for any measurable subset $E \subseteq [0,1]$ with E' being the set of infinite binary sequences which are binary expansions of the members of E.

To summarize, the Lebesgue measure on $[0,1]$ (i.e., the uniformly distributed probability measure on $[0,1]$) and the uniform probability measure on the Cantor space $\{0,1\}^{\mathbf{N}}$ are essentially the same thing.

The zero-one law. This important result (used later) asserts the following for any measurable $E \subseteq \{0,1\}^{\mathbf{N}}$. Suppose that whenever $x, y \in \{0,1\}^{\mathbf{N}}$ are sequences differing only at a finite number of places (i.e., $(\exists m)(\forall n > m)(x(n) = y(n))$), then $x \in E \iff y \in E$. Then E has measure either zero or one.

2.5 More general probability distributions

We stress the following: *The notion of randomness must be understood relative to an a priori fixed probability distribution for the outcome values.* Each probability distribution for the outcome values determines a specific set of random sequences, and a sequence which is random relative to a given probability distribution for the outcomes will not be random relative to a different probability distribution for those outcomes. Without an underlying fixed a priori probability distribution for the possible outcome values, the notion of a "random sequence" would not even make sense. It is thus useful to fix a specific probability distribution on a specific set of outcome values (a probability model) when discussing random sequences.

As mentioned earlier, in this article we will restrict our attention to the case where the underlying probability distribution for the blackbox is *the fair coin model:* Only two equiprobable outcomes 0 and 1 with $P(0) = P(1) = \frac{1}{2}$. We will then try to find out which sequences are random (i.e. investigate Question B of Section 1.2 and Problems 1 and 2 of Section 1.5) relative only to this fair coin model. This may at first appear to be too severe a restriction, but the extra generality obtained by considering more general probability distributions, while adding verbiage, would not provide much extra insight into the question of randomness. For the reader who is still worried about our restriction to the fair coin model alone, we now mention some technical results which show how this simple case can represent, at least for sequences, other more general distributions.

Suppose that our blackbox represents a more general probability distribution,

with n outcomes $\omega_1, \omega_2, \ldots, \omega_n$, and corresponding probabilities $P(\omega_i) = p_i$, $i = 1, 2, \ldots, n$, where $0 < p_i < 1$ and $\sum_{i=1}^{n} p_i = 1$. Then the probability space $P^{\mathbf{N}}$ of all infinite sequences of outcomes is still essentially identical to the Cantor space with Lebesgue measure. In particular, there is a measurable bijection between $P^{\mathbf{N}}$ and $\{0,1\}^{\mathbf{N}}$ which preserves measure. This is a technical result on finite Borel measures on complete separable metric spaces. If $p_1, p_2, \ldots, p_n$ are computable real numbers, then this bijection can be assumed to be an effective one.

If the probabilities $p_1, p_2, \ldots, p_n$ are dyadic rational numbers (as is the case, e.g., in computer representations of numbers), then one can use an especially simple effective coding by which any sequence (finite or infinite) of outcomes of the blackbox can be represented effectively by a corresponding binary sequence from the fair coin model, while still preserving probabilities of all events.

As another example, we mention the so called *von Neumann trick,* a simple method by which one can "turn a biased coin into an unbiased one", or, more precisely, simulate sequences of flips of a perfectly unbiased coin using sequences of flips of a biased coin (whose probability of heads differs from probability of tails). For every two consecutive flips, record the outcome as a single 0 if the (ordered) pair of flips is HT, record it as a single 1 if the pair of flips is TH, otherwise discard the pair (the cases HH and TT), and move to the next pair of flips of the coin.

We will see later that Martin-Löf's definition of randomness is general and flexible enough to be applied directly to very general probability distributions. However, the technical results mentioned here show that there are mathematically sound ways to restrict our attention to the simplified case of the fair coin model, without imposing any serious limitation to the study of random sequences. We adopt this simplification, which equivalently means that the underlying sequence space will always be the Cantor Space with the (uniform) Lebesgue measure.

3 CLASSICAL STOCHASTIC RANDOMNESS IN INFINITE SEQUENCES

We now begin the discussion of the fundamental question stated in Problem 1: *Given an infinite binary sequence $x = \langle x_1, x_2, \ldots, x_n, \ldots \rangle \in \{0,1\}^{\mathbf{N}}$, how do we determine if it is random?*

3.1 Key points

We start by noting several key points of randomness in infinite binary sequences. Recall that the Cantor space with Lebesgue measure is the underlying probabilistic model throughout. Our observations here will be heuristic and informal but of fundamental importance.

(a) *The probability that a sequence is random equals one,* i.e., the random sequences form a set of measure one (full-measure set). A basic intuition here is that the omission or addition of one single bit has no effect on the randomness of an infinite sequence: If x is the sequence $x_1x_2x_3\ldots$, and x' is the

sequence $x_2x_3x_4\ldots$ obtained by dropping only the first bit of x, then x' is random if and only if x is random. It follows by mathematical induction that the randomness of an *infinite* sequence should depend only on its "eventual behavior" and no amount of finite part can determine the randomness of the entire sequence.[7] Thus if two infinite sequences x and y agree on all but a finite number of places (i.e. $\exists m\,\forall n{>}m\,(x_n = y_n)$), then the randomness of x is equivalent to the randomness of y. This implies that the set of random sequences satisfies the condition for the zero-one law, and therefore must be either a measure-zero set or a full-measure set. As pointed out by Laplace, it seems natural that the random sequences should form the vast majority of sequences: If we "randomly pick" a sequence in $\{0,1\}^{\mathbf{N}}$, or equivalently, generate one by "randomly flipping" a fair coin infinitely many times, the probability that the result is random should be high, and so non-zero. It follows that the random sequences must form a set of measure one.

(b) Second, *if the sequence of outcomes of a gambling wheel with two equiprobable outcomes is random then no successful betting strategy can be devised against it.* Note that, here the bits of an infinite sequence x are thought to be generated by the independent turns of the gambling wheel, and that the randomness of x would imply a strong form of unpredictability for the future bit values of x (from previously observed bit values). More precisely, suppose that a gambling house is generating the infinite binary sequence x and offering the following fair game: A gambler can bet an amount b of money predicting the next bit of x; if the prediction is correct, the gambler wins an amount b, otherwise loses the same amount b. (This rule is tantamount to our underlying assumption of a fair coin model.) We say that the gambler is able to *devise a successful gambling system against x,* or *beat the house against x*, if by using a suitable strategy the gambler can start with a finite initial capital and win an arbitrarily large fortune without going bankrupt. By a strategy we mean a finitely specifiable rule which determines how much (and whether) to bet on particular turn based on the outcomes of the previous turns. (These notions will be made more precise later in subsections 5.1 and 11.2.) From the point of view of the house, the randomness of the sequence x of outcomes must imply that the bits of x should be so unpredictable that no gambler would be able to beat the house against x. We carefully note that this *impossibility of a successful gambling system is a fundamental necessary condition for randomness of an infinite sequence,* (a fact first recognized by von Mises, see subsection 5.1 for quoted comments of Feller).[8]

[7] This is quite similar to the notion of *the limit* of an infinite numerical sequence found in elementary calculus: No amount of alternation of the values of any finite number of terms of a sequence can affect its limit.

[8] Later, during the attempts in subsections 5.1 and 11.2 to find a suitable definition of randomness, things will be turned around to postulate this impossibility condition as also a sufficient condition for randomness.

(c) On the other hand, *randomness cannot be identified with complete and absolute lawlessness.* We may try to think of a sequence to be random if it satisfies no law whatsoever ("absolutely lawless"). But, as pointed out by Calude in [Calude, 2000] (see also [Volchan, 2002]), no such sequence can exist, since *every* digit-sequence satisfies the following Ramsey-type combinatorial law first proved by van der Waerden [1927]: The positions (indices) for at least one digit-value will contain arbitrarily long arithmetic progressions. Thus we have to abandon such ideas of "complete lawlessness".

(d) In fact, *random sequences will necessarily satisfy certain limiting or stochastic properties.* For example, given a sequence $x = \langle x_k \rangle$, let $\mathbf{S}_n[x]$ denote the number of 1s (or **S**uccesses) in the first n terms of the sequence x:

$$\mathbf{S}_n[x] \;=\; \sum_{k=1}^{n} x_k \;=\; \text{Number of 1s in first } n \text{ terms of } x,$$

so that the quantity $\frac{1}{n}\mathbf{S}_n[x]$ represents the *proportion* of 1s in the first n terms of x. This proportion is called the *relative frequency* or simply the *frequency* (of successes). If for a sequence x this proportion exceeds $\frac{2}{3}$ (say) infinitely often (i.e., $\frac{1}{n}\mathbf{S}_n[x] > \frac{2}{3}$ for infinitely n), then x cannot be random (under the fair coin model), because a gambler would then be able to exploit this "bias within x" to devise a strategy which (starting with a finite initial capital) can return an arbitrarily large fortune without going bankrupt.

More precisely, if x is to be random so that no gambling system would be successful against it, then it can be shown that the following condition must hold:

> *For no positive number p should the proportion $\frac{1}{n}\mathbf{S}_n[x]$ exceed $\frac{1}{2}+p$ (or fall below $\frac{1}{2}-p$) infinitely often.*[9]

This condition on x is equivalent to requiring that the relative frequency $\frac{1}{n}\mathbf{S}_n[x]$ approach the value $\frac{1}{2}$ in the limit as n approaches infinity.

Thus, for every random x we have $\lim_{n\to\infty} \frac{1}{n}\mathbf{S}_n[x] = \frac{1}{2}$. This exemplifies that *in order to be random, a sequence, instead of being "totally lawless", must actually satisfy certain stochastic laws of "unbiasedness".* In the rest of this section we will discuss further stochastic laws that should be satisfied by every random sequence, starting with the *Borel strong law,* whose basic condition is the same as the one in the above example.

[9] If $\frac{1}{n}\mathbf{S}_n[x]$ exceeds $\frac{1}{2}+p$ infinitely often for a positive p, then the gambler can beat the house by betting a constant fraction r of his remaining capital at each turn predicting a bit outcome of 1, where r is a constant fraction with $0 < r < p/(1+p)$. We omit the details of calculation.

3.2 The Borel Strong Law

In 1909, Émile Borel [1909] proved a remarkable fact about infinite binary sequences which was later generalized by many mathematicians (from Cantelli to Kolmogorov) and became a fundamental result of probability theory called *The Strong Law of Large Numbers.* Borel established that with probability one, the proportion of 1s among the first n terms ($\frac{1}{n}\mathbf{S}_n$) approaches the value $\frac{1}{2}$ in the limit.

THEOREM 1 The Borel strong law. *For independent infinite sequences of flips of a fair coin, let B denote the event that the proportion of successes among the first n flips, $\frac{1}{n}\mathbf{S}_n$, approaches the limit $\frac{1}{2}$ as $n \to \infty$, or formally, put*

$$B \;=\; \left\{ x \in \{0,1\}^{\mathbf{N}} \;:\; \lim_{n\to\infty} \frac{\mathbf{S}_n[x]}{n} = \frac{1}{2} \right\}.$$

Then the probability of the event B is 1, i.e. the set B has Lebesgue measure 1.

We saw that due to impossibility of successful betting strategies, a random sequence must satisfy the condition of the Borel strong law. That, together with our earlier observation that the random sequences form a set of measure one, provides an informal proof for the Borel strong law. (For a formal proof, see [Feller, 1968].)

Informally, the Borel strong law asserts that if we randomly pick a member x from $\{0,1\}^{\mathbf{N}}$, the probability is one that the relative frequency of 1s among initial parts of x approach a limiting value called *the limiting frequency,* and this limiting frequency is equal to $\frac{1}{2}$, so that "randomly picked sequences are unbiased".

Conversely, if for a sequence x this limiting frequency exists but is not equal to $1/2$, then, in view of our underlying fair coin model, x would clearly be biased, not random. And if the limiting frequency does not exist then either $\limsup_n \frac{1}{n}\mathbf{S}_n[x] > \frac{1}{2}$ or $\liminf_n \frac{1}{n}\mathbf{S}_n[x] < \frac{1}{2}$, and in either case arbitrarily large segments of x would be biased with arbitrarily great statistical significance, so x again would be non-random (successful gambling systems could be devised against x in these cases).

Thus, it is natural to view this "stochastic law of unbiasedness" as a "stochastic law of randomness".

Note that satisfying this law is only a basic necessary condition for being random. For example the sequence $010101010101\cdots$ of alternating 0s and 1s satisfies the condition of the strong law but this sequence is clearly not random. We therefore look for stronger stochastic laws which would exclude such simple examples.

3.3 Borel Normality

Let σ be a fixed binary string with length $|\sigma| = k$ (e.g., if $\sigma = 00110101$ then its length is $|\sigma| = k = 8$). If a fair coin is flipped k times, the probability of obtaining σ as the resulting outcome equals $1/2^k$. For an infinite binary sequence x, consider a block of k bits starting at bit position n: This is the block segment of x given by $x_n x_{n+1} \ldots x_{n+k-1}$. It is not hard to see that with probability 1 the string σ

must occur infinitely many times in an infinite binary sequence.[10] Adding this condition (that every finite binary string must occur in x with infinite frequency) as a stochastic law certainly excludes simple sequences satisfying the Borel strong law, such as $01010101\cdots$, as being random.

In [Borel, 1909], Borel had established an even stronger fact. Call an infinite binary sequence x to be *Borel normal in base* 2 if for each fixed binary string σ of length $|\sigma| = k$, the frequency of occurrences of σ among the first n bits of x (as a fraction of n) approaches $1/2^k$ as $n \to \infty$. (For $|\sigma| = k = 1$, this reduces the condition of the Borel strong law.) Borel proved that *the probability that x is normal in base* 2 *is one, i.e. almost all infinite binary sequences are normal in base* 2. Any infinite binary sequence satisfying this property will clearly be "much more random" than simple sequences like $01010101\cdots$.

An immediate consequence of the Borel normality is another form of unpredictability for random sequences: *The observation of any particular bit pattern in a random sequence does not influence the value of the next bit.* More precisely, for almost all infinite binary sequences x and any fixed finite binary string σ (bit-pattern), the limiting frequency that σ is immediately followed by 0 equals the limiting frequency that σ is followed by 1.

An explicit example of an infinite binary sequence which is normal in base 2 is the *Champernowne binary sequence* obtained by concatenating the binary representation of every non-negative integer (taken in their natural order):

$$0110111001011101111000100110101011110011011110111 1\cdots$$

As far as randomness is concerned, this is a big improvement over the simple $01010101\cdots$, but it is impossible to call this sequence random, since it is also generated by a simple effective procedure.

Recall that one can identify $\{0,1\}^{\mathbf{N}}$ with the unit interval $[0,1]$ by viewing infinite binary sequences as binary expansions of reals in $[0,1]$ (after disregarding a negligible countable subset of $\{0,1\}^{\mathbf{N}}$). This identification preserves measurable sets and the measure of every such set. Moreover, given an integer base $b > 1$, every real $x \in [0,1]$ can be expanded in base b as:

$$x = \sum_{k=1}^{\infty} x_k b^k, \quad \langle x_k \rangle \in \{0,1,\ldots,b-1\}^{\mathbf{N}}.$$

The terms of the infinite sequence $\langle x_k \rangle \in \{0,1,\ldots,b-1\}^{\mathbf{N}}$ above are known as the *b-ary digits of x*. E. g., $b = 2$ for binary and $b = 10$ for decimal expansion.

For $x \in [0,1]$ with base b expansion $\langle x_k \rangle$, we say that x *is Borel normal in base b* if for each fixed finite string $\sigma \in \{0,1,\ldots,b-1\}^*$ of length $|\sigma| = k$, the

[10] Proof: Divide x into consecutive blocks of size k each. For any n, the probability that σ is not the n-th block equals the constant $r = 1 - 1/2^k$. So the probability that σ does not occur as one in a run of m consecutive blocks is r^m. Since $r < 1$, $r^m \to 0$ as $m \to \infty$, and thus for any n the probability that σ does not occur after the n-th block is zero. It follows that the probability that σ occurs in only finitely many blocks is also zero, QED.

frequency of occurrences of σ among the first n b-ary digits of x (as a fraction of n) approaches $1/b^k$ as $n \to \infty$. Finally, x *is absolutely normal* if x is Borel normal in every base $b > 1$. Note that by the identification of $\{0, 1, \ldots, b-1\}^{\mathbf{N}}$ with $[0, 1]$, these definitions apply to infinite sequences as well.

Two examples numbers Borel normal in base 10 are the following reals shown in decimal expansion:

$$0.12345678910111213\cdots \text{ (Decimal Champernowne number)},$$
$$0.235711131719232931\cdots \text{ (Decimal Copeland-Erdös number)}.$$

For the first number above, the decimal digits after the decimal point are formed by concatenating the positive integers written in decimal notation (in their natural order), while the second one has decimal digits formed by concatenating the positive primes written in decimal notation. While both these are Borel normal in base 10, it is not clear if they are Borel normal in any other base.

Borel's result implies that almost all reals in $[0, 1]$ (almost all infinite binary sequences) are actually absolutely normal. But it is harder to come up with examples of absolutely normal numbers, as they have an even higher degree of randomness compared to the Champernowne number or the Copeland-Erdös number. It is an old open question if the number π (or $\sqrt{2}$, or e) is absolutely normal, or even normal in some base whatsoever.

Following early work of Sierpinski [1917] (also Turing [1992]), Becher and Figueira [2002] have constructed absolutely normal numbers using an effective procedure (these are real numbers which are computable, although somewhat complicated to define).

3.4 Laws of random walk ("wandering drunkard" laws)

One can identify the infinite binary sequences (or, equivalently, all possible outcomes of an infinite sequence of coin flips) with the set of all random walks. Consider the number line infinitely extended in both directions and indexed by the integers $\mathbf{Z}$, and a person starting at 0 taking a sequence of steps determined by $x \in \{0, 1\}^{\mathbf{N}}$ as follows: The first step is one unit to the right (in the positive direction) if $x_1 = 1$ and one unit to the left if $x_1 = 0$; and the n-th step of the person is similarly determined by the value of x_n. If x is random, then this results in a *random walk* (sometimes called the drunkard's walk). If $\mathbf{S}_n[x]$ denotes the number of 1s (Successes) in the first n bits of x and $\mathbf{F}_n[x] = n - \mathbf{S}_n[x]$ the number of 0s (Failures) in the first n bits of x, then the position of the person on the number line after step n is given by:

$$\mathbf{S}_n[x] - \mathbf{F}_n[x] = 2\mathbf{S}_n[x] - n.$$

It is not hard to see that with probability 1, the person must move away an arbitrarily large distance to the right of the starting point, and also an arbitrarily

large distance to the left of the starting point.[11] Formally, in terms of infinite binary sequences, the set

$$\left\{x \in \{0,1\}^{\mathbf{N}} : \sup_n \mathbf{S}_n[x] - \mathbf{F}_n[x] = +\infty \text{ and } \inf_n \mathbf{S}_n[x] - \mathbf{F}_n[x] = -\infty\right\}$$

must have measure (probability) one.

In particular, this stochastic law implies that for a walk to be random, the person must oscillate about the origin with "arbitrarily large amplitudes", and must "cross the origin" from right to left and from left to right infinitely many times:

THEOREM 2 Law of Symmetric Oscillation in Random Walks. *If* $x \in \{0,1\}^{\mathbf{N}}$ *is random, then we must have:*

$$\frac{\mathbf{S}_n[x]}{n} > \frac{1}{2} \text{ for infinitely many } n, \text{ as well as: } \frac{\mathbf{S}_n[x]}{n} < \frac{1}{2} \text{ for infinitely many } n.$$

Thus a walk in which the person eventually stays on one side of the origin (eventually to the right or eventually to the left) is "biased" and cannot be "random".

3.5 The Law of Iterated Logarithms and Strong Normality

The Borel strong law says that $\frac{1}{n}\mathbf{S}_n$ converges to the expected value $\frac{1}{2}$ with probability 1, but it does not say how the variance (or standard deviation) of $\mathbf{S}_n$ behaves asymptotically.

A much more precise and stronger theorem (than the Borel strong law) is the *Law of Iterated Logarithms,* which gives an exact asymptotic bound on the deviation of $\mathbf{S}_n$: *With probability one, the values of* $\mathbf{S}_n$ *spread around its mean to an asymptotic distance of* $\sqrt{2\log\log n}$ *times its standard deviation.* To state it more precisely, note that the mean of $\mathbf{S}_n$ is $\mu_n = n/2$, and its standard deviation $\sigma_n = \sqrt{n}/2$. Then the Law of Iterated Logarithms asserts the following: For any $\lambda > 1$, the probability is one that $\mathbf{S}_n < \mu_n + \lambda\sqrt{2\log\log n}\,\sigma_n$ for all but finitely many n, and for $\lambda < 1$, the probability is one that $\mathbf{S}_n > \mu_n + \lambda\sqrt{2\log\log n}\,\sigma_n$ for infinitely many n; and similarly for the lower bounds of $\mathbf{S}_n$. (See Feller [1968] for proof).

The Law of Iterated Logarithms has both the Borel strong law and the Law of Symmetric Oscillation in 3.4 as immediate corollaries.

[11] Proof (outline): Computing binomial probabilities, the probability that $|\mathbf{S}_n[x] - \mathbf{F}_n[x]| \leq k$ for all but finitely many n equals zero for each k, and hence the probability that the sequence $\langle \mathbf{S}_n[x] - \mathbf{F}_n[x] : n \in \mathbf{N}\rangle$ is unbounded equals one. Now the events $\sup_n \mathbf{S}_n[x] - \mathbf{F}_n[x] = +\infty$ and $\inf_n \mathbf{S}_n[x] - \mathbf{F}_n[x] = -\infty$ are equiprobable and satisfy the conditions of the zero-one law. So either both events have probability zero, or both have probability one. But it is impossible for both these events to have probability zero, as that would imply boundedness of the sequence $\langle \mathbf{S}_n[x] - \mathbf{F}_n[x] : n \in \mathbf{N}\rangle$ with probability one (a contradiction), and the result follows.

Belshaw and Borwein [2005;] have used a slightly weaker version of the Law of Iterated Logarithms to define the notion of *Strong Normality*. They show that requiring a sequence to be strongly normal makes it more random than requiring it to be simply Borel normal. This is established both by graphic empirical evidence and by proving that the Champernowne binary sequence is not strongly normal. Since every sequence satisfying the Law of Iterated Logarithms is strongly normal, it follows that the Champernowne binary sequence violates the Law of Iterated Logarithms.

3.6 Equidistribution laws and the Ergodic Frequency Theorem

For any statement P, we use the notation $[\![P]\!]$ to denote the binary truth value of the statement P, i.e.,

$$[\![P]\!] = \begin{cases} 1 & \text{if } P \text{ is true} \\ 0 & \text{if } P \text{ is false} \end{cases}.$$

Let $\langle x_n \rangle$ be an infinite sequence of real numbers in $[0,1)$. We say that $\langle x_n \rangle$ is *equidistributed* (or *uniformly distributed*) if for all $0 \leq a < b \leq 1$,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} [\![x_k \in [a,b)]\!] \;=\; \text{measure of } [a,b) \;\; (= b-a).$$

In other words, $\langle x_n \rangle$ is equidistributed if the limiting frequency with which x_n enters the interval $[a,b)$ equals the size of $[a,b)$.

Similarly, a sequence $\langle x_n \rangle$ of infinite binary sequences (i.e. each $x_n \in \{0,1\}^{\mathbf{N}}$) is equidistributed if for each finite binary string σ,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} [\![x_k \in N(\sigma)]\!] \;=\; \text{measure of } N(\sigma) \;\; \left(= \frac{1}{2^{|\sigma|}}\right),$$

where $N(\sigma)$ is the set of all infinite binary sequences having σ as an initial segment.

Equidistribution can also viewed as a form of unbiasedness: Every subinterval asymptotically gets its "proper share" of the sequence.

A great deal of classical mathematical literature exists on equidistribution (see [Kuipers and Niederreiter, 1974]). We mention a theorem of Weyl: *If $\langle a_n \rangle$ is a sequence of distinct integers, then the sequence $\langle \mathrm{FRAC}(a_n x) \rangle$ is equidistributed for almost all x in the unit interval.* (Here $\mathrm{FRAC}(x)$ denotes the fractional part of $x \in \mathbf{R}$: $\mathrm{FRAC}(x) = x - \lfloor x \rfloor$, where $\lfloor x \rfloor$ is the floor of x, or the greatest integer not greater than x.)

Borel normality can be viewed as a special case of the Weyl equidistribution theorem just mentioned. Taking $a_n = 2^{n-1}$, we see that $\langle \mathrm{FRAC}(2^{n-1}x) \rangle$ is equidistributed over the unit interval for almost all x. Moving to $\{0,1\}^{\mathbf{N}}$, this implies that for almost all $x = \langle x_n \rangle \in \{0,1\}^{\mathbf{N}}$, the sequences $\langle x_1, x_2, x_3, \ldots \rangle$, $\langle x_2, x_3, x_4, \ldots \rangle$, $\langle x_3, x_4, x_5, \ldots \rangle$, etc, are equidistributed over $\{0,1\}^{\mathbf{N}}$.

Let $L\colon \{0,1\}^{\mathbf{N}} \to \{0,1\}^{\mathbf{N}}$ denote the left-shift operator:

$$L(\langle x_1, x_2, x_3, \dots \rangle) = \langle x_2, x_3, x_4, \dots \rangle.$$

Then we may restate this equidistribution by saying that for almost all $x \in \{0,1\}^{\mathbf{N}}$, the sequence $\langle x, Lx, L^2x, \dots \rangle$ is equidistributed. In other words, for every finite binary string σ,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} [\![L^k x \in N(\sigma)]\!] \;=\; \text{measure of } N(\sigma) \quad \left(= \frac{1}{2^{|\sigma|}} \right),$$

for almost all $x \in \{0,1\}^{\mathbf{N}}$. But this is precisely the statement that almost all infinite binary strings are normal in base 2.

The left-shift operator L is an example of an ergodic operator, and the last displayed equation is a special case of the Birkhoff Ergodic Theorem.

In subsection 11.6, we further discuss how this approach can be used as as a stochastic law for specifying randomness.

3.7 General probabilistic laws for specifying randomness

In general, suppose that a specific law L for sequences (i.e. a specific property L of sequences) is satisfied with probability one, i.e. satisfied by almost all sequences. We have seen many examples such laws ("laws of large numbers"): The Borel strong law, Borel normality, Symmetric Oscillations, Iterated Logarithms, etc, form an increasingly stringent sequence of such laws. Let us call any such law to be a *probabilistic law of randomness.* Formally, a *probabilistic law of randomness,* or simply a *randomness law,* is an *explicitly defined* set L of binary sequences having full-measure. Given such a law L, the probability that a random sequence will satisfy it is 1, so we can make L to be another law for randomness.

By specifying more and more stringent randomness laws, we can try to verify that a given sequence is sufficiently random in this sense. In fact, any algorithm for generating pseudo-random sequences needs to be subjected to a series of such theoretical stochastic tests[12] to estimate the "quality of randomness" for sequences generated by the algorithm.

After a sufficient number of such stages of refinement, we can hope to arrive at the "right" definition of randomness for infinite sequences — a definition which best matches our intuition. This is the approach taken by Knuth in his fascinating article *What Is a Random Sequence?* (Section 3.5 in [Knuth, 1998]). Knuth gives a series of definitions, **R1**–**R6**, and claims (or hopes, as he clarifies in a later footnote) that the final refinement (definition **R6**) is an appropriate definition of randomness.

[12]This should not be confused with statistical tests for randomness in a finite sequence of digits. As mentioned earlier, for an infinite sequence, no amount of finite part can determine if the entire sequence is random. Statistical tests, in particular the χ^2-tests, are important for estimating "randomness confidence" for a finite set of data. See Knuth [Knuth, 1998] for a list of statistical tests of randomness, and p. 80 of [Knuth, 1998] for theoretical tests. John Walker distributes a suite of statistical tests for randomness.

3.8 Is absolute probabilistic randomness possible?

What happens if we define a sequence to be random if it satisfies *all* explicitly defined probabilistic randomness laws? Provisionally, let us call such a sequence *absolutely random.*

Unfortunately, such a definition causes problems.

First, we have an apparent paradox: For an arbitrary binary sequence y, consider the randomness law $Ly := \{x \in \{0,1\}^{\mathbf{N}} : x \neq y\}$. If x satisfies all randomness laws, then $x \in L_y$ for all $y \in \{0,1\}^{\mathbf{N}}$, but this implies $x \in L_x$, so $x \neq x$, a contradiction! It follows that no binary sequence can satisfy all randomness laws!

However, this is not a real paradox. Since every law can be written using a finite sequence of symbols from a finite alphabet, there are at most countably many laws that can be *explicitly defined.* For all but a countably many members $y \in \{0,1\}^{\mathbf{N}}$, the law L_y cannot even be explicitly stated, and hence will not count.

But, second, we run into a more serious technical metamathematical problem: The notion of "all randomness laws", which makes good intuitive sense, is not formalizable in the standard system of axiomatic mathematics (ZFC). The reason for this is that the notion of a randomness law really refers to a *definable* subset of $\{0,1\}^{\mathbf{N}}$ of full-measure, but the notion of *definability* itself can only be defined in terms of *satisfaction*, or *truth.* By a classic result of Tarski, it is impossible to formalize the notion of truth in a formal system within the system.

One way out from this metamathematical difficulty is to not talk about the entire class of all definable subsets of $\{0,1\}^{\mathbf{N}}$, but restrict ourselves to those which can be defined using formulas with a limited number of quantifiers. By restricting the number and scope of the quantifiers in the defining formulas, various hierarchies of definable subsets of $\{0,1\}^{\mathbf{N}}$ are obtained, such as the *arithmetical hierarchy* (scope of quantifiers limited to the set of natural numbers), and the *analytical hierarchy* (scope of quantifiers limited to natural and real numbers).

This means that we must be satisfied with relative degrees of randomness, and Absolute Randomness, like truth in formal systems, must remain elusive forever.

This is the approach taken by modern research. Very roughly, the n-th level of the arithmetical hierarchy consists of sets defined by formulas with n alternating quantifiers ranging over natural numbers. For each n, a notion of n-randomness can be defined appropriately. The higher the value of n, the stronger is the randomness. The lowest level of this hierarchy, which defines 1-randomness by suitably considering algorithmic randomness laws, is especially important. It captures the random sequences by defining them as the ones which satisfy all "algorithmic stochastic laws converging algorithmically." But we first need to rigorously define the concept of algorithm to talk about these notions precisely.

4 ALGORITHMS AND POST MACHINES

Like randomness, the intuitive notion of algorithm (or effective computation) was not easy to capture. During the early part of the 20th century, in response to

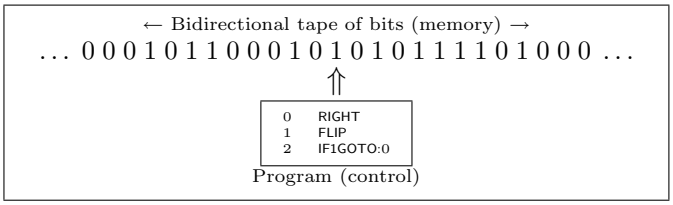
Hilbert's program, there was a great deal of effort by mathematicians to come up with a precise definition of the term "algorithm". During the 1930's decade, which began with Gödel's celebrated negative answer to Hilbert's program, several mathematicians independently came up with such definitions. These included, in addition to Gödel, Herbrand, Church, Kleene, Post, and of course, Turing.

After sorting out more restrictive definitions (such as those now known as primitive recursive computation), mathematicians converged on a definition that Church (in 1936) announced as the appropriate definition for algorithm. Church's assertion is known as *Church's Thesis* or the *Church-Turing thesis.* This is the definition that is accepted today, and so the Church-Turing thesis turned out to be highly successful. Most remarkably, almost all independent definitions coincided and produced the same characterization of the notion of algorithm!

We present this definition as a variant of Post's original version (see also [Uspensky, 1983; Davis, 1980]).

4.1 Post machines and programs

A *Post Machine* consists of a control part containing a *program* (a finite list of *instructions* described below), a bidirectional infinite *tape* of memory bits, and a *head* (shown as ⇑) which at any time is located at some unique bit position (the "current bit") of the tape.



The Post Machine

The head can read the bit at the current position as either 0 or 1 and inform this value to the control, and at the direction of control it can either change the bit value at the current position, or move to one bit position to the left, or to one position to the right. The value of every tape bit is always either 0 or 1 (there are no blank symbols).

Instruction	Code	Function
FLIP	0	Complement current bit: 0 becomes 1 and 1 becomes 0
LEFT	1	Move the head one bit position to the left
RIGHT	2	Move the head one bit position to the right
IF1GOTO:n	$n+3$	Conditional jump: If current bit = 1, GO TO instruction # n

Table 1. Post Machine Instruction set and their numeric encodings

The Post Machine has four[13] types of instructions, as listed in the table.

A Post Machine *program* is simply a finite list of instructions as given in Table 1. Equivalently, since every instruction corresponds uniquely to a natural number code in Table 1, a program can be defined as a finite sequence of natural numbers. In addition, every instruction in the program is assumed to be labeled serially by its "line number", starting with 0 for the first instruction. Program execution starts at line number 0 and sequentially proceeds to the following line, except possibly for the IF1GOTO:n instruction. There is no explicit STOP or HALT instruction, and the program terminates whenever it is unable to execute the "next instruction".

Here is an example program and its code as a finite sequence of natural numbers:

0	RIGHT
1	IF1GOTO:0
2	FLIP
3	LEFT
4	IF1GOTO:3

$= \quad \langle 2, 3, 0, 1, 6 \rangle.$

We encode natural numbers as strings by the correspondence $n \leftrightarrow 1^{n+1}0$:

$$0 \leftrightarrow 10,\ 1 \leftrightarrow 110,\ 2 \leftrightarrow 1110,\ 3 \leftrightarrow 11110,\ \ldots$$

Given a program P and a positive integer k, we will now define a partial function φ of k natural number arguments. To find $\varphi(n_1, \ldots, n_k)$, we start the program with the input string $1^{n_1+1}0 \ldots 1^{n_k+1}0$ on the tape and the head, as in:

$$\ldots 0\,\underset{\boxed{P}}{\overset{\Uparrow}{1}}\overbrace{1 \ldots 1}^{n_1} 0\, 1 \overbrace{1 \ldots 1}^{n_2} 0 \ldots \quad \ldots\ 1 \overbrace{1 \ldots 1}^{n_k} 00 \ldots \quad \text{(other tape bits are 0).}$$

There are now two possibilities, and we define $\varphi(n_1, \ldots, n_k)$ accordingly:

Case 1 *The program P terminates when started as above.* We then define $\varphi(n_1, \ldots, n_k) = m$, where m is the length of the run of 1s to the right of the head, before the first 0 to the right of the head, as in:

$$\ldots * * \underset{\boxed{P}}{\overset{\Uparrow}{\overbrace{1\ 1 \ldots 1}^{m}}}\ 0\ * \ldots$$

(We express the situation in this case by saying that *the program P halts on the inputs $n_1, \ldots, n_k$ with output m.*)

Case 2 *The program P started as above does not terminate ("loops forever").* In this case we leave $\varphi(n_1, \ldots, n_k)$ undefined.

[13] One can combine the first two instructions into a single one to have an adequate version of the Post Machine with only three types of instructions.

Thus, for every program P and positive integer k, there is a unique k-ary partial function φ determined as above, and we express this by saying *φ is the k-ary partial function computed by P,* or *P computes the k-ary partial function φ.*

For example, the example program above computes the 1-ary function $f(n) = n + 2$, but it also computes the 2-ary function $f(m, n) = m + n + 3$. As simpler examples, note that the single line program with the only instruction LEFT computes the (1-ary) successor function $s(n) = n + 1$, and the empty program computes the identity function $f(n) = n$.

We can now formally define the notion of "algorithm" or *effective computability*:

DEFINITION 3. A k-ary partial function $f = f(n_1, n_2, \ldots, n_k)$ is *effectively partial computable* or simply *partial computable* if there is Post machine program P such that P computes f. If f is total, we say that f is *effectively computable,* or simply *computable.*

A subset $R \subseteq \mathbf{N}^k$ (i.e., a k-ary relation R, which is simply a set of natural numbers if $k = 1$) is called *effectively computable* or *algorithmically decidable* or simply *computable* if its characteristic function is a computable function.

$R \subseteq \mathbf{N}^k$ is called *computably enumerable* (or *c.e.*) if it equals the domain of some k-ary partial computable function.

(In older literature the term "recursive" is used in place of "computable".)

It can be shown that a set A is c.e. iff it equals the range of a partial computable function f. If the set A is non-empty, the function f can be assumed to be total. This explains the terminology: A set E is computably enumerable if its elements can be enumerated by a computable function f, as in $E = \{f(1), f(2), f(3), \ldots\}$.

Another important fact is that a set is computable iff both the set and its complement are c.e.

Most importantly, we want to note that the notions of effectively computable functions and computable sets, c.e. sets, etc, are all independent of the particular model of computation. In particular, if a function can be computed by some other computer, however powerful, it will be also computable by some Post machine program.

Since there are only countably many programs but uncountably many sets and functions of natural numbers, it follows that most functions are not computable and most sets are not decidable (not even c.e.). We will see some specific examples soon.

4.2 Gödel numbering Post machine programs

Since a program P is a finite sequence of natural numbers, say $P = \langle p_1, p_2, \ldots, p_m \rangle$, each program is easily coded into a single natural number in an effective manner.

One way to do this would be to build an integer $e(P)$ from $P = \langle p_1, p_2, \ldots, p_m \rangle$ in binary notation as follows: Write 1 followed by p_m zeros, followed by another 1 and p_{m-1} more zeros, and so on, ending with 1 followed by p_1 zeros. Finally,

convert this binary string into an integer in the usual way. In other words,

$$e(P) = 2^{p_1} + 2^{p_1+p_2+1} + 2^{p_1+p_2+p_3+2} + \cdots + 2^{p_1+p_2+\cdots+p_m+(m-1)}.$$

$e(P)$ is called the *Gödel number of P.* For example, the Gödel number of the example program $P = \langle 2, 3, 0, 1, 6\rangle$ is 66244 (in decimal). Note that Gödel number of the empty program (which computes the identity function) is 0.

There is an equally effective procedure to convert any natural number to the corresponding Post machine program. For example, given the integer 140 (in decimal), we first write it in binary as 10001100, and then read off the number of consecutive zeros after each 1, starting from the rightmost 1 and moving to the left. This gives us the sequence $\langle 2, 0, 3\rangle$, which decodes into the program

$$\langle 2, 0, 3\rangle \quad = \quad \begin{array}{|ll|}\hline 0 & \mathsf{RIGHT} \\ 1 & \mathsf{FLIP} \\ 2 & \mathsf{IF1GOTO{:}0} \\ \hline \end{array}\ .$$

It is another remarkable fact that there is Post machine program U which, given a finite sequence of numbers as input arguments, treats the first argument as the Gödel number of a program P, and then (by decoding the first argument) is able to simulate P on the remaining arguments. Such programs U are called *universal.*

THEOREM 4 Universal Programs and Computable Functions. *There is a (universal) Post machine program U which computes a partial function $\Phi(m, n)$ of two variables with the following property:* For every Post machine program P with Gödel number $e = e(P)$, the 1-ary function φ computed by P equals the function Φ_e defined by $\Phi_e(n) = \Phi(e, n)$, i.e. $\varphi = \Phi_e$. In particular, every single-variable partial computable function f equals Φ_e for some e.

Now put $W_e = \operatorname{dom} \Phi_e$, and recall that a set is called c.e. if it equals the domain of some partial computable function. Since the sequence $\langle \Phi_e \colon e \in \mathbf{N}\rangle$ contains all partial computable functions, so the sequence $\langle W_e \colon e \in \mathbf{N}\rangle$ forms a list of all c.e. sets.

We define the special set $\emptyset'$ as $\emptyset' = \{e \colon e \in W_e\}$. This is an example of a c.e. set whose complement is not c.e. To see this suppose the complement of $\emptyset'$ is a c.e. set so that for some e we have $n \in W_e \iff n \notin \emptyset'$ for all n. Taking $n = e$ we get

$$e \in W_e \iff e \notin \emptyset' \iff e \notin W_e,$$

a contradiction. It follows that $\emptyset'$ is not computable.

The domain of the function Φ is called HALT, since $\Phi(m, n)$ is defined iff the program with Gödel number m halts on input n. HALT is a c.e. set by definition, with

$$n \in \emptyset' \iff (n, n) \in \mathrm{HALT}, \quad \text{for all } n.$$

Therefore if HALT were computable, so would be $\emptyset'$, which we have seen to be non-computable. Hence HALT itself is non-computable. This is often expressed by saying that *the Halting Problem is uncomputable.*

We thus have examples of c.e. sets HALT and $\emptyset'$ which are not computable, and so their characteristic functions would be examples of non-computable functions.

4.3 Computation on strings

We have been using numbers for computability notions. We now fix an effective one-to-one correspondence between the natural numbers and the binary strings so that computability notions can be extended to strings.

DEFINITION 5. We fix the following one-to-one correspondence between the natural numbers and the binary strings:

$$0 \leftrightarrow \mathbf{\Lambda},\ \ 1 \leftrightarrow 0,\ \ 2 \leftrightarrow 1,\ \ 3 \leftrightarrow 00,\ \ 4 \leftrightarrow 01,\ \ 5 \leftrightarrow 10,\ \ 6 \leftrightarrow 11,\ \ 7 \leftrightarrow 000, \ldots$$

Given a number n, the corresponding binary string is denoted by $\mathbf{str}(n)$. *Given a binary string σ, the corresponding number is denoted by* $\mathbf{num}(\sigma)$.

This correspondence is effective: An algorithm for converting n to $\mathbf{str}(n)$ is obtained as "write $n+1$ in binary notation (without leading zero) and then drop (erase) the leading 1". An algorithm for converting σ to $\mathbf{num}(\sigma)$ is obtained as "prefix σ with an additional 1, regard the resulting string as and integer m written in binary notation, and have the final result to be $m-1$."

With this effective correspondence, we can extend every computability notion for numbers into one for strings, *by converting (translating) between the two types back and forth as needed.*

In particular, we will say "the program P halts on input string δ with output string σ" to really mean "the program P halts on input $\mathbf{num}(\delta)$ with output $\mathbf{num}(\sigma)$".

Another example: We will say that a function $f\colon \{0,1\}^* \to \{0,1\}^*$ is effectively computable to really mean that the function $g\colon \mathbf{N} \to \mathbf{N}$ defined by the "translations" $g = \mathbf{num} \circ f \circ \mathbf{str}$ (i.e., $g(n) = \mathbf{num}(f(\mathbf{str}(n)))$) is effectively computable.

We will say that a set of strings is c.e., if the corresponding set of numbers is c.e.

We could also as easily define computability notions for functions from numbers to strings and vice versa, and in general for subsets of $\mathbf{N}^m \times (\{0,1\}^*)^n$.

4.4 Effective topological notions

The idea of effective computability has been extended to topological notions and is used extensively in an area called *descriptive set theory.* Many classical notions of analysis get refined this way. Here we describe two such notions, *effective open* and *uniformly effective open.*

Recall that an open set G in the Cantor space is a union of basic intervals, $G = \bigcup_{\sigma \in S} N(\sigma)$ for some set S of strings. We refine this definition by requiring the index set over which the union is taken to be computably enumerable:

DEFINITION 6 Effective Open Sets. A subset G of $\{0,1\}^{\mathbf{N}}$ is called *effective open* or *computably enumerable* iff there is a c.e. set S of strings such that

$$G = \bigcup_{\sigma \in S} N(\sigma).$$

A c.e. set of strings is one whose members can be listed by a program (or a partial computable function), so a set G is effective open iff there is a program which prints a list of basic intervals whose union equals G. More precisely, *G is effective open iff there is a partial computable function $\sigma\colon \mathbf{N} \to \{0,1\}^*$ such that*

$$G = \bigcup_{n=1}^{\infty} N(\sigma(n)),$$

where we take $N(\sigma(n)) = \varnothing$ if $\sigma(n)$ is not defined.

Finally we define what it means to say that the sets $G_1, G_2, G_3, \ldots$ are *uniformly effective open.* To be uniformly effective open, it is not enough that each set G_n in the sequence be individually effective open, but the entire sequence must be listed together in an effective way, i.e., there should be a single program listing a double-sequence of basic intervals whose unions form these sets. More precisely:

DEFINITION 7 Uniformly Effective Open Sets. A sequence of sets $G_1, G_2, \ldots$ is *uniformly effective open* if there is a partial computable function $\sigma\colon \mathbf{N} \times \mathbf{N} \to \{0,1\}^*$ such that

$$G_n = \bigcup_{m=1}^{\infty} N(\sigma(m,n)),$$

where we take $N(\sigma(m,n)) = \varnothing$ if $\sigma(m,n)$ is undefined.

5 VON MISES' DEFINITION OF RANDOM SEQUENCE

Undoubtedly, one of the most successful achievements of twentieth century mathematics was the measure-theoretic axiomatization of probability. Proposed by Kolmogorov in 1933, it soon became the essential basis for the study of mathematical probability theory. The generality and elegance of this abstract axiomatic approach — a hallmark of modern formalism mathematics — found wide applicability in almost every situation in probability theory. However, this method with its formalist nature does not directly involve the notion of randomness in any fundamental way, and the rather intuitionistic problem of defining randomness did not arise in its course of development.

Richard von Mises was the first person to focus clearly and deeply on the mathematical essence of randomness in sequences. His pioneering work early in the twentieth century on developing a *frequentist theory of probability* ([Von Mises, 1919; Von Mises, 1981], already shed considerable light on the heart of the matter of

randomness, as Kolmogorov himself has remarked [Li and Vitanyi, 2008, p.50]. For von Mises, random sequences, which he called "collectives", formed the essential basis of his frequentist theory. Subsequently, his work led other mathematicians to carry out further investigations that have clarified the notion of randomness significantly. The far reaching implications of his work is still deeply influencing current research on randomness. Kolmogorov is known as the father of modern mathematical probability theory, and it is perhaps appropriate to call von Mises the founder of the modern mathematical theory of randomness in infinite sequences.

Here we only briefly outline the central ideas of von Mises concerning random sequences. For more details, see his own works [Von Mises, 1919; Von Mises, 1981] and works of van Lambalgen [1987a; 1987b; 1996; 1990].

5.1 Von Mises' definition

The condition of the Borel strong law can be stated as saying that the relative frequency of successes in initial parts of the infinite binary sequence under consideration should have a limiting frequency of 1/2. Recall that this is a condition for *unbiasedness* (subsection 3.2).

The fundamental intuition of von Mises is often summarized as *the invariance of limiting frequency under (admissible) place selections.*

To understand what this means, consider an infinite sequence of turns of a gambling wheel — turn 1, turn 2, and so on — with each turn having two equiprobable outcomes 0 or 1. Let the sequence $x = \langle x_n \rangle$ denote the outcomes of the turns (if the n-th turn of the wheel produces a 0, then $x_n = 0$, else $x_n = 1$).

Suppose that a gambler is observing the outcomes of the turns, and before every turn the gambler decides whether to bet on that turn or not, *perhaps basing the decision on the finite past history of earlier outcomes of the turns.*

Example 1: The gambler chooses to bet on every third turn (turns 3, 6, 9, etc), disregarding earlier history of outcomes altogether ("lucky-third" rule).

Example 2: The gambler chooses to bet after any run of five consecutive 0s.

In any case, bets may not get placed on every turn, but rather on selected turns determined by the gambler's strategy. This results in a subsequence of turns selected for placing bets, say turns $n_1 < n_2 < \cdots < n_k < \ldots$, as shown below:

1	2	...	n_1-1	$\mathbf{n_1}$	n_1+1	n_1+2	...	n_2-1	$\mathbf{n_2}$	···
skip	skip	...	skip	**bet**	skip	skip	...	skip	**bet**	···

Using the strategy, the gambler selects turn n_1 for placing his first bet, turn n_2 for the second bet, etc. In the first example, the subsequence of selected turns for betting is $3 < 6 < 9 < \ldots$, independent of the outcome of the turns. In the second example, the turns on which the gambler bets will depend on the particular infinite sequence of turn outcomes.

Once the turns for betting are all selected, suppose we restrict to the outcome values at only these selected turns — discarding (erasing) the outcome values of those turns on which bets are not placed — and then compute the limiting

frequency for this new restricted subsequence of outcomes. According to von Mises, if the original sequence of outcomes were random, then this new limiting frequency would still equal 1/2, regardless of the gambling strategy being used! Moreover, this crucial property is the essence of randomness, and therefore characterizes it:

DEFINITION 8 Von Mises Randomness, Initial Version. An infinite binary sequence $x \in \{0,1\}^{\mathbf{N}}$ is random if whatever be the gambler's strategy and the resulting turns $n_1 < n_2 < \ldots$ selected for placing bets, the subsequence of x obtained by restricting to these selected turns still has limiting frequency 1/2, i.e.:

$$\lim_{m\to\infty} \frac{1}{m} \sum_{k=1}^{m} x_{n_k} = \frac{1}{2}.$$

In von Mises' terminology, a "turn" (on which the gambler may or may not bet) is called a *place*, and a "strategy" by which the gambler selects which turns to bet on, is called a *place selection rule.*

The "invariance of limiting frequency under admissible place selections" can now be understood as a form of *unpredictability* arising from *unbiasedness*: *No betting strategy of place selections can succeed by improving predictability within a random sequence, since such selections will leave unbiasedness intact (identical limiting frequency for the resulting subsequence).* In other words, not only is the entire sequence unbiased (limiting frequency of 1/2), but there is no hidden biased or unstable subsequence that can be found (by a gambler) using a suitable strategy of place selection.[14]

We quote some relevant remarks of Feller:

> "The painful experience of many gamblers have taught us the lesson that no system of betting is successful in improving the gambler's chances." [Feller, 1968, VIII.2, p. 198]
>
> ... "*[U]nder any system the successive bets form a sequence of Bernoulli trials with unchanged probability of success.* ... The importance of this statement was first recognized by von Mises, who introduced the impossibility of a successful gambling system as a fundamental axiom." [Feller, 1968, VIII.2, p. 199]
>
> "Taken in conjunction with [the] theorem on impossibility of gambling systems, the law of the large numbers implies the existence of the [limiting frequency] not only for the original sequence of trials but also for all subsequences obtained in accordance with the rules of [place selection]." [Feller, 1968, VIII.4, p. 204]

[14] For simplicity we are restricting (by using the von Neumann trick if necessary) only to the special limiting frequency value of 1/2 instead of the general value p $(0 < p < 1)$ used by von Mises. For us, this does not cause much loss of generality as we are focusing only on randomness. Von Mises' chief objective was to develop *the frequentist theory of probability.*

5.2 Mises-Wald-Church randomness

We first formalize the earlier definition of von Mises Randomness.

A *place selection rule* or *betting strategy* is a partial function $\varphi\colon \{0,1\}^* \to \{0,1\}$. (It tells the gambler when to bet: For a binary sequence $\langle x_1, x_2, \ldots, x_n, \ldots\rangle$ of outcomes, the n-th turn is selected for betting according to the strategy φ if and only if $\varphi(\langle x_1, x_2, \ldots, x_{n-1}\rangle) = 1$.)

Given a place selection rule φ and $x \in \{0,1\}^{\mathbf{N}}$ such that $\varphi(\langle x_1, x_2, \ldots, x_{n-1}\rangle) = 1$ for infinitely many n, let n_1 be the least n for which $\varphi(\langle x_1, x_2, \ldots, x_{n-1}\rangle) = 1$, n_2 be the next such n, etc. Then the sequence $\langle x_{n_1}, x_{n_2}, \ldots, x_{n_k}, \ldots\rangle$ is called *the φ-selected part of x.* (It is the subsequence of x obtained by restricting x to those indexes which are selected for betting according to φ.) Thus we say that *the φ-selected part of x has limiting frequency* $1/2$ iff

$$\lim_{m\to\infty} \frac{1}{m} \sum_{k=1}^{m} x_{n_k} = \frac{1}{2}.$$

We can now try to restate the definition of von Mises randomness as follows. *A sequence $x \in \{0,1\}^{\mathbf{N}}$ is random iff for all place selection rules φ for which $\varphi(\langle x_1, x_2, \ldots, x_{n-1}\rangle) = 1$ for infinitely many n, the φ-selected part of x has limiting frequency* $1/2$.

We run into a problem with this definition with its unrestricted use of the universal quantifier in the clause "for all place selection rules". Given any sequence $x \in \{0,1\}^{\mathbf{N}}$, put $A = \{n\colon x_n = 1\}$ if this set is infinite, otherwise put $A = \{n\colon x_n = 0\}$. Define φ by the condition that $\varphi(\langle y_1, \ldots, y_{n-1}\rangle) = 1$ if $n \in A$ and $= 0$ otherwise. It is easy to that the the φ-selected part of x has a limiting frequency equaling either 0 or 1, violating randomness. It follows that no sequence is random!

However, as von Mises points out, this is not a real problem. The defect of the argument is that the rule φ used in the argument *selects a place n based on the outcome value x_n,* and such rules are of course not allowed [Von Mises, 1981, p. 25]. The place selection rules in the definition of randomness are restricted to only certain *admissible* rules, instead being completely arbitrary, and the problem is resolved.[15]

Here is the corrected definition in its original intended form:

DEFINITION 9 Von Mises Randomness. A sequence $x \in \{0,1\}^{\mathbf{N}}$ is random iff for all *admissible* place selection rules φ for which $\varphi(\langle x_1, x_2, \ldots, x_{n-1}\rangle) = 1$ for infinitely many n, the φ-selection of x has limiting frequency $1/2$.

[15] From certain sections of von Mises' detailed discussion of the concept [Von Mises, 1981], it is also clear that he wants a place selection rule, or betting strategy, to be "specifiable in some effective manner". In the modern language of mathematical logic, that could be interpreted as some notion of effective description (say as being effectively computable or at least being definable in some definite language), but von Mises does not precisely specify any such rigorous criterion for defining "admissible".

Abraham Wald [1936] then showed that whenever the set of admissible place selection rules is countable, random sequences according to the von Mises definition do exist, and form a set of full-measure.

This implies that if "admissible" is taken to mean any form effective specifiability using finite sequence of symbols from a finite alphabet (such as effective computability), then the set of admissible rules remains countable, and so von Mises random sequences would exist.

In 1940, Alonzo Church [1940] proposed the use of *effectively computable* (total) place selection rules for precisely defining von Mises randomness. Such random sequences are now called *Church randomness* or *Church stochastic.*

Using *partial computable* place selection rules, we have the following definition:

DEFINITION 10 Mises-Wald-Church Randomness. A sequence $x \in \{0,1\}^{\mathbf{N}}$ is *Mises-Wald-Church random* iff for all *partial computable* place selection rules φ for which $\varphi(\langle x_1, x_2, \ldots, x_{n-1}\rangle) = 1$ for infinitely many n, the limiting frequency of x under φ-selection is $1/2$.

Thus, in 1940, the first precise and rigorous mathematical definition of randomness for infinite sequences was found.

In recent literature, the term *stochastic* is used for randomness defined using limiting frequency after place selections, and so Mises-Wald-Church random sequences are now also called *Mises-Wald-Church Stochastic.*

As explained in Section 3, any notion of randomness must be subjected to the fundamental stochastic laws such as Borel strong law, Borel normality, Symmetric Oscillation, etc.

It is easy to see that every Mises-Wald-Church random sequence satisfies the condition of the Borel strong law, since the place selection rule φ defined by $\varphi(\sigma) = 1$ for all σ is computable, and the resulting subsequence is simply the entire original sequence.

It can also be shown that Mises-Wald-Church random sequences are Borel normal.

But a big blow to the definition came when in 1939 Ville [1939] proved that there are Mises-Wald-Church random sequences which do not satisfy the Law of Symmetric Oscillations: For certain Mises-Wald-Church random sequences x, the relative frequency satisfies $\frac{1}{n}\mathbf{S}_n[x] > 1/2$ for all n. In terms of random walk, this means the position of the walking person stays always to the right of the origin, a violation of the Law of Symmetric Oscillations.

The definition of Mises-Wald-Church randomness can be viewed as the impossibility of any successful algorithmic betting strategy of place selections. Unfortunately, Ville's result shows that this condition is *not sufficient* to guarantee the randomness of a sequence (recall from subsection 3.1 that the condition must be necessary for randomness).

The method outlined in subsection 3.8 of capturing the random sequences using "effective stochastic laws converging effectively" has turned out to be more

successful, and we discuss it in Section 6.

If, instead of considering betting strategies of place-selection and the resulting limiting frequency, we consider *capital betting strategies* (martingales), then the corresponding analog of the Mises-Wald-Church definition — namely the impossibility of any successful suitably algorithmic capital betting strategy — turns out to be more well behaved (see Section 11).

6 MARTIN-LÖF AND SOLOVAY RANDOMNESS

6.1 Martin-Löf randomness

Subsection 3.8 outlined the program of defining a sequence x to be random if it satisfies all "effective probabilistic laws of randomness", where an "effective probabilistic randomness law" is simply an "effective full-measure" set (or, what we called a "typical property" in subsection 1.6). Going to complements, this means that x should not belong to any "effective measure-zero" set (or, in the language of subsection 1.6, that x should not have any "special property".)

The question now, therefore, is how to precisely define "effective measure-zero". Twenty-five years after the Mises-Wald-Church definition, in 1965, a satisfactory solution to this crucial problem was found by the Swedish mathematician Martin-Löf [Martin-Löf, 1966] , which we now describe.[16]

Recall that (Section 2) a set E has measure-zero if there is a sequence of open sets $G_1, G_2, G_3, \ldots$ with each G_n covering E and $\mu(G_n) < 1/n$. Recall also that such a sequence of sets $G_1, G_2, G_3, \ldots$ is uniformly effective open if there is a single program listing the basic intervals whose unions form these sets (subsection 4.4).

Martin-Löf's fundamental idea was that by simply taking the sequence of covering sets G_n to be uniformly effective open, we get the correct notion of "effective measure-zero".

A constructive proof of a probabilistic law of randomness (such as the Borel strong law) would usually proceed this way: Given n, one uniformly builds an effective open set G_n of measure less than $1/n$ such that every sequence in the complement of G_n satisfies the law in question, which immediately establishes that the set of sequences satisfying the law has full measure. The strongest probabilistic law of randomness that we have mentioned, the Law of Iterated Logarithms, is known to have such a constructive proof [van Lambalgen, 1987b, p. 733].

We thus have the following definitions of *effective measure-zero* and *effective full-measure* sets:

[16] Martin-Löf was visiting Kolmogorov in Moscow during 1964–65 and they were working on randomness and complexity of finite objects (Kolmogorov Complexity). Note also that during the twenty-five year period 1940–65, computability theory (recursion theory) was progressing vigorously and expanding its domain into classical analysis, leading to highly refined development of the notion of effectiveness, including effective open sets and effective Borel sets by Kleene, Addison, Mostowski, and others. See [Moschovakis, 1980] for more details.

DEFINITION 11 Martin-Löf. A set $E \subseteq \{0,1\}^{\mathbf{N}}$ is *effective measure-zero* iff there is a uniformly effective sequence open sets, say $G_1, G_2, \ldots$, such that for all n:

(a) $E \subseteq G_n$, and

(b) $\mu(G_n) < 1/n$.[17]

A set has *effective full-measure* if its complement is effective measure-zero.

For example, the set of all sequences with bit value 0 at every third position is effective measure-zero.

Another example of an effective measure-zero set is the set of all sequences in which the bit pattern 0110110 does not occur.

One can think of the sets $G_1, G_2, G_3, \ldots$ as providing a uniformly effective sequence of statistical tests for randomness with stronger and stronger significance.

Finally, we define *Martin-Löf Randomness.*

DEFINITION 12 Martin-Löf. A sequence $x \in \{0,1\}^{\mathbf{N}}$ is *Martin-Löf Random* iff x does not belong to any effective measure-zero set, i.e., iff x belongs to every effective full-measure set.

We can think of an *effective probabilistic randomness law* L to be simply an effective full-measure set L , and think of a sequence x *to be satisfying the law* L iff $x \in L$. We can then restate the definition of Martin-Löf randomness as: x *is Martin-Löf Random iff it satisfies all effective probabilistic randomness laws.*

Martin-Löf also established the remarkable fact that the the set of all Martin-Löf random sequences itself has effective full-measure, that is, the set U of non-Martin-Löf-random sequences is effective measure-zero. This means that there is a sequence of uniformly effective open sets $U_1, U_2, U_3, \ldots$ such that $\mu(U_n) < 1/n$ and $U \subseteq \bigcap_n U_n$. But also $U \supseteq \bigcap_n U_n$ by definition. Hence $U = \bigcap_n U_n$, and thus the sequence $\langle U_n \rangle$ acts as a *universal test* for Martin-Löf randomness: x *is Martin-Löf random iff* $x \notin U_n$ *for some* n. This universal test condition gives an especially simple characterization of Martin-Löf randomness.

6.2 Solovay's characterization of randomness

The Borel-Cantelli Lemma of probability theory implies that if $G_1, G_2, \ldots, G_n, \ldots$ is an infinite sequence of events and the sum of their probabilities converges (as an infinite series), then with probability one, only finitely many of these events can occur.

[17] Of course, instead of the measure bounds $1/n$ for G_n, one can use any sequence of positive rational numbers ϵ_n so long as ϵ_n can be effectively computed from n and the sequence $\langle \epsilon_n \rangle$ converges to zero. It is not hard to see that that if a computable sequence of positive rationals $\langle r_n \rangle$ converges to zero, then given any other computable sequence of positive rationals $\langle s_n \rangle \to 0$, there is a computable subsequence $\langle r_{n_k} \rangle$ of the original sequence with $r_{n_k} < s_k$ for all k. Therefore, the choice of the bounding sequence is completely arbitrary, so long as it forms a computable sequence of positive rationals converging to zero.

The following remarkable result of Solovay shows that this "Borel-Cantelli condition" characterizes Martin-Löf randomness, provided that we restrict the sequence of open sets to be uniformly effective.

THEOREM 13 Solovay. *An infinite binary sequence is Martin-Löf random iff for every uniformly effective sequence $G_1, G_2, \ldots, G_n, \ldots$ of open sets,*

$$\sum_{n=1}^{\infty} \mu(G_n) < \infty \implies x \text{ belongs to only finitely many } G_n s.$$

See Chaitin [1992] for a proof.

The above condition (in the theorem) for characterizing Martin-Löf randomness is known as *Solovay Randomness.*

Note that in Solovay's characterization, the infinite series $\sum_n \mu(G_n)$ is simply assumed to converge, there is no need to assume that it converges in any effective way.

6.3 More general probability spaces

Martin-Löf's and Solovay's definitions for randomness are so general and flexible that they can be applied to any effective separable complete metric space with an effective probability measure. This includes a very large class of probability spaces, including many (perhaps most) probability spaces arising in practice.

Thus the Martin-Löf definition provides a way of assigning a precise meaning of the word "random" in quite general settings. See [González, 2008; Hoyrup, 2008] for more on this.

7 RANDOMNESS OF FINITE STRINGS: KOLMOGOROV COMPLEXITY

We now turn to the "Laplacian Problem" mentioned in the introduction, namely, that of defining randomness for finite sequences. Laplace's observation was that among all the sequences of a fixed large size, only a few "regular" ones have a "rule that is easy to grasp" and Laplace attributes this to those sequences having "a regular cause". The other sequences, "incomparably more numerous", are irregular and we therefore take them to be the random ones.

If we follow this "Laplacian Program" then our problem reduces to precisely isolating the notion of a sequence having a "regular cause" behind it (or being generated by a "rule that is easy to grasp"). This is precisely the philosophical point missing from classical probability theory, which fails to distinguish the strings which we think of being "regular".

This problem was resolved quite satisfactorily in the mid 1960s by Solomonoff, Kolmogorov, and Chaitin. Their theory provided a measure for the *information content* or the *complexity* of a binary string (or more generally of a finite object which can be represented by a binary string) by taking it to be the length of the

"shortest possible complete description" of the string, or its *description-complexity.* The idea is based on our intuition that a relatively simple object will have a short complete description, while a highly complex one will lack a short description which can completely specify it.

Moreover, the related notion of *algorithmic probability* invented by Solomonoff assigns a form of a priori universal probability to binary strings. But unlike classical probability, it takes into account the information-complexity of the string when assigning the probability. As a result, its probability assignments sharply discriminates between the regular strings and the random ones, and provides an explanation of Laplace's intuition into why the regular strings are more likely to have a "cause".

We have been using the term "description" freely, without much qualification. It is important to precisely specify what is meant by a "description". Unrestricted use of the term, as done in natural languages, causes problems, as shown by the following.

7.1 The Berry paradox

Consider the definition:

> *The Berry number is the smallest positive integer that cannot be described in less than eighteen words.*

Since only a finite number of positive integers can be described using less than eighteen words, the Berry number is well-defined, and by definition it cannot be described in less than eighteen words. Yet the above definition describes it using only seventeen words. This is the Berry paradox.

The problem here is with the use of the property of "a number being *described* in certain words", which is not precisely defined, and cannot be defined, as used in the definition of the Berry number, without being circular.

Whatever be its resolution, the Berry paradox reminds us that we have to be careful when talking about the "description" of a number or a string. Once again the concept of algorithm or effective computation allows us to make this precise: We restrict only to "algorithmic descriptions" or "effective descriptions" as defined below. Formally, by an algorithm or program we mean a Post machine program.

DEFINITION 14 Algorithmic Description. Let P be a program and σ be a binary string. We say that a string δ *is a* P*-description of* σ*,* or that δ P*-describes* σ*,* if the program P on the input δ halts with output σ.

The idea here is that whenever the program P halts on the input δ with output σ, we think of the string δ as being an *algorithmic description* of the string σ, according to algorithm P. The computation of σ from input δ by P is thought of as P reconstructing σ from its description δ. This is equivalently called the *decompression* of δ by P (into σ). The description δ is intended to be shorter

than the object being described (the string σ), and therefore it can be viewed as a "compressed" version of the string σ.

Given a program P and a string σ, we now look for the *shortest* string(s) P-describing σ, and take the length of such string(s) as a complexity-measure of the string σ with respect to P. Of course for certain programs P, a string σ may not have any P-description, in which case the complexity of σ (with respect to P) is considered to be infinite.

DEFINITION 15 Algorithmic P-Complexity. The *plain algorithmic complexity of a string σ with respect to the program P*, or simply *the P-complexity of σ,* denoted by $C_P(\sigma)$, is the length of the shortest string(s) P-describing σ, provided that there are such strings. If there is no string which P-describes σ, we let $C_P(\sigma) = \infty$.

There are programs P such that the P-complexity $C_P(\sigma)$ is a finite number for all strings σ (i.e., P has the property that $\forall\sigma\exists\delta(\delta\ P\text{-describes }\sigma)$). Let us call such programs P to be *complexity-finite.*

An example of a complexity-finite program is the empty program E (the program with no instructions), which computes the identity function: Every string E-describes itself.

If $C_P(\sigma)$ is much smaller than the length of σ we think of σ being "well-compressed" by P, since there are P-descriptions of σ much shorter than σ. The ratio $|\sigma|/C_P(\sigma)$ is the "compression factor" for the string σ, with respect to P. For the empty program E, $C_E(\sigma) = |\sigma|$ for all σ, and so the compression factor is 1 for all strings, meaning no string is really "compressed" by E.

On the other hand, there are complexity-finite programs P which compress infinitely many strings by arbitrarily large factors.

Example. Let P be the program informally described as follows. If the first symbol of the input string σ is 0, then P erases this leading 0, shortening its length by 1, outputs the resulting string, and halts; else P outputs the string consisting of $2^{|\sigma|}$ 1s and halts. Then $C_P(\sigma) \leq |\sigma| + 1$ for all σ, since 0σ is a P-description of σ, so P is complexity-finite. But for any n, if we put $\delta_n = 1^n$ and $\sigma_n = 1^m$ where $m = 2^n$, then δ_n is a P-description of σ_n, so $C_P(\sigma_n) \leq |\delta_n| = n$, so the compression factor for σ_n is $|\sigma_n|/C_P(\sigma_n) \geq 2^n/n$. Given any number a, we can find n such that $2^n/n > a$, and so the strings $\sigma_n, \sigma_{n+1}, \ldots$ are all compressed by a factor more than a. In particular, the string σ_{10} (string of 1024 1s) is compressed by P to the string 1111111111, thus by a factor more than 100. More drastically, the string σ_{64} is compressed by a factor of 288230376151711744.

However, a simple but important counting argument (an example of the so-called pigeon-hole principle) shows that no method can compress strings too uniformly.

Given an arbitrary partial function $f\colon \{0,1\}^* \to \{0,1\}^*$, we think of f being a general method of string-description, and think of δ being an f-description of σ if $f(\delta) = \sigma$. (The P-descriptions given by programs P are special "algorithmic" case of this.) We say that a string σ is compressed by f if there is an f-description of σ which is shorter than σ. More generally, given a positive integer b, we say that

σ is b-compressed by f if there is an f-description of σ which is shorter than σ by at least b bits (i.e., if $\exists\delta(f(\delta) = \sigma \wedge |\delta| \leq |\sigma| - b)$). Thus σ is compressed by f iff σ is 1-compressed, that is by at least 1 bit.

THEOREM 16 "Only a small minority of strings compress". *For any string-description method f, less than half of all the strings of length $\leq k$ can be compressed (i.e. 1-compressed) by f. More generally, less than a fraction of $1/2^b$ of all the strings of length $\leq k$ can be b-compressed by f.*

Proof. For any k and b, let A be the set of strings of length $\leq k$, B be the set of strings of length $\leq k - b$, and C be the subset of A consisting of those members which can be b-compressed by f. Note that there is a one-to-one correspondence between C and a subset of B (for each $\sigma \in C$, fix $\delta_\sigma \in B$ with $f(\delta_\sigma) = \sigma$; then the correspondence $\sigma \leftrightarrow \delta_\sigma$ is one-to-one). Also note that $|A|$ (number of members of A) equals $2^{k+1} - 1$ and $|B| = 2^{k-b+1} - 1$, hence the fraction of those strings in A which are b-compressed equals

$$\frac{|C|}{|A|} \leq \frac{|B|}{|A|} = \frac{2^{k-b+1} - 1}{2^{k+1} - 1} < \frac{2^{k-b+1}}{2^{k+1}} = \frac{1}{2^b}.$$

■

For example, among the strings of length not exceeding a thousand (or any other number) bits, more than 99.9% will either not compress at all or compress by at most 9 bits, whatever be the method of string description.

We are of course interested in *shortest possible descriptions* or *best possible compression factor*. Therefore, among the complexity-finite programs we prefer the ones which tend to give *overall* shorter descriptions (better compression) for strings. In other words, given complexity-finite programs P and Q, we regard P to be "better" than Q if the P-complexity of σ, $C_P(\sigma)$ is lower than the Q-complexity $C_Q(\sigma)$ for all strings σ. We can then try to choose a "best" complexity-measure and use it as a standard.

Unfortunately, it is impossible to get such a "best" complexity-finite program in the uniformly strict sense above, because for every complexity-finite program one can find another one which lowers complexity (compresses better) for an arbitrarily large number of strings by an arbitrarily large amount.[18]

Therefore, we will compare programs using a "general overall sense" rather than the uniformly strict sense above, relaxing the relation of one program being better

[18] More precisely, for every complexity-finite P and every m and n, there is another complexity-finite program Q such that $C_Q(\sigma) < C_P(\sigma) - n$ for at least m strings σ. Proof: Without loss of generality assume that $m = 2^k - 1$ for some k, and fix distinct $\sigma_1, \ldots, \sigma_m$ such that $C_P(\sigma_j) > n + k$ for $j = 1, \ldots, m$. This can be done since there are only finitely many strings σ with $C_P(\sigma) \leq n + k$. Also let $\alpha_1, \ldots, \alpha_m$ be a non-repeating listing of all strings of length less than k. Now a program P' can be so designed that it contains coded copies of $\sigma_1, \ldots, \sigma_m$ inside it, and behaves in the following way: If the input is α_j $(1 \leq j \leq m)$, then output the string σ_j and halt; otherwise, truncate the input string by removing its first k symbols and emulate the program P with this truncated input.

than another as follows. For programs P and Q, let us define:

$$P \textit{ matches } Q \text{ (in terms of complexity)} \iff \exists k \forall \sigma (C_P(\sigma) \leq C_Q(\sigma) + k),$$

i.e., from a complexity viewpoint, the program P is regarded to be as good as the program Q or better ("P matches Q") if the P-complexity of every string is less than its Q-complexity *modulo some constant independent of the string.* Let us also call programs P and Q to be *complexity-equivalent* if each one matches the other, i.e., if the difference between P-complexity and Q-complexity is bounded uniformly by some constant.

It now turns out that in this sense, there is indeed a "best compressing" or *optimal* program, which is also unique in the sense that any other optimal program is complexity-equivalent to it.

DEFINITION 17 Optimal or Universal Programs. A program is called *universal,* or *complexity-optimal,* or simply *optimal* if it matches every program.

THEOREM 18 Solomonoff-Kolmogorov-Chaitin Invariance Theorem. *There is an optimal program U. Moreover, every optimal program is equivalent to U.*

Proof. A program U is defined as follows. Given an input string σ, the program U finds the length m of the longest prefix of σ consisting of 0s ($m = 0$ if σ begins with 1 or is empty), and erases this initial run of 0s in σ. If the remaining string begins with a 1, that symbol is also erased, with the final result being the string ρ. U then runs the program with Gödel number m on the input ρ.

To show that U matches P for any P, let $e = e(P)$ be the Gödel number of P and let $k_P = e + 1$. For any string σ, let δ be a shortest P-description of σ, so that $C_P(\sigma) = |\delta|$. Put $\tau = 0^e 1 \delta$, then τ is a U-description of σ, so

$$C_U(\sigma) \leq |\tau| \leq e + 1 + |\delta| = C_P(\sigma) + k_P,$$

where $k_P = e + 1$ is independent of σ. ■

We now fix a particular optimal U and define the *plain algorithmic complexity of a string σ, $C(\sigma)$,* to be $C_U(\sigma)$.

DEFINITION 19 Universal Plain Algorithmic Complexity C. Define the plain algorithmic complexity function C by $C(\sigma) = C_U(\sigma)$, where U is an optimal program fixed permanently.

Being optimal, U matches E, where E is the empty program with E-complexity $C_E(\sigma) = |\sigma|$. So:

COROLLARY 20. *There is k such that for all σ, $C(\sigma) \leq |\sigma| + k$.*

The plain algorithmic complexity $C(\sigma)$ of a string σ is also known as *Kolmogorov complexity.* The term "Kolmogorov complexity" is used in a wide and general sense as synonym for algorithmic complexity, and so *prefix-free complexity* (described next) is also known as Kolmogorov complexity.

The invariance theorem is quite remarkable as it shows that the concept of plain algorithmic complexity is essentially unique and independent of the particular model of computation being used.

The plain algorithmic complexity $C(\sigma)$ of a string σ, can usefully be viewed as a measure of the (algorithmic) information content of the string σ. We therefore have a formal definition for the somewhat vague notion of *information contained in a finite object.*[19]

DEFINITION 21 Randomness and Compressibility for Finite Strings. Let σ be a string and b be a positive integer. We then define:

(a) σ is *b-compressible* if $C(\sigma) \leq |\sigma| - b$;

(b) σ is *compressible* if it is 1-compressible, i.e., if $C(\sigma) < |\sigma|$; and

(c) σ is *random* if σ is not compressible, i.e., if $C(\sigma) \geq |\sigma|$.

THEOREM 22 Existence of Random Strings. *For every n there are random strings of length n. More generally, for any n and $b > 0$ at least $2^n - 2^{n-b+1} + 1$ strings of length n are b-incompressible.*

Proof. Fix n and for each σ with $|\sigma| = n$, pick a string δ_σ of smallest length describing σ. The strings δ_σ are distinct for distinct σ, and there are less than 2^n strings of length less than n, hence by the pigeon-hole principle $|\delta_\sigma| \geq n$ for some σ of length n, so that $C(\sigma) \geq n$, and so σ is a random string of length n.

The second statement is proved similarly. ■

For finite strings, note that we really have relative degrees of randomness. If we have two thousand-bit strings one of which is not compressible and the other, say, 2-compressible but not 3-compressible, then the first one is more random than the second one, but only slightly more so. The complexity measure $C(\sigma)$ therefore provides a measure for the degree of randomness in σ: The smaller the value of $C(\sigma)$ compared to $|\sigma|$, the less random it is. Among all binary strings of a fixed length, the most random are the ones on which the function C achieves its maximum value, and the most non-random ones are the ones on which C is minimized.

Also, short finite strings (such as 111) can be random yet quite simple. This is not surprising when we regard the strings as being produced by a fixed number of flips of a fair coin: If a series of three flips produces 111, there is no reason to the suspect the randomness of the process and so it is easy to also accept the outcome 111 as random. However, a long string of all 1s (say a million bits, all

[19]Information content as defined by *algorithmic* complexity should be contrasted with the one known as Shannon entropy in "classical" information theory, where it is defined in *probabilistic* terms for random variables. While Shannon's theory of information focuses on an entire set of strings associated with varying probabilities, the Kolmogorov theory focuses on an individual string. However, the two notions are closely related, see [Li and Vitanyi, 2008, p.603–608].

1s) is dramatically non-random and will cause us to question the randomness of the process generating it.

THEOREM 23. *The complexity function C is not computable.*

Proof. (The proof resembles the argument of the Berry paradox.)

If C were computable then one can define a program P which given any string σ as input computes a string σ^* such that $C(\sigma^*) > 2|\sigma|$.

By the invariance theorem, there is a constant k such that $C(\sigma) < C_P(\sigma) + k$ for all σ.

Let $\sigma = 1^{k+1}$. Then $C(\sigma^*) > 2|\sigma|$, but since σ is a P-description of σ^* so $C(\sigma^*) \leq C_P(\sigma^*) + k \leq |\sigma| + k < 2|\sigma|$, a contradiction. ■

The plain complexity function C also satisfies much of the intuitive concepts relating to "information content of a finite object." For example, if σ is a long string with significant information content, the information content will not double in the string $\sigma\sigma$, because of redundancy of information: If σ can be described, so can $\sigma\sigma$, with little additional verbiage. To put it differently, every program can be modified by adding only a few lines where the final output is duplicated by "post-processing". We therefore have:

THEOREM 24. *For some k, $C(\sigma\sigma) \leq C(\sigma) + k$ for all σ.*

Random finite strings also satisfy a very general "stochastic" property. Let R be a property of binary strings, that is $R \subseteq \{0,1\}^*$. We say that a binary string σ *satisfies the property* R if $\sigma \in R$, and we say that *almost all strings satisfy the property* R if the fraction of the strings of length n which satisfies R, $|\{\sigma \in R\colon |\sigma| = n\}|/2^n$, approaches 1 as n approaches ∞. For a proof of the following result, see [Sipser, 1997, p.219].

THEOREM 25 General Stochasticity for Finite Random Strings. *If almost all strings satisfy a computable property R, then all except a finite number of random strings satisfy R. The result also holds for b-incompressible strings (for any $b > 0$).*

More properties of general algorithmic complexity will be stated in the next section using a variant of the plain complexity function C that we just described. The new complexity function K will be obtained essentially by restricting the class of strings that are allowed to be algorithmic descriptions: Descriptions must now have a particular form of "unique readability" under some rule.

In either case, we see that by using the concept of algorithm to formalize the fundamental idea of *incompressibility* or *lack of shorter descriptions,* we arrive at a precise and invariant definition of randomness for finite strings that remarkably captures our intuition as described by Laplace. See subsection 8.2 for more discussion on this topic.

We also note that while we used Post-Turing computability (Post Machine programs) as the model of computation, any other model of computation could be used satisfactorily. A particularly recent new approach is to use *Binary Lambda Calculus,* due to John Tromp [2009], to study Kolmogorov Complexity.

The literature of modern research in the subject of Kolmogorov Complexity is vast, see [Li and Vitanyi, 2008].

We end this section with a slight digression by giving a specific example of an application of algorithmic complexity.

7.2 An application to Gödel incompleteness

While most strings are random, only a finite number of them can be proved to be so. For each string σ, let n_σ denote its complexity, i.e., $n_\sigma = C(\sigma)$. In any formalization of mathematics (say ZFC), the relation "the complexity of σ is n" can be formally expressed. Using a variant of the argument of the Berry paradox and the fact that the set of theorems is computably enumerable, we now show that while the sentences

$$\text{"the complexity of } \sigma \text{ is } n_\sigma\text{"}, \qquad \sigma = \mathbf{\Lambda}, 0, 1, 00, 01, 10, 11, 000, 001, \ldots$$

are all true, only finitely many of them can be proved in the theory. We therefore have an "information theoretic version" of Gödel's Incompleteness Theorem. It also follows that there is a "maximum provable complexity": For some m, no string can be proved to have complexity more than m. [Boolos and Jeffrey, 1989; Davis, 1980]

THEOREM 26 Gödel's Incompleteness Theorem, Information-Complexity Version. *Only finitely many of the sentences of the form "$C(\sigma) = n$", where σ ranges over binary strings and n over natural numbers, are theorems of mathematics.*

Proof. Let $\mathbf{C}(x, y)$ be the formula expressing "the complexity of x is y" in the formal theory,[20] and also for each string σ and natural number n, let $\ulcorner\sigma\urcorner$ and $\ulcorner n\urcorner$ be their formal names in the theory. Since the set of theorems can be computably enumerated, there is a program (Post machine) P which, on input string δ, searches through all the theorems to check if any of them is of the form $\mathbf{C}(\ulcorner\sigma\urcorner, \ulcorner n\urcorner)$ with $n > 2|\delta|$, and if one such theorem is found, then outputs the string σ and halts. Using the invariance theorem fix k such that $C(\sigma) \leq C_P(\sigma) + k$ for all σ.

Now run the program P on the input string $\delta_k = 1^k$. Since there are infinitely many n for which $\mathbf{C}(\ulcorner\sigma\urcorner, \ulcorner n\urcorner)$ is a theorem, so P eventually finds such a one with $n = n_0 > 2k$ and halts with an output string $\sigma = \sigma_0$. Since $\mathbf{C}(\ulcorner\sigma_0\urcorner, \ulcorner n_0\urcorner)$ is true, so $C(\sigma_0) = n_0$. But $C(\sigma_0) \leq C_P(\sigma_0) + k \leq |\delta_k| + k = 2k < n_0$, a contradiction. ■

Interestingly, while only finitely many of the true statements "the complexity of σ is n_σ" can be proved (where, again, n_σ denotes the unique natural number

[20] To do this, recall our fixed optimal program U, and let a relation H_U be defined as $H_U(x, n, y) \iff U$ halts with output x in less than n program execution steps on some input string of length not exceeding y. Since H_U is computable, there is a formula $\psi(x, n, y)$ such that $\psi(x, n, y)$ is provable if $H_U(x, n, y)$ is true, else its negation $\neg\psi(x, n, y)$ is provable. Now let $\mathbf{C}(x, y)$ be $\exists n(\psi(x, n, y)) \wedge \neg\exists z < y(\exists n(\psi(x, n, z)))$.

equal to the complexity of σ, i.e. $n_\sigma = C(\sigma)$), it is also easy to see that the true statement "the complexity of σ is $\leq n_\sigma$" can be proved for every string σ. In other words, while we cannot prove that the complexity of σ *equals* n_σ (except for finitely many strings σ), we can prove, for every string σ, that the complexity of σ *does not exceed* n_σ (its true value), without being able to recognize that n_σ is indeed the true value for the complexity of σ.

For a critical discussion of the information-complexity version of Gödel incompleteness, see [van Lambalgen, 1989].

8 THE PREFIX-FREE COMPLEXITY K

While the plain complexity measure $C(\sigma)$ yields a quite satisfactory theory of randomness for finite strings, the function C has some defects. One such defect is about how it relates to the randomness of infinite sequences. If we look at the initial segments $\sigma_n = \langle x_1, x_2, \ldots, x_n \rangle$ of an infinite sequence x, the complexity of the n-th initial segment $C(\sigma_n)$ drops by an undesirable amount for infinitely many n, a phenomenon known as *complexity oscillation* [Li and Vitanyi, 2008, p.143].

Also, for the needs of Solomonoff's theory of Algorithmic Probability, using the literal value of plain complexity measure was not the correct formulation.

Several ideas were developed for dealing variously with such problems, such as monotone complexity, process complexity, decision complexity, and uniform complexity, but it is *prefix-free complexity,* also called *prefix complexity* in short, due to Levin [Kolmogorov, 1974], Gács [1974] and Chaitin (see [Chaitin, 1992] for references), that has now the become the standard for algorithmic (Kolmogorov) complexity, and is denoted by K. It is very much like C, but with a more restricted definition of "description", where only certain classes of strings having a form of unique readability are allowed to be descriptions.

Suppose that we want to pass a string σ directly to a Post machine program as input by placing the string on the tape (all other bits being zero) and starting the program with its head at the beginning bit of the string. Unfortunately, since the tape consists only of 0s and 1s and no special termination markers, there is no general way for any program to determine the end of the string. For example if the head is started on a single 1 with all other tape bits being 0, how can the machine know be sure that there is not another 1 after a trillion bits? Even simpler, how can the machine know if this is supposed to represent the string 1, or 10, or 10000? We express this problem by saying that the plain representation of a binary string is *not properly delimited.* We circumvented this problem by converting a string σ first to its number code $\mathbf{num}(\sigma) = n$ (say), and then passing the unary coded version of n, namely $1^{n+1}0$, to the machine. This coding is an example of a *prefix-free* or *self-delimiting* code, where no code string is a proper initial segment of another, and the input to the machine is uniquely determined. More generally, a set S of strings is said to be *prefix-free* if no string in S is a prefix of another member of S. The set of unary codes, $1^{n+1}0$, $n = 0, 1, 2 \ldots$, do form a prefix-free

set, but is exponentially more inefficient than passing the plain binary string. The following example describes a more efficient prefix-free coding.

Example 1 (The 1-code). Consider the scheme where each binary string σ is coded by $1^{|\sigma|}0\sigma$. The string $1^{|\sigma|}0\sigma$ will be called the 1-*code of* σ. For example, the 14-bit string $\tau = 00011101011100$ is the following 1-code:

$$\overbrace{1\,1\,1\,1\,1\,1\,1\,1\,1\,1\,1\,1\,1\,1\,0}^{1^{|\tau|}0}\ \overbrace{0\,0\,0\,1\,1\,1\,0\,1\,0\,1\,1\,1\,0\,0}^{\tau}$$

Prefix-free 1-coding of τ

The set of 1-codes form a prefix-free set. In fact, by placing the head at the beginning of the 1-code of σ, it can be uniquely decoded back to σ (by first decoding from the initial unary part $1^{|\sigma|}0$). It will take $2|\sigma| + 1$ bits to encode the plain binary string σ into its 1-code (e.g., in the above displayed example $|\tau| = 14$, so the 1-coded string has length 29).

Example 2 (The 2-code). An even more efficient prefix-free coding is be obtained by the following scheme. Given a binary string σ, first express the *length* $|\sigma|$ of the string in plain binary notation $\mathbf{bin}(|\sigma|)$ and then prefix σ with the 1-code of $\mathbf{bin}(|\sigma|)$ to get what we will call *the* 2-*code of* σ. For example, the the string $\tau = 00011101011100$ of the previous example has length 14, which in binary notation is 1110. Since 1110 has length 4, its 1-code is 111101110, and we prefix this to τ to get the 2-code of τ:

$$\overbrace{1\,1\,1\,1\,0}^{1^{|\,\mathbf{bin}(|\tau|)|}0}\ \overbrace{1\,1\,1\,0}^{\mathbf{bin}(|\tau|)}\ \overbrace{0\,0\,0\,1\,1\,1\,0\,1\,0\,1\,1\,1\,0\,0}^{\tau}$$

Prefix-free 2-coding of τ

The 2-code gives a prefix-free encoding which will encode the string σ using $|\sigma| + 2\log_2|\sigma| + 1$ bits. For example, the 2-code of τ shown above consists of 23 bits, a saving of 6 bits over its 1-code.

Further improvements can be made by iterating this method.

Note that all these encodings are effective, i.e., there are simple algorithms for decoding and encoding strings according to any of these schemes. Moreover, the above examples are *length-monotonic,* meaning that longer strings have longer codes.

A real world example of a prefix-free set (over the alphabet of decimal digits) is the set of country dialing codes in the international telephone system.

We will now define prefix-free complexity, which is quite similar to plain complexity and is defined as the length of the "shortest description". *The main difference is that while for plain complexity any binary string could possibly count as a "description", for prefix-free complexity only "prefix-free strings" (under an effective prefix-free encoding) are allowed to be descriptions.*

DEFINITION 27 Prefix-free Complexity Functions. A partial function $\psi\colon \{0,1\}^* \to \{0,1\}^*$ mapping strings to strings is *prefix-free* if its domain is prefix-free. Given

a partial computable prefix-free function ψ, the *prefix-free ψ-complexity function* K_ψ is defined by letting $K_\psi(\sigma)$ to be the length of the shortest string(s) δ for which $\psi(\delta) = \sigma$, and putting $K_\psi(\sigma) = \infty$ no such strings exist.

An example of a prefix-free complexity function ψ_2 is obtained by coupling the decoding function for the 2-codes with the optimal program U for plain complexity C as follows.

Let D_2 denote the set of all 2-coded strings, so that D_2 is a prefix-free set. The decoding function $d_2 \colon D_2 \to \{0,1\}^*$ for decoding 2-codes establishes a one-to-one correspondence between D_2 and $\{0,1\}^*$, but we regard it as a partial function from $\{0,1\}^*$ to $\{0,1\}^*$. If a string δ is not in D_2, then $d_2(\delta)$ is not defined, so that the domain of d_2 is D_2.

Now define ψ_2 to be the function which, given an input string δ, regards δ as 2-code for some string, decodes it into $d_2(\delta)$ and sends this decoded string to the program U as input. U, in turn, runs with $d_2(\delta)$ as input, and may halt with an output string σ, in which case we put $\psi_2(\delta) = \sigma$. If δ is not in D_2 so that $d_2(\delta)$ is not defined, or if U does not halt on input $d_2(\delta)$, then we leave $\psi_2(\delta)$ as undefined.

Clearly ψ_2 is partial computable and its domain is a subset of D_2, hence it is a partial computable prefix-free function. Let the corresponding prefix-free complexity function K_{ψ_2} be denoted simply by K_2.

We thus have an example of a prefix-free complexity function K_2.

How does K_2 compare with the plain complexity function C? Given a string σ with plain complexity $n = C(\sigma)$, let τ be a string of length n for which U outputs σ on input τ. Let δ be the 2-code for τ, so that $d_2(\delta) = \tau$, and thus $|\delta| = |\tau| + 2\log_2|\tau| + 1 = n + 2\log_2 n + 1$. Now by definition, $\psi_2(\delta) = \sigma$. Moreover, since encoding-decoding using the 2-code is length monotonic, so there is no string δ' shorted than δ with $\psi_2(\delta) = \sigma$, hence:

$$K_2(\sigma) = |\delta| = n + 2\log_2 n + 1 = C(\sigma) + 2\log_2 C(\sigma) + 1,$$

which shows that $K_2(\sigma)$ exceeds $C(\sigma)$ by $2\log_2 C(\sigma) + 1$.

Of course K_2 is not "optimal" and there are "better" prefix-free partial computable functions ψ giving lower lower complexity values. In order to get a "universal optimal" function, we need the following fundamental result.

THEOREM 28 Invariance Theorem for Prefix-free Complexity. *There is an partial computable* optimal *prefix-free function ξ. That is, for each partial computable prefix-free function ψ there is a constant k satisfying:*

$$K_\xi(\sigma) \leq K_\psi(\sigma) + k, \quad \textit{for all } \sigma.$$

DEFINITION 29 Universal Prefix-free Complexity. The *Prefix-free Algorithmic Complexity* $K(\sigma)$ is defined by taking $K(\sigma) = K_\xi(\sigma)$, where ξ is a permanently fixed optimal partial computable prefix-free function.

How does the prefix-free complexity value $K(\sigma)$ compare with the plain complexity value $C(\sigma)$?

Roughly speaking, since fewer strings are allowed to be descriptions of σ under prefix-free complexity K, we may expect a somewhat higher value for the prefix-free complexity $K(\sigma)$ than $C(\sigma)$ (modulo a constant). This indeed turns out to be true.

On the other hand, we can use 2-codes to convert any plain description δ into a prefix-free description of length $|\delta| + 2\log_2 |\delta| + 1$. More precisely, for the specific prefix-free complexity function K_2 in the example above, we saw that

$$K_2(\sigma) = C(\sigma) + 2\log_2 C(\sigma) + 1.$$

For the optimal prefix-free complexity K, we would get an overall lower value for $K(\sigma)$ than $K_2(\sigma)$ (modulo a constant), and so $C(\sigma) + \log_2 C(\sigma) + 1$ is only an upper bound for $K(\sigma)$. The following result gives a standard upper bound for $K(\sigma)$.

THEOREM 30. *Modulo additive constants,*

$$C(\sigma) \leq K(\sigma) \leq C(\sigma) + 2\log_2 |\sigma|.$$

Proof. The first inequality follows from the invariance theorem for C (optimality of C) since the partial computable function ξ used to define K is just one specific "program".

To prove the second inequality, note that we had earlier already established that $K_2(\sigma) \leq C(\sigma) + 2\log_2 C(\sigma)$ and $C(\sigma) \leq |\sigma|$, modulo constants. Combining the two, the result follows. ■

Of course, using more efficient prefix-free encodings, this result can be further sharpened.

K has many nice properties which are lacking in C.

From now on, we will be using the prefix-free complexity function K in place of C. In particular, the definitions for compressibility and randomness for finite strings are redefined as follows.

DEFINITION 31 Randomness and Compressibility for Finite Strings. Let σ be a string and b be a positive integer. We then define:

(a) σ is *b-compressible* if $K(\sigma) < |\sigma| - b$; otherwise it is *b-incompressible*;

(b) σ is *compressible* if it is 1-compressible, i.e., if $K(\sigma) < |\sigma|$; otherwise it is called *incompressible*; and

(c) σ is *random* if σ is incompressible, i.e., if $K(\sigma) \geq |\sigma|$.

Since K assigns a higher complexity value to strings than C (uniformly modulo a constant), strings compress a "little less" under K than under C, and so random strings now become "more numerous". In particular, the existence theorem for random strings (which holds for any complexity measure) remains valid.

THEOREM 32 Existence of Random Strings. *For every n there are random strings of length n. More generally, for any n and $b > 0$ at least $2^n - 2^{n-b+1} + 1$ strings of length n are b-incompressible.*

8.1 Properties of finite random strings

We mention two more "stochastic" properties of finite random strings which conform to our intuition as evidence that incompressibility is the correct definition of randomness for finite strings.

THEOREM 33. *Long random strings have a "balanced" number of 0s and 1s. More precisely, for any $\epsilon > 0$ there is k such that for all random strings σ of length $n > k$, we have*

$$\left| \frac{S_n[\sigma]}{n} - \frac{1}{2} \right| < \epsilon,$$

where $S_n[\sigma]$ denotes the number of 1s in σ.

THEOREM 34. *Any run of zeros or ones in a random string n is asymptotically bounded above by $O(\log n)$. That is there is k and a constant a such that for every random string σ of length $n > k$, the longest run of 0s in σ is less than $a \log n$.*

8.2 Kolmogorov Complexity as vindication of Laplace

We end our discussion of randomness for finite strings with the position that Kolmogorov Complexity provides a satisfactory solution to Problem 2 of the Introduction. As outlined and evidenced above, the *incompressibility* definition of randomness for finite strings conforms quite well to our intuition. In fact, an important test of randomness for finite strings now is to apply standard computational compression programs to the the string in question and check if it compresses or not.

Kolmogorov Complexity also provides a strong vindication of all of Laplace's intuitions, by classifying strings according to their complexity: The lower the K value, the more "regular" the strings, and the higher the K value the more "irregular" (or random) they are.

Cause and Regularity Laplace mentions that we perceive a "cause" in strings which are "regular", "those in which we observe a rule that is easy to grasp." Kolmogorov complexity provides a precise and objective way to define this idea, using *short effective descriptions.* Given a string σ with small complexity value $K(\sigma)$, its "cause" (or a "rule that is easy to grasp") is any of its minimal descriptions, i.e. a minimal length strings (of length $K(\sigma)$) which describes σ via the optimal algorithm. If $K(\sigma)$ is not small compared to $|\sigma|$, we regard the string σ as "irregular" or random. This interpretation of cause and regularity is obtained by classifying strings by their complexity, i.e., by measuring how short an effective description is possible. The invariance of

Kolmogorov complexity under all possible methods of effective description (for sufficiently large strings) shows that this is not an arbitrary measure of complexity, but is essentially an objective one.

Rarity of Regularity Laplace mentions that the "irregular sequence ... are incomparably more numerous" compared to the "regular" ones. This is again confirmed by that fact that if cause or regularity is defined as *descriptions which are sufficiently short,* then the irregular sequences automatically become "incomparably more numerous". We saw this in the theorem which showed that only a small minority of strings compress well.

Probability of Regular Strings Laplace explains that while the regular strings are much less numerous, if we observe a highly regular but long string, "we seek a cause whenever we perceive symmetry", and it is *"more probable"* that "this event ought to be the effect of a regular cause" than "that of chance". For example, let σ be the thousand bit long string containing the pattern $01010101\ldots$ throughout the entire string. If we observe σ, our intuition tells us that in a sense not explained by classical probability theory, it is remarkably different from another random string generated by a thousand coin flips. Classical probability, being information-blind, will assign the same probability $1/2^{|\sigma|} = 1/2^{1000}$ to σ, as well as to all other thousand bit string. But the algorithmic probability (using Kolmogorov complexity) of σ is $1/2^{K(\sigma)}$. If in addition to random coin flips we also consider "effective causes", then algorithmic probability remarkably explains Laplace's intuition. As the string σ is regular (has a short description), $K(\sigma)$ will be quite smaller than $|\sigma| = 1000$, and so its algorithmic probability $1/2^{K(\sigma)}$ will be much higher compared to another random string of the same length, since the algorithmic probability of a random string will be, by definition of randomness for finite strings, at most $1/2^{|\sigma|} = 1/2^{1000}$. In other words, the probability that σ was generated by an "effective cause" will be higher than the probability that it was generated randomly, by a factor of at least $2^{|\sigma|-K(\sigma)} = 2^{1000-K(\sigma)}$. With the conservative estimate of $K(\sigma) = 950$, this factor is 1125899906842624.

At last, the Laplace Program is realized.

9 KOLMOGOROV-CHAITIN RANDOMNESS AND SCHNORR'S THEOREM

We now return back to randomness for infinite sequences.

We will say that an infinite binary sequence x is *b-incompressible* if every initial segment $\langle x_1, x_2, \ldots, x_n\rangle$ of x is b-incompressible as a finite string, i.e. if

$$K(\langle x_1, x_2, \ldots, x_n\rangle) > n - b \quad \text{for all } n.$$

We say that the infinite binary sequence x is *incompressible* it is b-incompressible for some b, i.e., if *no initial segment of x can be compressed by more than a fixed number of bits.*

This property of incompressibility of an infinite binary sequence x can be regarded as an information-complexity definition of randomness for x.

DEFINITION 35 Kolmogorov-Chaitin Randomness. An infinite binary sequence x is *Kolmogorov-Chaitin Random* if x is incompressible (no initial segment of x can be compressed by more than a fixed number of bits), or in other words, if

$$\text{For some } b: \quad K(\langle x_1, x_2, \ldots, x_n \rangle) > n - b \quad \text{for all } n.$$

Remark: The term "Kolmogorov-Chaitin random" is not in standard use. In the literature it is known variously as "Chaitin random", "Levin-Chaitin random", "Levin-Chaitin-Schnorr random", K-incompressible, etc.

The definition of Kolmogorov-Chaitin randomness appears to be significantly different when compared to the definition of Martin-Löf randomness (or Solovay randomness).

The notion of Martin-Löf randomness is based on effective stochastic laws — or predicates (properties) which are satisfied almost surely, i.e. with probability one. Randomness of a sequence x in the Martin-Löf sense is defined not by looking at the individual sequence x alone, but using an entire collection of predicates of x, and the definition appears to be in the form a "second-order" definition, involving universal quantification over predicates of sequences.[21]

On the other hand, the Kolmogorov-Chaitin definition does not directly refer to any external object other than the sequence x itself and the complexity measure K. Instead of taking an "external top down" approach, it looks at x "from inside" in terms of initial segments (a purely internal view), measuring their information-complexity using K, and declares x to be random if none of the initial segments admit any substantially shorter description. The Kolmogorov-Chaitin definition therefore reduces the definition of randomness for infinite sequences to that for finite strings, establishing a fundamental connection between the two notions.

This striking dissimilarity makes the following celebrated theorem of Schnorr truly remarkable.

THEOREM 36 Schnorr's Theorem. *A sequence is Martin-Löf random if and only if it is Kolmogorov-Chaitin random.*

For a proof, see any of [Nies, 2009; Downey and Hirschfeldt, 2010; Li and Vitanyi, 2008; Chaitin, 1992].

The equivalence of Martin-Löf randomness with Kolmogorov-Chaitin randomness forms the basis of the assertion that Martin-Löf's definition has truly captured the notion of randomness for infinite sequences, and therefore gives a satisfactory

[21] Of course, since the predicates in question are effectively enumerated, in actuality the universal quantifier is reduced to range over natural numbers, but we are referring to the form of the definition in classical terms.

solution to this classic problem in the philosophy of mathematics and statistics (Problem 1 of the Introduction).

Moreover, as the Kolmogorov-Chaitin definition shows, the notions of randomness for finite and infinite sequences are fundamentally linked, and therefore the solutions to both Problems of the introduction can be given simultaneously in an interconnected fashion. (A characterization of Martin-Löf randomness in terms of plain complexity C has also been obtained, but it is a much more complicated condition compared to the one for K.)

From now on, by a *random infinite sequence* we will mean a Martin-Löf random or equivalently Kolmogorov-Chaitin random sequence.

DEFINITION 37 Randomness for Infinite Sequences, Final Version. An infinite binary sequence x will be called *random* if it is Martin-Löf random or equivalently Kolmogorov-Chaitin random.

The assertion that Martin-Löf randomness or equivalently Kolmogorov-Chaitin randomness captures the "true notion of randomness" conforming to our intuition is sometimes called the *Martin-Löf-Chaitin thesis.* The Martin-Löf-Chaitin thesis, like the Church-Turing thesis for the definition of algorithm, is not a mathematical proposition that can be be proved or refuted. We discuss it further in Section 12.

9.1 Properties of infinite random sequences

We list here some regularity properties of infinite random sequences as evidence that we have the correct definition of randomness for infinite sequences. Recall that when we say "random" without qualification, we mean Martin-Löf random, or equivalently Kolmogorov-Chaitin random.

For proofs and further details of the following facts, see [Calude, 1994; Li and Vitanyi, 2008; Nies, 2009; Downey and Hirschfeldt, 2010].

THEOREM 38 Effective Place Selections Preserve Randomness. *Let $\langle x_1, x_2, \dots \rangle$ be a random infinite binary sequence and $\varphi\colon \{0,1\}^* \to \{0,1\}$ be a partial computable function. Suppose that $\varphi(\langle x_1, \dots, x_{n-1}\rangle) = 1$ for infinitely n, and $n_1 =$ the least n such that $\varphi(\langle x_1, \dots, x_{n-1}\rangle) = 1$, $n_2 =$ the next such n, etc. Then the subsequence $\langle x_{n_1}, x_{n_2}, \dots, x_{n_k}, \dots \rangle$ is also random.*

COROLLARY 39. *Every random sequence is Mises-Wald-Church random.*

COROLLARY 40 Computable Restrictions Preserve Randomness. *If x is random, and $n_1 < n_2 < n_3 < \dots$ form a computable sequence of strictly increasing numbers then the sequence $\langle x_{n_1}, x_{n_2}, x_{n_3}, \dots \rangle$ is also random.*

The above result remains true if $n_1, n_2, \dots$ form a computable sequence of distinct numbers (not necessarily increasing).

COROLLARY 41. *If x is random, then neither the set $\{n\colon x_n = 1\}$ nor its complement $\{n\colon x_n = 0\}$ can contain any infinite computably enumerable set (they are* immune*). In particular, neither these sets nor the sequence x is computable.*

A real number a is called *computable* if the set $\{(m,n)\colon m/n < |a|\}$ is a computable subset of $\mathbf{N} \times \mathbf{N}$.

A real number $a \in [0,1]$ is called *random* if there is a random sequence $x \in \{0,1\}^{\mathbf{N}}$ such that x is the sequence of digits in a binary expansion of a, i.e. $a = \sum_{n=1}^{\infty} x_n/2^n$.

COROLLARY 42. *If $a \in [0,1]$ is a random real number, then a is not computable. In particular, a is irrational, and in fact transcendental, since all algebraic real numbers are computable.*

The following result shows that like being convergent, being random is an eventual property.

THEOREM 43. *The randomness of a sequence is a "tail" property. In particular:*

(a) If the sequence y is obtained from x by altering only finitely many values of x, then x is random iff y is random.

(b) If $x = \langle x_1, x_2, \ldots, x_n, \ldots \rangle$ and $y = \langle x_{n+1}, x_{n+2}, \ldots, x_{n+k}, \ldots \rangle$ is obtained by removing the first n terms of x (an n-step shift), then x is random iff y is random.

THEOREM 44. *If x is random, then:*

(a) x satisfies the law of iterated logarithms.

(b) x is absolutely Borel normal: If the real number $\hat{x}$ having x as its binary expansion digits is expanded in base $b > 1$, then the resulting expansion is Borel normal in base b.

9.2 An example of a specific random sequence: $\boldsymbol{\Omega}$

It is clear that a set S of strings is prefix-free iff the set of basic intervals $N(\sigma)$ (in the Cantor space) indexed by strings from S form a disjoint family, and so the open set formed by their union has a measure equal to the sum of the measures of the basic intervals, i.e., $\mu\left(\cup_{\sigma\in S} N(\sigma)\right) = \sum_{\sigma\in S} \mu(N(\sigma)) = \sum_{\sigma\in S} 1/2^{|\sigma|}$. Therefore:

$$\text{For any prefix-free set } S \text{ of strings, we have:} \qquad \sum_{\sigma\in S} \frac{1}{2^{|\sigma|}} \leq 1,$$

an important fact known as *the Kraft inequality.*

Recall now the partial computable prefix-free function ξ that used to define the optimal prefix-free complexity K. We define a real number $\boldsymbol{\Omega}$, called the *Halting Probability* or *Chaitin's Omega,* by:

$$\boldsymbol{\Omega} = \sum_{\sigma\in \mathrm{dom}(\xi)} \frac{1}{2^{|\sigma|}}.$$

Since the domain of ξ is a prefix-free set, $\mathbf{\Omega} \leq 1$ by the Kraft inequality.

Fix a program Q which computes the partial computable function ξ. Then Q halts on an input binary string δ iff $\xi(\delta)$ is defined. Suppose now that a fair coin is flipped until some initial segment of the sequence of flips is found to be in the domain of ξ, or equivalently until it generates a string having an initial segment on which Q halts. Of course, in many cases no such string will be generated (i.e. we may have an infinite sequence of flips for which there is no initial segment string on which Q ever halts). In this sense, $\mathbf{\Omega}$ denotes the probability that Q halts if its input is generated by a random sequence of coin flips. This is the reason $\mathbf{\Omega}$ is known as the Halting Probability. It is a non-computable real number, and so is transcendental.

Alternatively, define an open set by:

$$G_\Omega = \bigcup_{\sigma \in \mathrm{dom}(\xi)} N(\sigma).$$

Then G_Ω is effective open and $\mathbf{\Omega}$ equals the Lebesgue measure of G_Ω.

The infinite sequence of bits forming the binary expansion of $\mathbf{\Omega}$ is also denoted by $\mathbf{\Omega}$. $\mathbf{\Omega}$ is then a random infinite sequence, our first example of a specific random infinite sequence.

$\mathbf{\Omega}$ has many remarkable properties, see [Bennett, 1979; Calude, 1994; Chaitin, 1992].

10 RELATIVE AND STRONGER RANDOMNESS. HIERARCHIES

Given two sequences $x = \langle x_1, x_2, \dots \rangle$ and $y = \langle y_1, y_2, \dots \rangle$, we merge them into a single sequence $x \oplus y$ by intertwining the terms as follows:

$$x \oplus y \ = \ \langle x_1, y_1, x_2, y_2, \dots, x_n, y_n, \dots \rangle.$$

We call $x \oplus y$ the *join of x and y*. Since each of x and y can be extracted from $x \oplus y$, we can, from an information content point of view, think of $x \oplus y$ as perfectly combining the *information* contained in x and that in y, without any "information loss".

From the properties of randomness given earlier, it is immediate that if $x \oplus y$ is random then so are both x and y. However, if x and y are random, it does not follow that $x \oplus y$ is random. As a drastic example, if $x = y$, then $x \oplus y$ cannot be random since every alternate pair of consecutive bits of $x \oplus y$ would be identical, which allows one to devise a simple successful gambling system against it (it also violates Borel normality as neither the pattern 010 nor the pattern 101 occurs in it).

So the question arises: *Under what conditions on x and y do we have $x \oplus y$ random?*

This was answered beautifully by van Lambalgen in terms of the notion of *relative randomness,* which we now discuss. Roughly speaking, x is random relative to

y (abbreviated x is random in y) if even a complete knowledge of y does not improve the predictability of the bits of x. This is exactly opposite of the situation in our drastic example of $x = y$, where knowledge of y allows us to perfectly predict x, or, in other words, the information about x can be obtained (in this case completely) from information of y. We may therefore expect, that the relation of one sequence being random relatively to another is some form of "information-independence", although it is not a priori clear that this relation should be symmetric (the drastic example shows that the relation must be irreflexive).

We proceed to formalize this idea.

DEFINITION 45 Effective Open, Relative Version. Let $z \in \{0,1\}^{\mathbf{N}}$. A set G is said to be *effective open relative to z,* or simply *effective open in z,* or in symbols G *is* $\Sigma_1^0(z)$*,* if there is an effective open set H such that for all $x \in \{0,1\}^{\mathbf{N}}$:

$$x \in G \iff x \oplus z \in H.$$

Notice how, in the formation of G, the information of z becomes available: As before, G is still the union of basic intervals which are enumerated by some computation, but now that computation is also allowed to use any additional information from z as needed.

Thus, if G is effective open, then G is effective open in z, for any z (additional information from z is available, but not used, in the computation which enumerates basic intervals forming G). On the other hand if G is effective open in z and z is computable, then G is already effective open (since z can be computed by some program P, the computation which enumerates basic intervals forming G does not get any additional help by knowing z, since any information about z could also be obtained by calling P as a subprogram).

Similarly, we define a sequence of sets being "uniformly effective open in z".

DEFINITION 46 Uniformly Effective Open (Relative). A sequence $G_1, G_2, \ldots$ of sets is *uniformly effective open in z,* or *uniformly* $\Sigma_1^0(z)$*,* if there are sets $H_1, H_2, \ldots$, uniformly effective open, such that for all n,

$$x \in G_n \iff x \oplus z \in H_n.$$

Again, we define a set being "effective measure-zero in z" just by changing the old definition with "effective open" replaced by "effective open in z". (This process, known as *relativization,* can actually be carried out fruitfully throughout most of computability theory.)

DEFINITION 47 Effective Measure-Zero, Relative Version. A set E is *effective measure-zero in z* if there are sets $G_1, G_2, \ldots$, uniformly effective open in z, such that $\mu(G_n) < 1/n$ and $E \subseteq \bigcap_n G_n$.

And, finally:

DEFINITION 48 Relative Randomness. *x is random relative to y* (or *x is random in y*) iff x does not belong to any set effective measure-zero in y.

Now we can state van Lambalgen's theorem:

THEOREM 49 Van Lambalgen. *For $x, y \in \{0,1\}^{\mathbf{N}}$, the following conditions are all equivalent to each other:*

(a) $x \oplus y$ is random;

(b) y is random in x and x is random in y;

(c) y is random and x is random in y;

Van Lambalgen's theorem is remarkable, because the apparently weaker third condition in the theorem implies the second. In particular, if x and y are random, then

$$x \text{ is random in } y \implies y \text{ is random in } x,$$

which is surprising, because, as we mentioned earlier, this symmetry is not at all clear a priori.

The existence theorems all remain valid under relative randomness. E.g., for any y, the set of sequences random in y form a full-measure set whose complement is effective measure-zero in y. If y is computable, this does not give a new collection, as then the set of sequences random in y equals the set of random sequences. But if y is not computable, we may have stronger versions of randomness. E.g., if y is the characteristic function of the uncomputable set $\emptyset'$, or if y is $\mathbf{\Omega}$ (more precisely, the sequence y consists of the digits in the binary expansion of $\mathbf{\Omega}$), then the collection of sequences random in y cannot contain $\mathbf{\Omega}$ anymore, and therefore form a strictly smaller subclass of the random sequences, known as the 2-random sequences. Using a prefix free function $\xi_{\mathbf{\Omega}}$ partial computable in $\mathbf{\Omega}$ which is optimal for all prefix free functions partial computable in $\mathbf{\Omega}$, one can now define the *halting probability relative to* $\mathbf{\Omega}$ as:

$$\mathbf{\Omega}_2 = \sum_{\sigma \in \mathrm{dom}(\xi_{\mathbf{\Omega}})} \frac{1}{2^{|\sigma|}}.$$

Then $\mathbf{\Omega}_2$ is 2-random, while $\mathbf{\Omega}$ is random but not 2-random.

This process can be iterated to stronger and stronger versions of randomness, which we now describe in greater generality.

10.1 The arithmetical hierarchy and n-randomness

We will use a notation where relations are identified with predicates: If R is a 3-place relation, then we abbreviate "$(a, b, c) \in R$" by simply writing $R(a, b, c)$, etc.

A relation $A \subseteq \mathbf{N}^k \times \{0,1\}^{\mathbf{N}}$ is called *effective open* if there is a computably enumerable $E \subseteq \mathbf{N}^k \times \{0,1\}^*$ such that for all $m_1, \ldots, m_k \in \mathbf{N}$ and $x \in \{0,1\}^{\mathbf{N}}$,

$$A(m_1, \ldots, m_k, x) \iff E(m_1, \ldots, m_k, \sigma) \text{ for some initial segment } \sigma \text{ of } x.$$

A relation is called *effective closed* if its complement is effective open, and it is called *computable* (or *effective clopen*) if it is both effective open and effective closed.

Starting with the computable relations as a basis, we can define relations of higher complexity by adding a series n "alternating quantifiers" ranging over natural numbers as follows. We define a relation A to be Σ^0_n $(n \geq 1)$ if there is a computable relation R such that for all $m_1, \ldots, m_k \in \mathbf{N}$ and $x \in \{0,1\}^{\mathbf{N}}$,

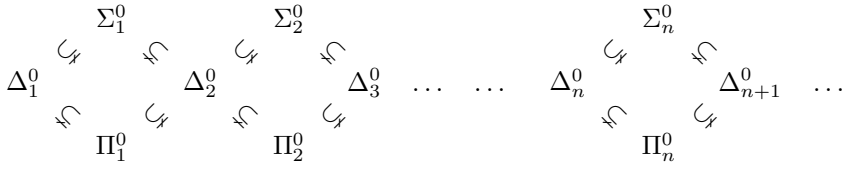
$$A(m_1, \ldots, m_k, x) \iff (\exists p_1)(\forall p_2) \ldots (Qp_n) R(p_1, \ldots, p_n, m_1, \ldots, m_k, x),$$

where Q stands for "$\exists$" if n is odd and for "$\forall$" if n is even.

We also define a relation to be Π^0_n if its complement is Σ^0_n, and a relation to be Δ^0_n if it is both Σ^0_n and Π^0_n.

It then turns out that the class Σ^0_1 coincides with the class of effective open relations, and Δ^0_1 is same as the class of computable relations.

Moreover, this indeed gives a strict hierarchy of classes of relations defined by their definitional complexity (the number of alternating quantifiers), with the class Δ^0_n strictly contained in each of Σ^0_n and Π^0_n, both of which are strictly contained in Δ^0_{n+1}, as shown below.



This hierarchy is called the *Arithmetical Hierarchy.* (It is a refinement of the finite levels of the classical hierarchy of Borel sets in analysis, but only *effective* countable unions and intersections are allowed. See [Rogers Jr, 1987; Odifreddi, 1992; Moschovakis, 1980] for more details.)

Finally, we can define n-randomness.

DEFINITION 50. A sequence z is called *n-random* iff there is no Σ^0_n-effective measure-zero set containing z, or more precisely if there is no sequence $H_1, H_2, \ldots$ of sets such that the relation H defined by $H(n,x) \iff x \in H_n$ is Σ^0_n, $\mu(H_n) < 1/n$ for all n, and $z \in H_n$ for all n.

We say z is *arithmetically random* iff z is n-random for all $n = 1, 2, \ldots$.

One can show that it does not make any difference to the definition whether we require the sets H_n to be all open or not.

Thus, 1-random is same as being random (i.e., Martin-Löf-random), and we have a sequence of stronger and stronger notions of randomness, corresponding to the levels of the arithmetical hierarchy. This hierarchy is indeed strict, meaning that $(n+1)$-randomness implies n-randomness, but for each n there is an n-random sequence which is not $(n+1)$-random. E.g., $\boldsymbol{\Omega}$ is 1-random but not 2-random, and

$\mathbf{\Omega}_2$ is 2-random but not 3-random, and so on. Moreover,

$$\begin{aligned} x \text{ is 2-random} &\iff x \text{ is random relative to } \mathbf{\Omega} \\ &\iff x \text{ is random and } \mathbf{\Omega} \text{ is random relative to } x. \end{aligned}$$

The first of these equivalences follow from standard facts about the arithmetical hierarchy, and the second from van Lambalgen's theorem.

Recall that an infinite binary sequence $x = \langle x_1, x_2, \dots \rangle$ was defined to be Kolmogorov-Chaitin random if for some $b > 0$, $K(\langle x_1, x_2, \dots, x_{n-1} \rangle) > n - b$ for all n. If prefix complexity K is replaced by plain complexity C, then one produces an empty definition, as was proved by Martin-Löf: *There is no sequence* $x = \langle x_1, x_2, \dots \rangle$ *such that for some* $b > 0$, $C(\langle x_1, x_2, \dots, x_{n-1} \rangle) > n - b$ *for all* n. However, it was also shown that the sequences x which satisfy the condition:

$$\text{for some } b > 0,\ C(\langle x_1, x_2, \dots, x_{n-1} \rangle) > n - b \text{ for infinitely many } n,$$

form a full-measure set which is contained in the set of random sequences. Such sequences are sometimes called *Kolmogorov Random.* Remarkably, it was established recently that the Kolmogorov random sequences are precisely the 2-random ones.

10.2 Other stronger notions of randomness

Many other notions of randomness stronger than 1-randomness have been studied. E.g., by relaxing the condition that $\mu(H_n) < 1/n$ in the definition of n-randomness to $\lim_n \mu(H_n) = 0$, one obtains the notion of weak-$(n+1)$-randomness, which lies strictly between n-randomness and $(n+1)$-randomness.

One can extend the arithmetical hierarchy into the transfinite using computable ordinals, which results in the *hyperarithmetical hierarchy.* Martin-Löf first suggested the notion of hyperarithmetical randomness. Beyond the hyperarithmetical classes, there is an even more comprehensive hierarchy known as the *analytical hierarchy.* At the first level of this hierarchy are the Π^1_1 relations. A relation A is called Π^1_1 if there is an arithmetical relation (Σ^0_2 is enough) B such that

$$A(n_1, \dots, n_k, x) \iff \forall y\, B(n_1, \dots, n_k, x \oplus y).$$

(The class of Π^1_1 relations includes all hyperarithmetical sets and more, and hence certainly all arithmetical sets as well.)

The notion of randomness in the sense of Martin-Löf has recently been extended to the class of Π^1_1 sets (and has been named Π^1_1-ML-randomness) fruitfully by Hjorth and Nies [Hjorth and Nies, 2007], and they have established the analog of Schnorr's theorem and other results.

The *strongest notion of randomness* that appears to have been studied so far is called Π^1_1*-randomness* by Hjorth and Nies [Hjorth and Nies, 2007]. The union of all Π^1_1 measure-zero sets is itself a Π^1_1 measure-zero set (the largest Π^1_1 measure-zero set),, so Π^1_1-randomness is defined as membership in the complement of the largest Π^1_1 measure-zero set.

10.3 Reducibility and degrees of computability

Given $A \subseteq \mathbf{N}$ and $z \in \{0,1\}^{\mathbf{N}}$, we say that A *is computably enumerable in* z, or in symbols $A \in \Sigma^0_1(z)$ if there is a c.e. set $B \subseteq \mathbf{N} \times \{0,1\}^*$ such that $\forall n(A(n) \iff \exists k B(n, \langle z_1, z_2, \ldots, z_k\rangle))$.

We say that A *is computable in* z, or A *is Turing-reducible to* z, in symbols $A \leq_T z$, if both A and its complement are computably enumerable in z. Finally a sequence $x \in \{0,1\}^{\mathbf{N}}$ is *computable in* z or *Turing-reducible to* z if the set $\{n\colon x_n = 1\}$ is computable in z. Roughly speaking, $x \leq_T y$ means that x can be computed by a program which has access to the bits of y in order, or even more vaguely y is computationally at least as complex as x.

The notion of Turing-reducibility is reflexive and transitive, and the corresponding equivalence relation, called *Turing equivalence,* $x \equiv_T y \iff x \leq_T y \wedge y \leq_T x$ generates equivalence classes known as *Turing degrees.* The study of Turing-reducibility and degrees has been one of the most important areas of classical recursion theory. There are several other types of computational reducibilities, generally stronger than Turing-reducibility, that are important for the theory of computability.

The interaction between randomness and Turing-reducibility (and other computational reducibilities not introduced here) has also been studied, and has generated fruitful applications in both directions.

Another notion straddling both randomness and computability theory that has been studied extensively is that of K-triviality. A sequence $x = \langle x_1, x_2, \ldots\rangle$ is *K-trivial* if $\forall n(K(\langle x_1, \ldots, x_n\rangle) \leq K(n) + b$ for some constant b. This property is quite orthogonal to that of randomness. It is known that x is K-trivial iff every random sequence is random relative to x.

Many other notions that interact with both computability theory and randomness form a part of current research, which is progressing vigorously. We refer the reader to [Nies, 2009; Downey and Hirschfeldt, 2010] where further extensive references can be found.

11 RANDOMNESS VIA MARTINGALES. OTHER FREQUENTIST DEFINITIONS

In the previous section, we considered randomness notions stronger than Martin-Löf randomness. Now we will focus on weaker notions of randomness, which are perhaps more important from a philosophical viewpoint.

11.1 Schnorr randomness

A critique of Martin-Löf randomness by Schnorr was that it yields too strong a notion of randomness as its notion of defining effective measure-zero is not effective enough. In order for a sequence of uniformly effective open sets $G_1, G_2, \ldots$ to define an effective measure-zero set via intersection, it is not enough, according

to Schnorr, that their measures effectively approach zero by just having *effective bounding* (e.g. as $\mu(G_n) < 1/n$), but we need the measures $\mu(G_n)$ of the sets themselves to be computable real numbers (uniformly in the index n). Using this stronger criteria for being effective measure-zero, we get a weaker notion of randomness, called *Schnorr randomness.*

Schnorr randomness has been studied extensively, but it fails to have certain regularity properties of Martin-Löf randomness. Two examples are:

(a) Unlike Martin-Löf randomness, Schnorr randomness does not possess a uniform test, i.e. the class of Schnorr random sequences cannot be defined as the complement of the intersection of a sequence of uniformly effective open sets $G_1, G_2, \ldots$ such that $\mu(G_n) < 1/n$ and such that $\mu(G_n)$ is a computable real number uniformly in n.

(b) The van Lambalgen theorem fails for Schnorr randomness. In fact, there is Schnorr random sequence $z = x \oplus y$ such that that the two halves x and y are Turing equivalent. This does not conform well with the intuitive notion of randomness.

11.2 Randomness defined by martingales

Recall that in von Mises type definitions of randomness, one uses a the concept of a betting strategy, or more precisely a *place selection rule,* to select places on which to bet, and after selections are all done, one checks if the limiting frequency value has become biased or not (subsections 5.1 and 5.2). The definition does not a priori have anything to do with the *amount of bet.* Such properties are called *stochasticity* (as opposed to randomness) in the modern mathematical literature, but note that this is essentially a matter of terminology.

We now introduce a concept of betting strategy involving the amount of bet, or equivalently the *capital* of the gambler ("total money in pocket"), at each stage of betting.

We think of the infinite sequence $x = \langle x_1, x_2, x_3, \ldots \rangle$ being revealed to the gambler, bit by bit, in order. Before each bit is revealed, the gambler may bet an amount a predicting the value of the bit (one of "next bit revealed will be 0" or "next bit revealed will be 1"). We will assume the *fairness condition* that if the prediction turns out to be correct, the gambler *gains* an amount of a (capital increases by a), otherwise the gambler *loses* the same amount a (capital decreases by a).

To formalize this type of strategy, we think of a finite binary string σ of length $|\sigma| = n - 1$ as representing the n-th stage of the game, so that σ consists of the bits of the infinite sequence that have been revealed so far (before the n-th bit is revealed), and let $F(\sigma)$ denote the gambler's capital at this stage.

Suppose that at stage σ, with capital $F(\sigma)$, the gambler bets an amount a predicting the next bit to be 0. If the gambler turns out to be correct, then σ is

extended to $\sigma 0$ and the capital goes up by a to $F(\sigma 0) = F(\sigma) + a$, but if incorrect, then σ is extend to $\sigma 1$ and the capital goes down by a to $F(\sigma 1) = F(\sigma) - a$. Similarly, if the gambler had predicted the next bit to be 1 (with same bet amount a) then we would have σ extending to σ_1 and $F(\sigma 1) = F(\sigma) + a$ if the gambler turns out to be correct, and σ extending to σ_0 and $F(\sigma 0) = F(\sigma) - a$ if the gambler is incorrect. A final case is when the gambler chooses not to bet at this stage, which is expressed by having $a = 0$ and $F(\sigma 0) = F(\sigma 1) = F(\sigma)$ (no change in capital).

All cases can be summarized using a single *zero sum condition*:

$$\overbrace{F(\sigma 0) - F(\sigma)}^{\text{capital change if next bit is 0}} + \overbrace{F(\sigma 1) - F(\sigma)}^{\text{capital change if next bit is 1}} = 0.$$

Therefore we make the following definition.

DEFINITION 51 Martingales. A *martingale* or *capital betting strategy* is a function $F\colon \{0,1\}^* \to \mathbf{R}$ satisfying two conditions:

(a) $F(\sigma) \geq 0$, $\forall \sigma$ (finiteness condition); and

(b) $[F(\sigma 0) - F(\sigma)] + [F(\sigma 1) - F(\sigma)] = 0$, $\forall \sigma$ (zero-sum or fairness condition).

Given $x \in \{0,1\}^{\mathbf{N}}$ and a martingale $F\colon \{0,1\}^* \to \mathbf{R}$, we say that the martingale F *succeeds on* x if the capital becomes unbounded on outcome sequence x, i.e., if

$$\sup_n F(\langle x_1, x_2, \ldots, x_{n-1}\rangle) = +\infty.$$

An example of a martingale is where the gambler always predicts an outcome of 0 with the bet being a fixed fraction r $(0 < r < 1)$ of the available capital. This martingale is the function recursively defined as $F(\sigma 0) = (1+r)F(\sigma)$ and $F(\sigma 1) = (1-r)F(\sigma)$; or, if initial capital is 1, more explicitly as $F(\sigma) = (1+r)^m(1-r)^n$ if σ is a string of length $m+n$ with m zeros and n ones.

Another example is where the gambler always predicts an outcome of 0 betting the entire amount of available capital ("bold play"). If the initial capital is 1, this martingale is given as follows. If 1 does not occur in σ, then $F(\sigma) = 2^{|\sigma|}$, else $F(\sigma) = 0$.

It can be shown that the concept of a martingale is really a generalization of place selection rule. To each place selection rule φ one can assign an especially simple type of martingale F_φ (one which always uses a constant fraction of the existing capital as the next bet), such that the φ-selected part of x has limiting frequency $1/2$ iff F_φ does not succeed on x. A converse association is also possible.

Possible definitions of randomness either using place selection rules (as done by von Mises) or using martingales are both examples of *characterizations of randomness via impossibility of gambling systems,* with the place selection method known as the *frequentist* approach, while the martingale method may be called non-frequentist.

11.3 A martingale characterization of randomness

A martingale F is said to be *computably enumerable* if the relation R defined by

$$R(m, n, \sigma) \iff \frac{m}{n+1} < F(\sigma)$$

is computably enumerable as a subset of $\mathbf{N} \times \mathbf{N} \times \{0,1\}^*$.

THEOREM 52 Martingale Characterization of Martin-Löf Randomness. *A sequence x is random (i.e. Martin-Löf random) iff no computably enumerable martingale succeeds on x.*

Thus we now have three different but equivalent definitions for randomness. This last characterization in terms of martingales gives a definition of randomness using the *unpredictability* approach.

If, instead of computably enumerable martingales, we require stronger effectiveness conditions on the martingales, we obtain weaker notions of randomness, as we will see now.

In the following, we consider only rational valued martingales.

DEFINITION 53. A *partial computable martingale* is a partial function $F\colon \{0,1\}^* \to \mathbf{Q}$ satisfying, for all σ:

(a) $F(\sigma) \geq 0$;

(b) If $F(\sigma)$ is defined, so is $F(\tau)$ for any prefix τ of σ;

(c) $F(\sigma 0)$ is defined iff $F(\sigma 1)$ is defined, and if so, then:

$$[F(\sigma 0) - F(\sigma)] + [F(\sigma 1) - F(\sigma)] = 0;$$

(d) The relation R defined by

$$R(\sigma, m, n) \iff F(\sigma) = m/(n+1)$$

is a computably enumerable subset of $\{0,1\}^* \times \mathbf{N} \times \mathbf{N}$.

A *computable martingale* is a partial computable martingale which is total (i.e., whose domain is $\{0,1\}^*$). (It can be seen that the graph of a computable martingale is computable, not just computably enumerable.)

Here are the main notions of randomness arising out of these types of martingales.

DEFINITION 54. A sequence is *partial computably random* if no partial computable martingale succeeds on it.

A sequence is *computably random* if no computable martingale succeeds on it.

Of course, every partial computably random sequence is computably random. It can be shown that

$$\begin{aligned} \text{(Martin-Löf) Random} &\implies \text{Partial Computably Random} \\ &\implies \text{Computably Random} \\ &\implies \text{Schnorr Random,} \end{aligned}$$

but none of these implications can be reversed [Nies, 2009].

11.4 Non-monotonic betting strategies

So far, we have seen two distinct types of betting strategies giving rise to notions of randomness:

- *Martingales,* or capital betting strategies, which lead to definitions of randomness based on the failure of the martingale, such as *partial computable randomness* and *computable randomness* that we just saw.
- *Place selection rules,* which lead to definitions of randomness based on the limiting frequency of the selected part, such as *Mises-Wald-Church stochasticity (or randomness)*, and *Church stochasticity (or randomness)* that we saw in subsection 5.2.

These four types just mentioned all are notions weaker than (Martin-Löf) randomness.

There is a further generalization possible in the type of betting allowed — called *non-monotonic betting* — that actually tightens the notions further, and makes them more robust.

To understand non-monotonic betting, both for martingales and for place selection rules, suppose that the bits of the sequence $x = \langle x_1, x_2, \ldots, x_n, \ldots \rangle$ lay covered on an infinitely long table (instead of being revealed serially one by one). The gambler now uncovers the bits in a not-necessarily increasing order, and along the way decides which places to select or to bet on.

For example, the gambler may choose to first uncover the ninth place to find the value of x_9, then uncover the fourth place to find x_4, and then, based on these two observations, decide to select or bet on the seventeenth position (before uncovering it). After x_{17} is uncovered, the gambler may choose to next uncover either x_3 or x_7, depending on whether x_{17} turns out to be 0 or 1, and so on.

We omit the formal details and hope that the above informal description makes it intuitively clear what non-monotonic betting is. In particular, it can be applied both to martingales and to place selection rules as follows.

DEFINITION 55. A sequence x is called *Kolmogorov-Loveland Random* if no computable non-monotonic martingale succeeds on it.

A sequence x is called *Kolmogorov-Loveland Stochastic* if for every computable non-monotonic place selection rule, the selected part has limiting frequency $= \frac{1}{2}$.

A nice feature of non-monotonic betting strategies is that it does not matter whether we use "computable" or "partial computable" in the definitions above, as in each case it results in an equivalent notion. In other words, changing the above definitions to "partial computable" will not give us a stronger notion of randomness. Thus the non-monotonic forms have a kind of robustness that was not there for monotonic betting strategies, since partial computable randomness is a strictly stronger notion of randomness compared to computable randomness, and Mises-Wald-Church stochasticity is strictly stronger than Church stochasticity.

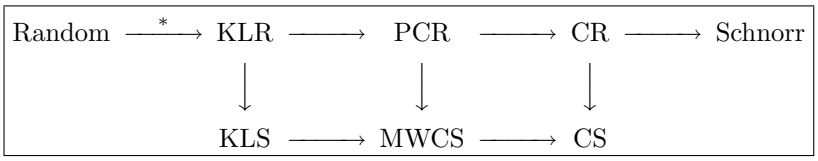
Since martingales are more general than place selection rules, it follows that a martingale version will give a stronger notion of randomness than the corresponding stochastic (i.e limiting frequency after place selections) version. Thus Kolmogorov-Loveland randomness is strictly stronger than Kolmogorov-Loveland stochasticity, partial computable randomness is strictly stronger than Mises-Wald-Church stochasticity, and computable randomness strictly stronger than Church stochasticity.

We summarize these notions in the following table. Each notion in this table implies the one directly below it and also the one to the right of it.

Strategy type	Monotonicity and computability of strategy function		
	Non-monotonic	Monotonic	
	Partial or total	Partial	Total
Martingale	*Kolmogorov-Loveland Random* (KLR)	*Partial Computably Random* (PCR)	*Computably Random* (CR)
Place selection	*Kolmogorov-Loveland Stochastic* (KLS)	*Mises-Wald-Church Stochastic* (MWCS)	*Church Stochastic* (CS)

Six Randomness Notions for Various Types of Betting Strategies

In fact, with "Random" standing for Martin-Löf randomness, "Schnorr" for Schnorr randomness, and the rest of the abbreviations as in the table, we have the following implications.



Implication Diagram for Weak Randomness Notions

Each arrow represents an implication, and almost all the implications above are strict (they cannot be reversed). However, for the implication marked with "$*$", it is not known whether the implication can be reversed. In other words we have:

Question. *While every Martin-Löf random sequence is Kolmogorov-Loveland random, is the converse true?*

This is perhaps the biggest open problem in current research on randomness. Many researchers feel that the answer is no, although work of Merkle et al [Merkle *et al.*, 2006] have shown that the two notions are rather close.

Nies and Miller have published a list of open problems [Miller and Nies, 2006] in the area, some of which have been solved since then.

11.5 Can we resurrect von Mises?

While Kolmogorov-Loveland randomness is the only major notion of randomness that remains close to Martin-Löf randomness, it does not have the true spirit of von Mises' idea of randomness, since von Mises's definition, which is based on place selections, is a *truly frequentist* one, that is, it is defined in terms of limiting frequency, while the martingale notions are all defined in terms of capital growth. Therefore, even if it turns out that Kolmogorov-Loveland randomness is equivalent to Martin-Löf randomness, one would still be looking for a characterization of Martin-Löf randomness in terms of a frequentist condition.

In recent literature, the term *stochasticity* is used for randomness defined in terms of a frequentist condition, or more precisely using the limiting frequency of place selections. Among these, the one closest to Martin-Löf randomness is *Kolmogorov-Loveland stochasticity,* as the above diagram of implications indicates. Unfortunately, like Mises-Wald-Church stochasticity, Kolmogorov-Loveland stochasticity gives random sequences which do not satisfy the Law of Symmetric Oscillations, and therefore is rather far from Martin-Löf randomness.

Li and Vitanyi writes in their 2008 book [Li and Vitanyi, 2008, p. 158]:

> "[T]he problem of giving a satisfactory definition of infinite Martin-Löf random sequences in the form proposed by von Mises has not yet been solved."

Thus, the search for a true frequentist characterization of Martin-Löf randomness continues.

11.6 The Ergodic Theorem as a Frequentist Definition

The general version of the Birkhoff Ergodic Theorem is not fully effective (see [Avigad, 2009; Hoyrup, 2008]). However, we consider the version of the theorem which is known as the "law of frequencies". We restrict to Lebesgue measure on the Cantor space.

Say that a measurable map $U\colon \{0,1\}^{\mathbf{N}} \to \{0,1\}^{\mathbf{N}}$ is *measure preserving* if $\mu(U^{-1}[A]) = \mu(A)$ for all measurable A, and a map $T\colon \{0,1\}^{\mathbf{N}} \to \{0,1\}^{\mathbf{N}}$ is *ergodic* if T is measure preserving and for all measurable A, if $T^{-1}[A]\Delta A$ is measure-zero then $\mu(A) = 0$ or $\mu(A) = 1$.

We now state the Birkhoff Ergodic Theorem in a slightly variant form:

THEOREM 56 The Birkhoff Ergodic Theorem as the Law of Frequencies. *If* $T\colon \{0,1\}^{\mathbf{N}} \to \{0,1\}^{\mathbf{N}}$ *is ergodic,* E *is measurable, and* $U\colon \{0,1\}^{\mathbf{N}} \to \{0,1\}^{\mathbf{N}}$ *is measure preserving, then (we put* $[\![P]\!] = 1$ *if the statement* P *is true, else* $[\![P]\!] = 0$*):*

$$\textit{For almost all } x\textit{:} \qquad \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} [\![T^k(U(x)) \in E]\!] \;=\; \mu(E),$$

i.e., the frequency with which the T*-orbit of* $U(x)$ *enters* E *approaches* $\mu(E)$*.*

We also view of this theorem, the Law of Frequencies, as a form of *equidistribution* (as in Weyl equidistribution): The T-orbit of $U(x)$ is equidistributed.

Now if φ is a place selection rule satisfying the property that for almost all x, $\varphi(\langle x_1, x_2, \ldots, x_{n-1}\rangle) = 1$ for infinitely many n, then the map $U_\varphi\colon \{0,1\}^{\mathbf{N}} \to \{0,1\}^{\mathbf{N}}$

$$U_\varphi(x) = \varphi\text{-selected part of } x$$

is defined for almost all x and is a measure preserving (and continuous) map. Finally, take T to the left shift map $T(\langle x_1, x_2, \ldots\rangle) = \langle x_2, x_3, \ldots\rangle$, and $E = N(1) = \{x \in \{0,1\}^{\mathbf{N}}\colon x_1 = 1\}$ = the basic interval consisting of sequences with 1st bit = 1.

With the above setting, the condition in the von Mises definition coincides precisely with the condition in the Birkhoff Ergodic Theorem above, at least in the case when the domain of U_φ has full measure. Furthermore, if E is allowed to range of over the basic intervals, then the condition of Borel normality of is obtained, which we view as equidistribution of the T-orbit of $U(x)$.

Perhaps the main weakness of the von Mises definition is the lack of the requirement of general equidistribution, which is illustrated in the approach of [Knuth, 1998]. The notion of equidistribution is more general than the notion of limiting frequency, while still being in the frequentist spirit.

Suppose we try to strengthen the Mises-Wald-Church definition by requiring Borel normality of the φ-selected part, not just existence of unbiased limiting frequency, i.e., we demand that for x to be random, the φ-selected part of x has to be Borel normal in base 2 for every partial computable place selection rule φ. This does not really change anything, as all Mises-Wald-Church stochastic sequences are Borel normal. However, instead of being limited to limiting frequency, it casts the definition in terms of equidistribution of the T-orbit of $U(x)$, where T is the left-shift map. In other words, in Mises-Wald-Church stochasticity, the condition that the place-selected part y of x must satisfy is equivalent to the equidistribution of the following sequence of sequences obtained from y:

$$\langle y_1, y_2, y_3, \ldots\rangle,\ \langle y_2, y_3, y_4, \ldots\rangle,\ \langle y_3, y_4, y_5, \ldots\rangle,\ \ \ldots$$

We think that a frequentist definition of randomness should allow more general forms of equidistribution, so long as the method of forming the sequence is uniformly effective and ergodic (see [Knuth, 1998] for examples).

For example, suppose that the natural numbers are partitioned into an infinite number of infinite uniformly computable subsets, say into the sets $\{1, 3, 5, 7, \dots\}$, $\{2, 6, 10, 14, \dots\}$, $\{4, 12, 20, 28, \dots\}$, etc. Then for a random x, if y is obtained from x by effective place selection, we expect that the sequences obtained by restricting y to each of these subsets

$$\langle y_1, y_3, y_5, \dots \rangle,\ \langle y_2, y_6, y_{10}, \dots \rangle,\ \langle y_4, y_{12}, y_{20}, \dots \rangle,\ \ \dots$$

should be equidistributed (which follows from the Ergodic Theorem). It is not a priori clear that Mises-Wald-Church stochasticity guarantees this. (Kolmogorov-Loveland stochasticity allows more general forms of U, but the T operator is still the same left shift.)

Perhaps the following definition can be taken to be an "ergodic generalization" of von Mises' definition of randomness, where we think of U as "place selection":

DEFINITION 57. $x \in \{0,1\}^{\mathbf{N}}$ is *random* iff for all sufficiently effective $E \subseteq \{0,1\}^{\mathbf{N}}$, sufficiently effective ergodic T, and sufficiently effective measure preserving U, we have:

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} [\![T^k(U(x)) \in E]\!] \ = \ \mu(E),$$

i.e., the frequency with which the T-orbit of $U(x)$ enters E approaches $\mu(E)$.

The three instances of the phrase "sufficiently effective" are deliberately left open to interpretation. We ask:

> **Question.** *Are there interpretations of "sufficiently effective" in the above definition which characterize Martin-Löf randomness?*

For example, it is known that with "sufficiently effective" interpreted as Δ^0_1 (computable) in all three instances, every Martin-Löf random sequence is random in the sense of the above definition.[22] However, we are not aware of any result that positively answers the above question.

12 CONCLUSION. THE MARTIN-LÖF-CHAITIN THESIS

In this article, we introduced the reader to definitions of the notion of a random sequence using the three main ideas described in Section 1.6 that have dominated algorithmic randomness (cf. [Downey and Hirschfeldt, 2010]):

[22] See [González, 2008; Hoyrup, 2008] for a proof and a further list of problems. See also [V'yugin, 1997; V'yugin, 1999]. In [González, 2008; Hoyrup, 2008], this is studied in a more general setting (such as effective probability metric spaces), and results have been obtained for Schnorr randomness, but it is stated as an open problem for Martin-Löf randomness. It appears that the problem of characterization of Martin-Löf randomness in this ergodic way is open even in the specific case of the Cantor Space with Lebesgue measure.

- *Randomness as typicality.* According to this stochastic or measure-theoretic idea, randomness of a sequence means its membership in all effective full-measure sets, or equivalently that the sequence "passes" all effective stochastic tests. The first major stochastic laws, the Borel strong law and Borel normality go back to 1909. Martin-Löf randomness (1966), the first definition that is almost universally accepted now as the correct one, is defined using this approach.

- *Randomness as incompressibility.* This is an information-complexity approach that views the sequence "from inside". According to this idea, randomness means lack of short complete descriptions, or equivalently a high degree of algorithmic complexity, for all initial parts of the sequence.

- *Randomness as unpredictability.* According to this approach, randomness means the impossibility of devising a successful betting strategy against the sequence in question. In particular, this means that knowledge of some part of the sequence does not help to predict any other unknown bit. Two types of definitions of randomness arise from this deeply intuitive approach: The special *frequentist* type definitions first put forward by von Mises as "invariance of limiting frequency under admissible place selections", and the more general *non-frequentist* type found in the definitions of randomness using martingales.

The ideal definition of randomness would be one which naturally and simultaneously satisfies the criteria given by these three approaches.

12.1 The Martin-Löf-Chaitin thesis

Following Delahaye [Delahaye, 1993], we use the term *Martin-Löf-Chaitin Thesis* for the assertion that *Martin-Löf randomness and equivalently Kolmogorov-Chaitin randomness is the correct formulation of the intuitive notion of randomness for sequences.* In this sense, it parallels the classic Church-Turing thesis, and is not a mathematical proposition to be proved or disproved. The Church-Turing thesis turned out to be highly successful in capturing the intuitive notion of *algorithm.*

Delahaye has carried out a detailed comparison between the Church-Turing thesis and the Martin-Löf-Chaitin thesis, and concludes that in both cases, the resulting precise definitions provide "profound insights to the mathematical and philosophical understanding of our universe." Delahaye admits that the Church-Turing thesis is "more deeply attested" and that the definition of randomness of sequences is "more complicated" compared to the definition of algorithm, but hopes that with time the Martin-Löf-Chaitin thesis will reach a level of certainty similar to the Church-Turing thesis. We think that overall, Delahaye's assertions still remain valid.

In the past few decades, there has been a vast amount of research activity in the area of algorithmic randomness. Many definitions of randomness for sequences

have been studied extensively, but none was found to be clearly superior to the Martin-Löf definition. Compared with other notions, it appears to be of optimal strength: Weaker notions turn out to be too weak, and the stronger ones too strong. In this way, the Martin-Löf-Chaitin thesis has gained strength in a slow but steady fashion.

The proliferation of definitions of randomness for sequences makes the field harder for non-experts, but it should not be regarded negatively. It is an indication of the richness of the area, and the associated healthy and lively activity provides refinements and insights deep into the subject. Recall that while we consider the Church-Turing thesis as more satisfying, there was an even larger number of associated notions of computability, both stronger and weaker, that were (and still are) studied fruitfully.

Perhaps the strongest evidence for the Martin-Löf-Chaitin thesis available so far is Schnorr's theorem, which establishes the equivalence between a naturally formulated "typicality definition" (Martin-Löf randomness) and a naturally formulated "incompressibility definition" (Kolmogorov-Chaitin randomness).

Another justification of the Martin-Löf-Chaitin thesis is provided by the simplicity of the definition of Martin-Löf randomness within the arithmetical hierarchy. As seen in Schnorr's theorem,

$$x \text{ is random} \iff \exists b \, \forall n \, K(\langle x_1, \ldots, x_{n-1} \rangle) \geq n - b.$$

This shows that Martin-Löf randomness has a Σ^0_2 definition (which also follows from the existence of a universal test). Most other definitions of randomness are more complicated, and situated at higher levels of the arithmetical hierarchy. In fact, the definitional complexity of Martin-Löf randomness is at the lowest possible level of the arithmetical hierarchy, assuming that any definition of randomness must satisfy the two basic axioms:

(a) No random sequence should be computable.

(b) The set of random sequences has full-measure.

It then follows that no definition of randomness can be Π^0_2 or simpler, as it is a standard "basis theorem" that any Π^0_2 set of full-measure contains computable sequences.

We also doubt if the resolution of the question of whether there are Kolmogorov-Loveland random sequences which are not Martin-Löf random will have much impact on the Martin-Löf-Chaitin thesis.

However, a *purely frequentist* natural characterization of Martin-Löf randomness can substantially increase the strength of the Martin-Löf-Chaitin thesis. While the characterization in terms of computably enumerable martingales is a nice "unpredictability definition", it is not as intuitive nor as frequentist as the von Mises definition. This is perhaps the most unsatisfying gap in the current state of affairs.

To summarize, we believe that while the Martin-Löf-Chaitin thesis is not (yet) as strong as the Church-Turing thesis, the two problems of the introduction, namely defining randomness for sequences and strings that captures our mathematical intuition of these objects, have essentially been solved quite satisfactorily as described in this article. It is perhaps not too surprising that the definition of randomness, which in all cases presupposes the definition of algorithm, has turned out to be more complicated than the definition of algorithm itself.

ACKNOWLEDGMENTS

The author wishes to thank Ananda Sen for help with some of the references. The author is also indebted to Prasanta S. Bandyopadhyay and the anonymous referee for several useful suggestions.

BIBLIOGRAPHY

[Aczel, 2004] A.D. Aczel. *Chance: A Guide to Gambling, Love, the Stock Market & Just About Everything Else*. Thunder's Mouth Press, New York, 2004.

[Avigad, 2009] J. Avigad. The metamathematics of ergodic theory. *Annals of Pure and Applied Logic*, 157(2-3):64–76, 2009.

[Bailly and Longo, 2007] F. Bailly and G. Longo. Randomness and determinism in the interplay between the continuum and the discrete. *Mathematical Structures in Computer Science*, 17(02):289–305, 2007.

[Becher and Figueira, 2002] V. Becher and S. Figueira. An example of a computable absolutely normal number. *Theoretical Computer Science*, 270(1-2):947–958, 2002.

[Belshaw and Borwein,] A. Belshaw and P. Borwein. Strong Normality of Numbers. `http://www.cecm.sfu.ca/personal/pborwein/PAPERS/P211.pdf`. Contains material added and updated after Belshaw's masters thesis.

[Belshaw, 2005] A. Belshaw. On the normality of numbers. Master's thesis, Simon Fraser University, 2005.

[Beltrami, 1999] E.J. Beltrami. *What is random? Chance and Order in Mathematics and Life*. Copernicus (Springer-Verlag), New York, 1999.

[Bennett, 1979] C.H. Bennett. On Random and Hard-to-Describe Numbers. Technical Report RC-7483, IBM Watson Research Center, Yorktown Heights, New York, 1979. reprinted in *Randomness And Complexity, from Leibniz To Chaitin* (C. Calude, ed.), World Scientific 2007, pp 3–12.

[Bennett, 1998] D.J. Bennett. *Randomness*. Harvard University Press, Cambridge, Mass., and London, England, 1998.

[Boolos and Jeffrey, 1989] G. S. Boolos and R. C. Jeffrey. *Computability and Logic*. Cambridge University Press, 1989.

[Borel, 1909] E. Borel. Les probabilités dénombrables et leurs applications arithmétiques. *Rend. Circ. Mat. Palermo*, 27:247–271, 1909.

[Calude, 1994] C. S. Calude. *Information and Randomness: An Algorithmic Perspective*. Springer Verlag, 1994.

[Calude, 2000] C. S. Calude. Who is afraid of randomness? Technical Report CDMTCS-143, University of Auckland, New Zealand, 2000.

[Calude, 2005] C. S. Calude. Algorithmic randomness, quantum physics, and incompleteness. In *Proceedings of the Conference on Machines, Computations and Universality(MCU2004)*, volume 3354, pages 1–17. Springer, 2005.

[Chaitin, 1992] G. J. Chaitin. *Algorithmic Information Theory*. Cambridge University Press, 1992.

[Church, 1940] A. Church. On the concept of a random sequence. *Bull. Amer. Math. Soc.*, 46:130–135, 1940.

[Davis, 1980] M. Davis. What Is a Computation? In Lynn Arthur Steen, editor, *Mathematics Today*, pages 241–267. Vintage Books, New York, 1980.

[de Laplace, 18191952] Pierre-Simon de Laplace. *A Philosphical Essay on Probabilities*. Dover, translated from 6th french ed edition, 1819,1952.

[Delahaye, 1993] J.P. Delahaye. Randomness, Unpredictability and Absence of Order: The Identification by the Theory of Recursivity of the Mathematical Notion of Random Sequence. *Philosophy of Probability*, pages 145–167, 1993.

[Downey and Hirschfeldt, 2010] R. G. Downey and D. Hirschfeldt. *Algorithmic Randomness and Complexity*. Springer, 2010.

[Eagle, 2005] A. Eagle. Randomness is unpredictability. *The British Journal for the Philosophy of Science*, 56(4):749–790, 2005.

[Feller, 1968] W. Feller. *An Introduction to Probability Theory and its Applications. Vol. 1.* John Wiley & Sons, New York, 1968.

[Gács, 1974] P. Gács. On the symmetry of algorithmic information. *Soviet Math. Dokl.*, 15:1477–1480, 1974.

[González, 2008] Cristóbal Rojas González. *Randomness and Ergodic Theory: An Algorithmic Point of View*. PhD thesis, École Polytechnique, Paris, France, and Università di Pisa, Italy, 2008.

[Hjorth and Nies, 2007] G. Hjorth and A. Nies. Randomness via effective descriptive set theory. *Journal of the London Mathematical Society*, 75(2):495, 2007.

[Hoyrup, 2008] Mathieu Hoyrup. *Computability, Randomness and Ergodic Theory on Metric Spaces*. PhD thesis, University Paris Diderot, France, and Università di Pisa, Italy, 2008.

[Jauch, 1990] J. M. Jauch. *Are Quanta Real? A Galilean Dialogue*. Indiana University Press, 1990.

[Knuth, 1998] D.E. Knuth. *The Art of Computer Programming, Volume 2: Seminumerical Algorithms, 3rd ed.* Addison-Wesley Reading, Mass, 1998.

[Kolmogorov, 1963] A.N. Kolmogorov. On tables of random numbers. *Sankhyā, Ser. A*, 25:369–376, 1963.

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

[Kolmogorov, 1974] A.N. Kolmogorov. Laws of information conservation (non-growth) and aspects of the foundation of probability theory. *Problems Inform. Transmission*, 10:206–210, 1974.

[Kuipers and Niederreiter, 1974] L. Kuipers and H. Niederreiter. *Uniform distribution of sequences*. John Wiley & Sons, New York, 1974.

[Li and Vitanyi, 2008] M. Li and P. Vitanyi. *An introduction to Kolmogorov complexity and its applications*. Springer, 3rd edition, 2008.

[Longo, 2009] G. Longo. Randomness and Determination, from Physics and Computing towards Biology. In *SOFSEM 2009: Theory and Practice of Computer Science: 35th Conference on Current Trends in Theory and Practice of Computer Science, Špindleruv Mlỳn, Czech Republic, January 24-30, 2009. Proceedings*, page 49. Springer, 2009.

[MacHale, 1993] D. MacHale. *Comic Sections: The book of mathematical jokes, humour, wit, and wisdom*. Boole Press, Dublin, 1993.

[Martin-Löf, 1966] P. Martin-Löf. The definition of random sequences. *Inform. and Control*, 9:602–619, 1966.

[Merkle *et al.*, 2006] W. Merkle, J.S. Miller, A. Nies, J. Reimann, and F. Stephan. Kolmogorov–Loveland randomness and stochasticity. *Annals of pure and applied logic*, 138(1-3):183–210, 2006.

[Miller and Nies, 2006] J.S. Miller and A. Nies. Randomness and computability: open questions. *Bulletin of Symbolic Logic*, pages 390–410, 2006.

[Mlodinow, 2008] L. Mlodinow. *The Drunkard's Walk: How Randomness Rules Our Lives*. Pantheon Books, New York, 2008.

[Moschovakis, 1980] Y. N. Moschovakis. *Descriptive Set Theory*, volume 100 of *Studies in Logic and the Foundations of Mathematics*. North-Holland Publishing Company, Amsterdam, 1980.

[Nies, 2009] A. Nies. *Computability and Randomness*. Oxford University Press, 2009.

[Odifreddi, 1992] P. Odifreddi. *Classical Recursion Theory*, volume 125 of *Studies in Logic and the Foundations of Mathematics*. North-Holland (Elsevier), Amsterdam, 1992.

[Penrose, 1989] R. Penrose. *The Emperor's New Mind*. Oxford University Press, New York, 1989.

[Rogers Jr, 1987] H. Rogers Jr. *Theory of recursive functions and effective computability*. MIT Press, Cambridge, Mass., 1987.

[Shen *et al.*, 20??] A. K. Shen, Uspensky V. A., and Vereshchagin N. K. *Kolmogorov Complexity and Randomness*. To appear, 20??.

[Sierpinski, 1917] W. Sierpinski. Demonstrationelementaire du theoreme de M. Borel sur les nombres absolument normaux et determination effective dun tel nombre. *Bull. Soc. Math. France*, 45:127–132, 1917.

[Sipser, 1997] M. Sipser. *Introduction to the Theory of Computation*. PWS Publishing Company, 1997.

[Solomonoff, 1960] R.J. Solomonoff. A preliminary report on a general theory of inductive inference. Technical Report ZTB-138, Zator Company, Cambridge, Mass., 1960. (November, 1960).

[Solomonoff, 1964] R.J. Solomonoff. A formal theory of inductive inference, part 1 and part 2. *Inform. Contr.*, 7:1–22, and 224–254, 1964.

[Stewart, 2002] I. Stewart. *Does God Play Dice?: The New Mathematics of Chaos*. Blackwell Publishers, 2002.

[Svozil, 1993] K. Svozil. *Randomness & undecidability in physics*. World Scientific, 1993.

[Taleb, 2005] N.N. Taleb. *Fooled by Randomness: The Hidden Role of Chance in Life and in the Markets*. Random House Trade Paperbacks, New York, 2005.

[Tromp, 2009] J. Tromp. Binary Lambda Calculus and Combinatory Logic. `http://homepages.cwi.nl/~tromp/cl/cl.html`, 2009. PostScript file of paper updated on March 13, 2009.

[Turing, 1992] A.M. Turing. A note on normal numbers. *Collected Works of AM Turing, Pure Mathematics, North-Holland, Amsterdam*, pages 117–119, 1992.

[Uspensky, 1983] V.A. Uspensky. *Post's Machine*. Little Mathematics Library. Mir Publishers, Moscow, 1983.

[van der Waerden, 1927] B. L. van der Waerden. Beweis einer baudetschen vermutung. 15:212–216, 1927.

[van Lambalgen, 1987a] M. van Lambalgen. *Random Sequences*. PhD thesis, University of Amsterdam, 1987.

[van Lambalgen, 1987b] M. van Lambalgen. Von Mises' definition of random sequences reconsidered. *Journal of Symbolic Logic*, 4:725–755, 1987.

[van Lambalgen, 1989] M. van Lambalgen. Algorithmic information theory. *Journal of Symbolic Logic*, pages 1389–1400, 1989.

[van Lambalgen, 1990] M. van Lambalgen. The axiomatization of randomness. *Journal of Symbolic Logic*, pages 1143–1167, 1990.

[van Lambalgen, 1996] M. van Lambalgen. Von Mises' axiomatization of randomness reconsidered. In LS Shapley TS Ferguson and JB MacQueen, editors, *Statistics, Probability and Game Theory, papers in honor of David Blackwell*, volume 30 of *IMS Lecture Notes and Monograph series (Hayward, CA)*, 1996.

[Ville, 1939] J. Ville. *Etude Critique du Concept de Collectif*. Gauthier-Villars, Paris, 1939.

[Volchan, 2002] Sergio B. Volchan. What is a random sequence? 109:46–63, 2002.

[Von Mises, 1919] R. Von Mises. Grundlagen der Wahrscheinlichkeitsrechnung. *Math. Zeitschrift*, 5:52–99, 1919.

[Von Mises, 1981] R. Von Mises. *Probability, statistics and truth (1957)*. Dover Publications, New York, 1981.

[V'yugin, 1997] Vladimir V. V'yugin. Effective convergence in probability and an ergodic theorem for individual random sequences. *SIAM Theory of Probability and Its Applications*, 42(1):39–50, 1997.

[V'yugin, 1999] Vladimir V. V'yugin. Ergodic theorems for individual random sequences. *Theoretical Computer Science*, 207(4):343–361, 1999.

[Wald, 1936] A. Wald. Sur la notion de collectif dans le calcul des probabilités. *C. R. Acad. Sci.*, 202:1080–1083, 1936.

[Yurtsever, 2000] U. Yurtsever. Quantum mechanics and algorithmic randomness. *Arxiv preprint: arXiv:quant-ph/9806059v2*, pages 1–8, 2000.

INDEX