Proc. Natl. Acad. Sci. USA
Vol. 93, pp. 2083–2088, March 1996
Mathematics

Randomness and degrees of irregularity

(approximate entropy/computable/maximally random sequences/normal numbers/chaos)

STEVE PINCUS*† AND BURTON H. SINGER‡

*990 Moose Hill Road, Guilford, CT 06437; and ‡Office of Population Research, Princeton University, Princeton, NJ 08544

Contributed by Burton H. Singer, November 22, 1995

ABSTRACT **The fundamental question "Are sequential data random?" arises in myriad contexts, often with severe data length constraints. Furthermore, there is frequently a critical need to delineate nonrandom sequences in terms of closeness to randomness—e.g., to evaluate the efficacy of therapy in medicine. We address both these issues from a computable framework via a quantification of regularity. ApEn (approximate entropy), defining maximal randomness for sequences of arbitrary length, indicating the applicability to sequences as short as $N = 5$ points. An infinite sequence formulation of randomness is introduced that retains the operational (and computable) features of the finite case. In the infinite sequence setting, we indicate how the "foundational" definition of independence in probability theory, and the definition of normality in number theory, reduce to limit theorems without rates of convergence, from which we utilize ApEn to address rates of convergence (of a deficit from maximal randomness), refining the aforementioned concepts in a computationally essential manner. Representative applications among many are indicated to assess (*i*) random number generation output; (*ii*) well-shuffled arrangements; and (*iii*) (the quality of) bootstrap replicates.**

The expressions "picking at random," "chance," and "due to chance" are ambiguous terms. Probability theory, widely believed by nonmathematicians to conclusively fix these notions, retains fundamental deficits here, including inapplicability to finite, especially short sequences and moreover a vivid failure to delineate infinite sequences of clearly different type, highlighted below. The simplest place to discuss the question of randomness is in the context of trying to make the words "Random Sequence" precise. Attempts to do this for infinite sequences have a history via Borel, von Mises, Church, Kolmogorov, Chaitin, Martin-Löf, and others. Their ideas and a rigorous formulation are given by Kolmogorov and Usphenski (KU) (1). KU demonstrate that the following set of criteria, for which there is strong intuitive plausibility and historical consensus, can be made precise: (*i*) typicality (i.e., belonging to any reasonable majority, having no obvious pattern, not very special); (*ii*) "chaotic" (i.e., there is no simple law governing the alternation of its terms); (*iii*) stability of frequencies (i.e., for the full sequence and properly chosen infinite subsequences).

KU adapt their infinite-sequence algorithmic formulation of randomness to the finite case; however, they are careful to point out "the essentially asymptotic nature of all of our considerations" (ref. 1, p. 402). Despite the conceptual elegance of KU's theory, such basic problems as delineating a sufficiently patternless arrangement of agricultural field plots (2), a random assignment of patients to treatments in clinical trials in medicine (3), any notion of random card shuffle (4, 5), and selection of random Latin squares (6, 7) lie outside of the scope of their theory.

Furthermore, there appears to be a critical need to develop a (computable) formulation of "closer to random," to grade the large class of decidedly nonrandom sequences. It appears that many analysts looking for chaos simply require a means of calibrating (short) sequences from this perspective. As another example, the capability to linearly order by randomness may be essential, e.g., in evaluating the efficacy of therapy in medical settings (8). Thirdly, segmenting "junk" DNA (9) into distinct subclasses may bring new insights to genetic function. While classical statistical tests can reject a null hypothesis of randomness based on various types of violations (10–14), they fail to address this issue at all—they provide no metric for proximity to randomness.

The purposes of this paper are (*i*) to define a notion of randomness for sequences of arbitrary length, particularly the short sequences that are not encompassed by the KU theory, that additionally allows one to classify nonrandom sequences as closer to random; and (*ii*) to introduce an infinite sequence formulation of randomness that retains the operational features of the finite case, thereby emphasizing a notion of computable random sequence.

Degrees of Irregularity

We formulate a notion of randomness for finite sequences in terms of degrees of irregularity, which features the "approximate stability of frequencies" criterion from the KU list. An essential tool for this purpose is approximate entropy (ApEn) (15), specified according to

Definition 1: Given a positive integer N and nonnegative integer m, with $m \leq N$, a positive real number r and a sequence of real numbers $u: = (u(1), u(2), \ldots u(N))$, let the distance between two blocks $x(i)$ and $x(j)$, where $x(i) = (u(i), u(i+1), \ldots u(i+m-1))$, be defined by $d(x(i), x(j)) = \max_{k=1,2,\ldots,m}(|u(i+k-1) - u(j+k-1)|)$. Then let $C_i^m(r) =$ (number of $j \leq N - m + 1$ such that $d(x(i), x(j)) \leq r)/(N - m + 1)$.

Remark: The numerator of $C_i^m(r)$ counts, to within resolution r, the number of blocks of consecutive values of length m that are approximately the same as a given block of consecutive values.

Now define

$$\Phi^m(r) = \frac{1}{N-m+1}\sum_{i=1}^{N-m+1} \log C_i^m(r),$$

$$\text{and } \mathrm{ApEn}(m, r, N)(u) = \Phi^m(r) - \Phi^{m+1}(r), m \geq 1,$$

$$\text{with } \mathrm{ApEn}(0, r, N)(u) = -\Phi^1(r).$$

ApEn(m, r, N)(u) measures the logarithmic frequency with which blocks of length m that are close together remain close together for blocks augmented by one position. Thus, small values of ApEn imply strong regularity, or persistence, in a

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: KU, Kolmogorov and Usphenski; ApEn, approximate entropy; OW, Ornstein and Weiss; LIL, law of the iterated logarithm; i.i.d., independent identically distributed.

†To whom reprint requests should be addressed.

sequence u. Alternatively, large values of ApEn imply substantial fluctuation, or irregularity, in u.

We initially restrict attention to binary sequences of 0s and 1s in order to set forth the basic ideas in the simplest possible setting. We set $r < 1$ as our measure of resolution and observe that the values of the distance metric $d(x(i), x(j))$ will be only 0 or 1. Thus, we are monitoring matches in the blocks $x(i)$ and $x(j)$—i.e., whether $|u(i + k - 1) - u(j + k - 1)| = 1$ or 0. With this restriction at hand, we suppress the dependence of ApEn on r and introduce

Definition 2: A binary sequence of length N, $u_*^{(N)}$, is said to be $\{m, N\}$-random—or equivalently, $\{m, N\}$-irregular—if $\text{ApEn}(m, N)\ (u_*^{(N)}) = \max_u \text{ApEn}(m, N)\ (u)$, where the maximum is evaluated over the set of all 2^N binary sequences of length N.

Definition 3: $u_*^{(N)}$ is said to be N-random—or equivalently, N-irregular—if it is $\{m, N\}$-random for $m = 0, 1, 2, \ldots, m_{\text{crit}}(N)$, and the critical values of the nondecreasing in N integer-valued sequence $m_{\text{crit}}(N)$ are defined by the following rule:

$$m_{\text{crit}}(N) = \max(m: 2^{2^m} \leq N).$$

The specification of $m_{\text{crit}}(N)$ is motivated by an application of the methods of Ornstein and Weiss (OW) (16) to the present context in which if $u_N = (u(1), u(2), \ldots u(N))$, $N \geq 1$ is a so-called "typical realization" (ref. 16, p. 914) of a Bernoulli process, then $\lim_{N\to\infty} \text{ApEn}(m_{\text{crit}}(N),N)\ (u_N) = h$ = entropy of the process. In fact, OW show that if $L(m)$ is any sequence that grows faster than exponential—e.g., 2^{2^m} is one instance—then when N is in the range $L(m) \leq N < L(m + 1)$, we calculate $(1/m)$ log(minimum number of blocks of length m needed to cover 1/2 of the sequence $(u(1), u(2), \ldots u(N))$) and observe that as $m \to \infty$, this ratio converges to h.§ The intuitive idea here is that—asymptotically, as $m \to \infty$—scanning blocks of length m is enough to determine the entropy rate of a process from a typical realization. Remarkably, the same scanning criterion allows us to extract maximally irregular sequences—even when N is small. We interpret $m_{\text{crit}}(N)$ as imposing a limit of gradation as a function of sequence length by indicating a maximal order of joint distribution frequencies consistent with a convergent entropy estimate that furthermore avoids a "curse of dimensionality" by the superexponential growth rate of $L(m)$.

Example: Consider the cases $N = 5$ and $N = 6$, where $m_{\text{crit}}(N) = 1$, and we require that $\text{ApEn}(0, N)(u)$ and $\text{ApEn}(1, N)(u)$ be maximal. For $N = 5$, $\max \text{ApEn}(0, 5)(u) \approx 0.673$ is attained for 20 of the $2^5 = 32$ possible binary sequences of length 5. The $\{0, 5\}$-random sequences are those with 3 0s and 2 1s, or 3 1s and 2 0s. Of these 20 sequences, only 4 are $\{1, 5\}$-random: {1,1,0,0,1}, {1,0,0,1,1}, {0,0,1,1,0}, and {0, 1,1,0,0}, with $\max \text{ApEn}(1, 5)\ (u) \approx 0.7133$. For these $\{1, 5\}$-random sequences, the four length-2 blocks $\{0, 1\}$, $\{0, 0\}$, $\{1, 1\}$, and $\{1, 0\}$ each occur once, whereas for all 16 other $\{0, 5\}$ random sequences, at least one of these length-2 blocks is absent. This allows us to characterize the 5-random binary sequences as the equivalence class for subsequences of length 5 in realizations of a partially exchangeable process (17, 18) where approximate stability of frequencies holds. By approximate stability of frequencies, we mean $|1/N$ (no. of js in the sequence) $- 1/2|$ is as small as possible for $j = 0, 1$ and $|\frac{1}{(N-1)}$ (no. of $\{j, k\}$ blocks in the sequence) $- 1/4|$ is as small as possible for $j = 0, 1; k = 0, 1$.

Remark: For partially exchangeable processes, the equivalence classes of subsequences of realizations of length N are those whose members start and end with the same value and have the same number of {0,0}, {0,1}, {1,0}, and {1,1} blocks. However, the frequency of occurrence of the blocks within a sequence need not closely approximate 1/4 as in the case of N-random sequences.

Pursuing the relationship between members of equivalence classes of length N segments of realizations of partially exchangeable processes and 5-random sequences, notice that from the intuitive perspective of randomness, or irregularity, and equivalence class membership, {1,0,1,0,1} should not be considered random (and it is not $\{1, 5\}$-random, missing both {0,0} and {1,1} of length-2 blocks). It is not, however, so immediately evident that $u = \{1,0,1,1,0\}$ should be rejected. However, $\text{ApEn}(1, 5)\ (u) = 0.3667 < \max_u \text{ApEn}(1, 5)(u) \approx 0.7133$; more immediately $\{1, 0\}$ occurs twice while $\{0, 0\}$ does not occur, ruling out membership in an equivalence class with approximately stable frequencies.

In a more algebraic direction, observe that the four 5-random sequences can be derived from the single sequence {1,1,0,0,1}, by reversing order, and then negating the original and reversed versions. This suggests a more general question of identifying a minimal set of N-random sequences or generators from which all others can be derived by using a restricted set of permutations and their negations of the resulting sequences.

For $N = 6$, 20 of the 2^6 binary sequences are $\{0, 6\}$-random, those sequences with 3 0s and 3 1s. Of these 20 sequences, 12 are $\{1, 6\}$-random. They have the property that of the four length-2 blocks $\{0, 1\}$, $\{0, 0\}$, $\{1, 1\}$, and $\{1, 0\}$, three occur once, and the fourth occurs twice. Retaining the idea of equivalence classes of partially exchangeable processes where approximate stability of frequencies holds as a characterization of N-random sequences, we broaden the definition of equivalence class to mean those sequences for which

$$\max_{\{j,k\}} \left| \frac{1}{N-1} (\text{no. of } \{j,k\} \text{ blocks in the sequence}) - 1/4 \right|$$

is as small as possible for the given value of N and $j = 0, 1; k = 0, 1$. We then deduce that {1,0,1,0,1,0} is clearly not $\{1,6\}$-random, and perhaps less obviously, {1,0,1,0,0,1} is not $\{1,6\}$-random (2 occurrences of {1,0} and {0,1}, 1 occurrence of {0,0}, none of {1,1}).

$N = 6$ is the shortest sequence length for which we can construct approximately ergodic sequences, in the sense that there is the best approximation, for a given N, to equality of phase and time averages. To see this, consider the 6×6 array $u_i = (u_{i,1}, u_{i,2}, \ldots, u_{i,6})$, $1 \leq i \leq 6$ specified as follows.

$$\begin{aligned} u_1 &= 1\,1\,0\,0\,1\,0 \\ u_2 &= 0\,1\,1\,0\,0\,1 \\ u_3 &= 1\,0\,1\,1\,0\,0 \\ u_4 &= 1\,0\,0\,1\,1\,0 \\ u_5 &= 0\,1\,0\,0\,1\,1 \\ u_6 &= 0\,0\,1\,1\,0\,1 \end{aligned}$$

Identifying the successive values in a given sequence as a time ordering and the vectors u_i $(1 \leq i \leq 6)$ as points in a phase space, we have

$$1/6 \sum_{L=1}^{6} u_{i,L} = 1/6 \sum_{i=1}^{6} u_{i,L} = 1/2,$$

and

$$1/5[\text{no.}\{j,k\} \text{ blocks in row } i] \approx 1/5[\text{no.}\{j,k\} \text{blocks in column } L]$$

$$\text{for } 1 \leq i \leq 6;\ 1 \leq L \leq 6 \text{ and } j = 0, 1; k = 0, 1.$$

Approximate equality means that both sequence frequencies are as close as possible to 1/4 given $N = 6$. Observe that

§We can improve (i.e., increase) this choice for $m_{\text{crit}}(N)$ and retain consistency with asymptotics to estimate the rate of entropy. OW (ref. 16, Lemma 3, pp. 914–915) require only $\lim_{m\to\infty} L(m)/|A|^m = \infty$, where $|A|$ denotes the cardinality of the state space, rather than for all $A > 0$. Thus, in the binary case one can specify—e.g., $m_{\text{crit}}(N) = \max(m: (2.1)^m \leq N)$. Nonetheless, we retain our specified choice of $m_{\text{crit}}(N)$ to generate a slightly larger class of N-random sequences, for applications with small N (e.g., $N < 50$).

both the rows and columns of the above array are 6-random sequences.

As noted in the introduction, there is an essential need to make precise the notion of "more (or closer to) random" in comparing two typically nonrandom sequences.

Definition 4: A binary sequence of length N, u, is said to be more $\{m, N\}$-random than v if $\mathrm{ApEn}(m, N)(u) > \mathrm{ApEn}(m, N)(v)$; and u is said to be more N-random than v if $\mathrm{ApEn}(m, N)(u) \geq \mathrm{ApEn}(m, N)(v)$ for $m = 0, 1, 2, \ldots, m_{\mathrm{crit}}(N)$, with strict inequality for at least one value of $m \leq m_{\mathrm{crit}}(N)$.

With the above at hand, two general points should be made about N-random sequences. (*i*) As N increases, (number of N-random sequences)$/2^N \to 0$. To see this, it suffices to consider the 0-random sequences—i.e., single elements. For even $N = 2n$, maximally irregular sequences have, as a minimum requirement, precisely n 0s and n 1s. Hence, a crude upper bound on the fraction of such (maximally irregular) sequences is

$$\binom{2n}{n} \Big/ 2^{2n}.$$

This quantity $= (2n)!/((n!)^2 2^{2n})$, which by Stirling's formula

$$\approx \frac{\left(\frac{2n}{e}\right)^{2n} \sqrt{4\pi n}}{\left(\frac{n}{e}\right)^{2n} 2\pi n} \frac{1}{2^{2n}} = \frac{1}{\sqrt{\pi n}}.$$

Thus, the fraction of binary sequences that are maximally irregular is bounded above asymptotically by $O(1/\sqrt{n})$.

Now define $\{\text{excess of 0 over 1}\}_N(u) = \max(0, \text{no. of 0s in } (u) - \text{no. of 1s in } (u))$, and symmetrically for $\{\text{excess of 1 over 0}\}_N(u)$. By a virtually identical argument, the fraction of nearly maximally (small, bounded excess) irregular sequences $\to 0$ as $N \to \infty$. Thus, true maximal irregularity is atypical, increasingly so as sequence length increases, and we must include sequences of arbitrarily large excess (as $N \to \infty$) to ensure the possibility of a majority, one of KU's "conditions." Restated, this establishes an incompatibility of conditions *i* and *iii* in the KU list—i.e., to maintain stability of frequencies as best as possible, we must relinquish the requirement of belonging to any reasonable majority, for finite sequences.

(*ii*) An $N + 1$-random sequence does not necessarily have an N-random sequence as prefix. However, most N-random sequences can be made $N + 1$-random by an appropriate choice of 0 or 1 either preceding or after the N-random sequence. These properties make it particularly important to investigate a minimal set of generators, permutations and negations that can generate the full set of N-random sequences without carrying out (2^N − no. of obviously regular sequences) evaluations of $\mathrm{ApEn}(m, N)$ for $m = 0, 1, \ldots, m_{\mathrm{crit}}(N)$.

Correlatively problematic, historically, with notions of randomness is making precise what we mean by realizations from independent random variables. A useful formulation from the perspective of N-randomness is a quantification of maximal irregularity relative to a given sequence u. To make this precise, we need a conditional form of ApEn.

Definition 5: Let $u = (u(1), u(2), \ldots u(N))$ and $v = (v(1), v(2), \ldots v(N))$ be binary sequences. For given m, let $x(i) = (u(i), u(i+1), \ldots u(i+m-1))$ and $y(j) = (v(j), v(j+1), \ldots v(j+m-1))$. Set $C_i^m(v \,\|\, u) = (\text{no. of } j \leq N - m + 1 \text{ such that } d(x(i), y(j)) \leq r)/(N - m + 1)$, with $r < 1$, where $d(x(i), y(j)) = \max_{k=1,2,\ldots,m}(|u(i+k-1) - v(j+k-1)|)$ and

$$\Phi^m(v \,\|\, u) = \frac{1}{N-m+1} \sum_{i=1}^{N-m+1} \log C_i^m(v \| u).$$

Then define $\mathrm{ApEn}(m, N)(v \,\|\, u) = \Phi^m(v \,\|\, u) - \Phi^{m+1}(v \,\|\, u)$, $m \geq 1$, with

$$\mathrm{ApEn}(0, N)(v \,\|\, u) = -\Phi^1(v \,\|\, u).$$

We say that v^* is $\{m, N\}$-maximally irregular with respect to u if $\mathrm{ApEn}(m, N)(v^* \,\|\, u) = \max_v \mathrm{ApEn}(m, N)(v \,\|\, u)$. Similarly, we say that v^* is N-maximally irregular with respect to u if $\mathrm{ApEn}(m, N)(v^* \,\|\, u) = \max_v \mathrm{ApEn}(m, N)(v \,\|\, u)$ for $m = 0, 1, 2, \ldots, m_{\mathrm{crit}}(N)$. We denote u and v as mutually maximally irregular if each is maximally irregular with respect to the other.

Note that neither u nor v need be $\{m, N\}$-irregular for v to be $\{m, N\}$-maximally irregular with respect to u; thus, this definition applies to comparisons of correlated sequences. Furthermore, *Definition 5* extends to the continuous state space setting, thus formalizing a cross-ApEn notion in this broader context. Such a notion is potentially very important, not simply in assessing the independence of realizations but in comparing sequences from, e.g., two distinct yet intertwined variables in a network (e.g., heart rate and respiratory rate).

Infinite Binary Sequences

The Void: An Acute Need for a Computable Formulation of Randomness. KU (1), Martin-Löf (19), and Chaitin (20) start their theory of infinite random sequences with sets of computable sequences as members of a sample space that contains the random elements. However, in forming the constructive support of the uniform Bernoulli measure, the set that contains the infinite random sequences, identifying a specific realization becomes impossible because of the lack of computability that arises in formation of the intersection of all sets of effective measure 1 (ref. 1, p. 393). The essential point, as Turing (21) showed, is that limits of sequences of computable numbers—i.e., those generated by a finite length computer program—are in general noncomputable.

Martin-Löf and Kolmogorov also show that the set of random sequences—i.e., those in the constructive support of the uniform Bernoulli measure—coincide with the set of algorithmically complex numbers. Although the idea of arbitrarily complex algorithms as the basis of computation of the elements of a random sequence has great intuitive appeal, it also leads to the impossibility of designating and displaying a single random string. This impossibility was clarified in a striking series of papers by Chaitin (e.g., refs. 20 and 22). In his setting, if we let U be an optimal prefix computer and Ω be the probability that U halts on a realization of a random string, the impossibility problem is well-summarized in the quotation: "First, we will never learn more than a handful of those tantalizingly information-packed bits of Ω. Second, even if we had an oracle supplying all bits of Ω, we could not make any practical use of them, since the time to decompress Ω to a solution of the halting problem, X, grows non-recursively" (23).

Thus, it is no surprise that algorithmic probability theory as in KU's formulation, while even leading to a proof that the law of the iterated logarithm (LIL) holds for random Kolmogorov sequences (24), can never become operational in the sense of exhibiting even one such sequence.

To further motivate our formulation below, we consider notions of randomness directly linked to a number-theoretic perspective. Modern measure-theoretic probability (25, 26) together with the plausible contention that normal numbers (27, 28) are good candidates for realizations of infinite random sequences led to the theorem that almost all numbers in the unit interval are normal. But Turing (21) showed that almost all normal numbers cannot be computed by any possible algorithm. Now we can think of the method of continued fractions as a finite algorithm; i.e., the computation of each rational approximation, P_n/Q_n, to any irrational number,

needs finitely many steps. Formally, every member of the continuum has a continued fraction expansion (28, 29). Thus, it may appear that we can compute any normal number algorithmly. However, to compute a continued fraction one must also specify an initial condition, or seed. A consequence of Turing's proof (21) is that for almost all numbers that can be defined, no seed can be found, even in principle.

Operational Formulation. The previous subsection indicates limitations that are at once mathematically and philosophically significant, yet of the utmost practical importance, since procedures to generate arbitrarily long random sequences are necessarily constructive. We retain our focus on operationalization and bring in some definitions. For infinite binary sequences $u = (u(1), u(2), \ldots)$ and $r < 1$, define $u^{(N)} = (u(1), u(2), \ldots, u(N))$, and define $\mathrm{ApEn}(m, N)(u)$: = $\mathrm{ApEn}(m, N)(u^{(N)})$. Then define $\mathrm{ApEn}(m)(u)$: = $\lim_{N\to\infty} \mathrm{ApEn}(m, N)(u^{(N)})$, assuming this limit exists. Now introduce

Definition 6: An infinite binary sequence u is called computationally random, hereinafter denoted as C-random, iff $\mathrm{ApEn}(m)(u) = \log 2$ for all $m \geq 0$.

Some intuitive motivation for this definition is provided by the fact that log2 is a maximal entropy rate for binary sequences, with the desirable interpretation of maximal information conveyed per digit generated. Alternatively, if $\mathrm{ApEn}(m)(u) < \log 2$ for some m, then there exists a block $a_1, a_2, a_3, \ldots a_m, a_{m+1}$ such that $\lim_{N\to\infty}$ conditional frequency $(a_{m+1} \| a_1, a_2, a_3, \ldots a_m) > 1/2$; in other words, the finite prior history block biases the subsequent observation, giving some predictability. Conversely if $\mathrm{ApEn}(m)(u) = \log 2$ for all m, then no such conditional information gain exists for any finite block of a_is. Clearly, these properties are very plausible features of any sequence we purport to call random.

Importantly, observe that for an infinite sequence of binary random variables $\{X_i\}$, $i \geq 1$, with probability $P = 1/2$ each of 0 and 1, an assumption of joint independence as defined by classical probability theory reduces to C-randomness of realizations with probability one—i.e., that $\mathrm{ApEn}(m)(u) = \log 2$ for all m. Similarly, the normality of a binary number again reduces to the condition that $\mathrm{ApEn}(m)(u) = \log 2$ for all $m \geq 0$. Thus, both these fundamental notions are seen as limit statements, without rates of convergence, which we refine below.

It would seem *a priori* plausible that to specify examples of u satisfying *Definition 6*, we would take appropriate limits of sets of maximally irregular initial segments. However, this is out of the question because of

THEOREM 1. *There exists no infinite binary sequence* u *with the property that all initial segments* $\mathrm{u}^{(\mathrm{N})}$ *are maximally irregular.*

Proof: We argue by contradiction. If all initial segments $u^{(N)}$ are maximally irregular, in particular, for all n, $u^{(2n)}$ has n 0s and n 1s, by the maximality of $\mathrm{ApEn}(0, 2n)(u^{(2n)})$. It follows that for all n, each pair of contiguous elements $\{(u(2n-1), u(2n))\}$ must be either $\{(0,1)\}$ or $\{(1,0)\}$. Thus, neither the triple $\{(0,0,0)\}$ nor $\{(1,1,1)\}$ ever occurs for any $\{(u(n), u(n+1), u(n+2))\}$ in u, violating the required limiting frequency of 1/8 of both these triples to satisfy normality.

Turning to normal numbers, to the best of our knowledge, the only known computable binary normal number is 0.11011100101110111100 0 . . . , the binary version of 0.1234567891011. . . . To make progress on any formulation of computable random sequences we need a substantial catalogue of computable normal numbers. To this end, we prove an innocuous-looking theorem that should be quite useful.

THEOREM 2. *Assume we are given a normal binary number* u = (u(1), u(2), . . .), *a binary sequence* v = (v(1), v(2), . . .), *and* f: $\mathrm{Z}^+ \to \mathbf{R}^+$ *such that* $\mathrm{f(N)/N} \to 0$ *as* $\mathrm{N} \to \infty$. *If* {*the number of* $\mathrm{i} \leq \mathrm{N}$ *such that* $\mathrm{u(i)} \neq \mathrm{v(i)}\} \leq \mathrm{f(N)}$ *for all* N, *then* v *is a normal binary number.*

Proof: Let $\mathrm{diff}_m(N)(u, v)$: = the number of $i \leq N - m + 1$ such that the length m block $(u(i), u(i+1), \ldots u(i+m-1)) \neq (v(i), v(i+1), \ldots v(i+m-1))$. Then, $\mathrm{diff}_m(N)(u, v) \leq mf(N)$; since $f(N)/N \to 0$ as $N \to \infty$, the limiting frequency of each length m block of v equals that of u. Since u was assumed normal (each length m block with limiting frequency 2^{-m}), it follows that v is normal.

Deficit from Maximal Irregularity. Next, we indicate how close the first portion of u is to $\{m,N\}$-random, and then N-random, via

Definition 7: For an infinite binary sequence u, define $\mathrm{def}_m[u^{(N)}]$: = $\max_{|v|=N} \mathrm{ApEn}(m,N)(v) - \mathrm{ApEn}(m,N)(u^{(N)})$.

Definition 8: The deficit from maximal equidistribution $\mathrm{De}[u^{(N)}]$: = $\max_{m \leq m\mathrm{crit}(N)} (\mathrm{def}_m[u^{(N)}])$.

We apply $\mathrm{De}[u^{(N)}]$ to delineate binary normal numbers into subclasses, as follows. For normal numbers (sequences) u, $\mathrm{ApEn}(m)(u) = \log 2$ for all m by the equidistribution property, so $\mathrm{def}_m[u^{(N)}] \to 0$ as $N \to \infty$ for fixed m. Observe that there may be normal numbers for which $\mathrm{De}[u^{(N)}]$ does not converge to 0 as $N \to \infty$, since $m_{\mathrm{crit}}(N) \to \infty$ as $N \to \infty$; it would be interesting to construct examples of such numbers. We partition the remaining normal sequences into classes by the rate of convergence of $\limsup_{N\to\infty} \mathrm{De}[u^{(N)}]$ to 0 as a function of N.

We now construct families of extremely slowly convergent normal numbers from a single normal binary number.

THEOREM 3. *Pick a normal binary sequence* u. *For any* g: $\mathrm{Z}^+ \to \mathbf{R}^+$ *such that* (i) $\mathrm{g(N)} \to 0$ *as* $\mathrm{N} \to \infty$; (ii) $\mathrm{N}\sqrt{\mathrm{g(N)}}$ *monotonically increases to* ∞ *as* $\mathrm{N} \to \infty$; *and* (iii) *lim* $(\mathrm{g(N)}/\mathit{def}_0[u^{(\mathrm{N})}]) \to \infty$ *as* $\mathrm{N} \to \infty$, *we can construct a normal binary sequence* v *such that for arbitrarily large* N, *De* $[v^{(\mathrm{N})}] > \mathrm{g(N)}$.

Proof: Define $f(N)$: = $[2N\sqrt{g(N)}]$, where $[x]$: = the greatest integer $\leq x$, and a sequence v recursively, as follows. Set $v(1) = u(1)$; given $v(1), \ldots, v(N-1)$, (*i*) set $v(N) = u(N)$ if $u(N) = 0$; (*ii*) set $v(N) = 0$ if $u(N) = 1$ and $\mathrm{diff}_1(N-1)(u, v) \leq f(N) - 1$, otherwise set $v(N) = 1$ if $u(N) = 1$. Given a sequence u and function f, we denote the constructed sequence v: = $u_{\mathrm{const}}(f)$. Thus, in creating v from u, we are imposing a bias of excess 0s that decreases sufficiently with increasing sequence length so that limiting frequencies are unchanged. By Theorem 2, which applies directly by the monotonicity of f, v is normal. We claim that there exists arbitrarily large N with $\mathrm{def}_0[v^{(N)}] > g(N)$, for which it follows that $\mathrm{De}[v^{(N)}] > g(N)$.

Since $f(N)/N \to 0$ and (no. of 1s in $u^{(N)})/N \to 1/2$, it follows that $\limsup_{N\to\infty} (\mathrm{diff}_1(N)(u,v) - f(N)) = 0$. Now let ε_N: = $\max(\{\text{excess of 0 over 1}\}_N (u^{(N)}), \{\text{excess of 1 over 0}\}_N (u^{(N)}))/2N$; then upon unraveling of definitions, it is immediate that

$$\mathrm{def}_0[u^{(N)}] = [(0.5 + \varepsilon_N)\log(0.5 + \varepsilon_N) + (0.5 - \varepsilon_N)\log(0.5 - \varepsilon_N) - \log 0.5], \quad \textbf{[1]}$$

we denote the right-hand side of Eq. **1** by $\mathrm{logpert}(\varepsilon_N)$. By Taylor's Theorem, $\mathrm{logpert}(\varepsilon) \approx 2\varepsilon^2$ for ε sufficiently small. Since by assumption *iii*, $\lim(\sqrt{g(N)}/\sqrt{\mathrm{def}_0[u^{(N)}]}) \to \infty$ as $N \to \infty$, it follows that $\lim(\sqrt{g(N)}/\varepsilon_N\sqrt{2}) \to \infty$ as $N \to \infty$. In particular, we deduce that $\lim(N\sqrt{g(N)}/\{\text{excess of 1 over 0}\}_N (u^{(N)})) \to \infty$ as $N \to \infty$, hence $\lim(f(N)/\{\text{excess of 1 over 0}\}_N (u^{(N)})) \to \infty$ as $N \to \infty$. Therefore, for all N sufficiently large, $\{\text{excess of 1 over 0}\}_N (u^{(N)}) \leq 0.1f(N)$. Since there exists arbitrarily large N for which $\mathrm{diff}_1(N)(u,v) = f(N)$, we deduce from the construction of v that for such N, $\{\text{excess of 0 over 1}\}_N(v^{(N)}) \geq 0.9f(N) > 1.5\, N\sqrt{g(N)}$. Finally, reversing some arithmetic above, for these arbitrarily large N (by the normality of v),

$$\mathrm{def}_0[v^{(N)}] \approx \frac{1}{2}\left(\frac{\{\text{excess of 0 over1}\}_N(v^{(N)})}{N}\right)^2 > g(N),$$

verifying the claim and completing this constructive proof.

Remarks: (*i*) All members of the parametrized family of functions $g_\beta(N) = N^{-\beta} L(N)$, $0 < \beta < 2$, where $L(N)$ is a given slowly varying function (e.g., log$N$) satisfy conditions *i* and *ii* in

Theorem 3 (similarly for $g_\beta(N) = (\log N)^{-\beta}$). Thus, either the normal number u is already extremely slowly convergent—i.e., condition *iii* is not satisfied for any $g_\beta(N)$ in this family—or we can utilize $g_\beta(N)$ for arbitrarily small β in Theorem 3 to generate sequences of normal numbers with rates of convergence of frequencies of 0s, 1s, {0,0}s, {0,1}s, etc. to their limiting frequencies that can be made extraordinarily slow. These rates of convergence can be tuned to particular applications, providing a reservoir of operationally defined random numbers suited to particular calculations. The clear need for such a reservoir of numbers and clarity about what one means by the word random is exemplified by the recent editorial of Maddox (30).

(*ii*) A substantial literature exists in which two almost sure properties of independent identically distributed (i.i.d.) random variables, the LIL and the central limit theorem, are mandated as requirements for any sequence to be considered random. Recall the precise statement of LIL: Let $X_1, X_2, \ldots$ be i.i.d. random variables with $EX_1 = 0$ and $EX_1^2 = 1$. Let $S_N := X_1 + \ldots + X_N$, and let $\mathrm{LIL}(N) = (2N\mathrm{loglog}N)^{1/2}$. Then almost surely $\lim \sup_{N\to\infty} S_N/\mathrm{LIL}(N) = 1$ and $\lim \inf_{N\to\infty} S_N/\mathrm{LIL}(N) = -1$. Restated for a binary alphabet, LIL implies that lim sup of $\{\text{excess of 0 over 1}\}_N$ is asymptotic to $(2N\mathrm{loglog}N)^{1/2}$. Applying the relationship between the excess function and def_0 indicated above, we see that the LIL mandate requires that

$$\lim \sup_{N\to\infty} \mathrm{def}_0[u^{(N)}] \approx \frac{1}{2}\left(\frac{\sqrt{2N\mathrm{loglog}N}}{N}\right)^2 = (\mathrm{loglog}N)/N,$$

which is simply one (rate of convergence) subclass from the set of all C-random sequences.

We can now construct large classes of normal, C-random numbers violating the LIL, as follows. If even a single binary normal number u exists satisfying LIL, apply Theorem 3 with any $g(N) \to 0$ such that $N\sqrt{g(N)}$ monotonically increases $\to \infty$ and $(Ng(N))/(\mathrm{loglog}N) \to \infty$ as $N \to \infty$—e.g., $g(N) = N^{-1/2}$. If no binary numbers satisfy LIL, all applications of Theorems 2 and 3 constructively produce more such numbers. It remains an open question to construct even a single binary sequence with rate of convergence of $\lim \sup_{N\to\infty} \mathrm{def}_0[u^{(N)}]$ to 0 at least as fast as that specified by the LIL.

(*iii*) Having an explicit reservoir of computable random numbers motivates the importance of developing Rankin's formulation of computable probability spaces (31). Here a probability measure is defined on a Borel Algebra, B of subsets of S, sequences of computable normal numbers. P would be a finitely additive set function on B such that $P(S) = 1$. This kind of probability theory leads to at most countably valued random variables; however, their realizations are computable. Much of the more general noncomputable measure theoretic theory still holds in this domain; however, a full development with exploitation of computable realizations of stochastic processes, lies in the future.

Thus, a further critical deficiency of classical probability theory insofar as addressing randomness now glares out—the "independence" criterion is validated for each of the normal numbers constructed above, lumping all these sequences into a single "Random Sequence" category, irrespective of rates of convergence of $\mathrm{De}(u)$. Since we employ finite initial truncations of such putatively random sequences, an assessment of how far from maximally irregular these initial segments are is evidently crucial.

Finally, note that we hardly utilize the full power of De above. We anticipate that joint distributional considerations realized by the study of def_n for $n \geq 1$ will result in new approaches to further problems.

Applications and Generalizations

(*i*) A primary point above is that we develop a computable framework of randomness entirely apart from the axioms of probability theory. Of course, ApEn can be applied to the classical probability and statistics framework via consideration of almost sure realizations of discrete time processes, with analytic expressions to evaluate ApEn given, e.g., in ref. 15.

(*ii*) The binary development above can be mimicked for application to the k-state alphabet, with the major change that asymptotic ApEn values tend to log$k$ for maximally random sequences. In the discrete state, medium- or large-alphabet setting, one might typically employ ApEn by either directly clumping together states or effectively doing so with a metric (via choice of r) to achieve asymptotic computational economies given by small-state alphabet processes.

(*iii*) Standard linear congruential random number generators (32) are conceptually and algorithmically very simple, and tests to verify the randomness of sequences derived from such generators evaluate statistical correlation. For sequences of length, e.g., $N = 25, 50, 100$, employing $X_{i+1} = (aX_i + b)\mathrm{mod}(2^{31} - 1)$ for a and b as in Knuth's "good" random number generators (32), we can apply ApEn to determine whether these $\{X_i\}$ are (nearly maximally) N-random. We would employ a binary version of this generator, with $W_i = 1$ if $X_i/(2^{31} - 1) > 1/2$, 0 otherwise.

Similarly, we can apply ApEn to determine the quality of "independent replicates," either realized by shuffling or bootstrapping (33), by assessing how many of the replicate series are either N-random or nearly so.

(*iv*) Analysis of what it means to be a well-shuffled deck of cards (4, 5) generally addresses this notion via the study of shuffle (permutations and) groups and, in doing so, necessarily ties the final arrangement to the initial deck ordering. There seems to be a compelling need to evaluate what a well shuffled deck is *ex nihilo*, independent of a starting arrangement. While no doubt there should be further considerations, ApEn can be employed to provide a minimum requirement for a well shuffled deck. To illustrate the application, consider a one-suited 13-card deck, consisting of a permutation of the numbers 1–13, and note the pattern of sign changes from the ith to the i + 1st element of this permutation, denoting up by 1 and down by 0, labeling the resultant binary (length 12) vector by u. For this deck (permutation) to be maximally shuffled, at the very least we require that u be 12-random. Thus, there must be 6 ups and 6 downs and, furthermore, among the 4{up, up}, {down, down}, {up, down}, and {down, up} blocks, 3 each of 3 of the blocks, plus 2 of the 4th block. This extends to any length deck, with suit considerations a distinct issue that also may be addressed via ApEn.

(*v*) A common empirical question is Are data $\{u(i)\}$ atypical or abnormal in some sense? Detecting shifts in irregularity from relatively short sequences—e.g., ≤60 points—has been effectively applied by comparing $\mathrm{ApEn}(m, r, N)(u)$ and $\mathrm{ApEn}(m, r, N)(v)$, for length-$N$ sequences u and v, using small m and coarse r. The range of m and r for which stable estimates of ApEn are achievable defines the limits of resolution of this measure of irregularity for sequences of length N. Discrimination using small m and coarse r is equivalent, in a classical probabilistic setting, to saying that low-dimensional (e.g., 2 or 3) joint distributions of two processes on a coarse partition of the state space are qualitatively different. Any attempt to estimate and test a model of the underlying phenomena involving higher dimensional joint distributions would not be feasible with the given data. This is the case even if strong prior theory with some empirical support suggests that a more complex model is required for a meaningful conceptualization of the phenomena being investigated. The central point is that the question—Is there a shift, or a difference in irregularity?—

does not require a full process specification to obtain an answer.

(*vi*) Greater regularity (lower ApEn) generally corresponds to greater ensemble correlation in phase space diagrams. Such diagrams typically display plots of some system variable $u(i)$ vs. $u(i - K)$, for a fixed "time-lag" K, generally with $K = 1$. A plot of $(u(i), u(i + 1))$ may be interpreted as a graphical display of the support of a 2-dimensional joint steady state measure. Similarly, a plot of $(u(i), u(i + 1), u(i + 2))$ represents the support of a 3-dimensional joint steady state measure. ApEn(1, r) and ApEn(2, r) are functionals of these measures that assess underlying process irregularity.

(*vii*) Despite the fact that nearly all members of continua are noncomputable, some useful insights for empirical analyses can be gleaned by a theoretical analysis applying ApEn as a two-parameter family of statistics in m and r to continuous state, discrete-time stochastic processes. A central question is Given an infinite amount of data, can one say that process A is more regular than process B? In discrete state, the rate of entropy can be applied to answer this, while for continuous state processes, there is no nice answer in general. The flip-flop pair of processes (34) supplies the germane counterexample: in general, comparison of relative process randomness *at a prescribed level of resolution* is the best one can do. That is, process B may appear more random than process A on many choices of partitions, but not necessarily on all partitions of suitably small diameter. The flip-flop pair are two i.i.d. processes A and B with the property that for any integer m and any positive r, there exists $s < r$ such that ApEn(m,s)(A) < ApEn(m,s)(B), and there exists $t < s$ such that ApEn(m,t)(B) < ApEn(m,t)(A). Also note that this pair are i.i.d.—the issue is the continuum of the state space, not any correlation between successive observations.

Fortunately, for many continuous state processes A and B, we can assert more than relative regularity, even though both A and B will typically have infinite K–S entropy. For such pairs of processes, K–S denoted a completely consistent pair (34), whenever ApEn(m,r)(A) < ApEn(m, r)(B) for any specific choice of m and r, then it follows that ApEn(n,s)(A) < ApEn(n,s)(B) for all choices of n and s. For completely consistent pairs we can assert that process B is more irregular than process A, without needing to indicate m and r. Additionally, both theoretically and on observed data, we often observe a relative consistency of ApEn over a statistically valid range of (m,r) pairs, similar to that given by completely consistent pairs—whenever the statistical estimate ApEn(m,r,N)(A) < ApEn(m,r,N)(B) for an (m,r) pair, then ApEn(n,s,N)(A) < ApEn(n,s,N)(B) for all (n,s) pairs in the range. We indicate elsewhere (35) conjectures to ensure that A and B are either a completely consistent pair or are relatively consistent over a specified range in terms of the autocorrelation functions of A and B.

Assuming either complete or relative consistency as in the previous paragraph, we suggest the program to choose r relatively coarse in the ApEn(m,r) definition, approximating the discrete state, small alphabet setting, whereupon application with $m = 1$ or (typically and preferably) $m = 2$ is often statistically useful in discrimination (standard deviation of ApEn($m = 2$, $r = 20\%$ process SD, N) ≤ 0.06 for a range of weakly dependent processes, for $60 \leq N \leq 1000$; see ref. 35). Clear discrimination has been established via this protocol both for theoretical processes (15, 36) and in a number of clinical medical applications (8, 37–42), in which controls were compared to pathophysiologic subject populations for N as small as 60 data points, with relative consistency over a range of m and r explicitly shown in refs. 39–41.

1. Kolmogorov, A. N. & Uspenskii, V. A. (1987) *Theory Probab. Its Appl.* **32,** 389–412.
2. Bailey, R. A. (1983) *Biometrika* **70,** 183–198.
3. Pocock, S. J. (1979) *Biometrics* **35,** 183–197.
4. Aldous, D. & Diaconis, P. (1986) *Am. Math. Monthly* **93,** 333–348.
5. Diaconis, P., Graham, R. L. & Kantor, W. M. (1983) *Adv. Appl. Math.* **4,** 175–196.
6. Fisher, R. A. (1935) *The Design of Experiments* (Oliver & Boyd, Edinburgh).
7. Savage, L. J. (1962) *The Foundations of Statistical Inference* (Methuen, London), p. 88.
8. Hartman, M. L., Pincus, S. M., Johnson, M. L., Matthews, D. H., Faunt, L. M., Vance, M. L., Thorner, M. O. & Veldhuis, J. D. (1994) *J. Clin. Invest.* **94,** 1277–1288.
9. Britten, R. J. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 5992–5996.
10. Wald, A. & Wolfowitz, J. (1940) *Ann. Math. Stat.* **11,** 147–162.
11. Olmstead, P. S. (1946) *Ann. Math. Stat.* **17,** 24–33.
12. von Neumann, J. (1941) *Ann. Math. Stat.* **12,** 367–395.
13. Bartholomew, D. J. (1954) *Biometrika* **41,** 556–558.
14. Wallis, W. A. & Moore, G. H. (1941) *J. Am. Stat. Assoc.* **36,** 401–409.
15. Pincus, S. M. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 2297–2301.
16. Ornstein, D. S. & Weiss, B. (1990) *Ann. Prob.* **18,** 905–930.
17. Diaconis, P. & Freedman, D. (1984) *Proc. Ind. Stat. Inst. Conf.* 205–236.
18. Frydman, H. & Singer, B. (1979) *Math. Proc. Cambridge Philos. Soc.* **86,** 339–344.
19. Martin-Löf, P. (1966) *Inform. Control* **9,** 602–619.
20. Chaitin, G. J. (1966) *J. ACM* **13,** 547–569.
21. Turing, A. (1936) *Proc. London Math Soc.* **42,** 230–265.
22. Chaitin, G. J. (1975) *J. ACM* **22,** 329–340.
23. Cover, T. M., Gacs, P. & Gray, R. (1989) *Ann. Prob.* **17,** 863.
24. Vovk, G. G. (1987) *Theory Probab. Its Appl.* **32,** 413–425.
25. Kolmogorov, A. N. (1933) *Grundbegriffe der Wahrscheinlichkeitsrechnung* (Springer, Berlin).
26. Doob, J. L. (1953) *Stochastic Processes* (Wiley, New York).
27. Hardy, G. H. & Wright, E. M. (1983) *An Introduction to the Theory of Numbers* (Clarendon Press, Oxford, U.K.), 5th Ed., pp. 125–128.
28. Niven, I. (1956) *Irrational Numbers*, Carus Monograph 11 (Wiley, New York).
29. Stark, H. M. (1994) *An Introduction to Number Theory* (MIT Press, Cambridge, MA), 8th Paperback Ed.
30. Maddox, J. (1994) *Nature (London)* **372,** 403.
31. Rankin, B. (1960) *Acta Mathematica* **103,** 89–122.
32. Knuth, D. E. (1981) *Seminumerical Algorithms: The Art of Computer Programming* (Addison-Wesley, Reading, MA), 2nd Ed., Vol. 2, Chap. 3.
33. Efron, B. (1982) *The Jackknife, the Bootstrap, and Other Resampling Plans* (SIAM, Philadelphia).
34. Pincus, S. M. & Huang, W.-M. (1992) *Commun. Statist.-Theory Methods* **21,** 3061–3077.
35. Pincus, S. M. & Goldberger, A. L. (1994) *Am. J. Physiol.* **266,** H1643–H1656.
36. Pincus, S. M. & Keefe, D. L. (1992) *Am. J. Physiol.* **262,** E741–E754.
37. Fleisher, L. A., Pincus, S. M. & Rosenbaum, S. H. (1993) *Anesthesiology* **78,** 683–692.
38. Kaplan, D. T., Furman, M. I., Pincus, S. M., Ryan, S. M., Lipsitz, L. A. & Goldberger, A. L. (1991) *Biophys. J.* **59,** 945–949.
39. Pincus, S. M., Gevers, E., Robinson, I. C. A. F., van den Berg, G., Roelfsema, F., Hartman, M. L. & Veldhuis, J. D. (1996) *Am. J. Physiol.* **270,** E107–E115.
40. Pincus, S. M., Cummins, T. R. & Haddad, G. G. (1993) *Am. J. Physiol.* **264,** R638–R646.
41. Pincus, S. M., Gladstone, I. M. & Ehrenkranz, R. A. (1991) *J. Clin. Monit.* **7,** 335–345.
42. Pincus, S. M. & Viscarello, R. R. (1992) *Obstet. Gynecol.* **79,** 249–255.