Proc. Natl. Acad. Sci. USA
Vol. 94, pp. 3513–3518, April 1997
Mathematics

Not all (possibly) "random" sequences are created equal

(approximate entropy/maximally irregular sequences/normal numbers/deficit from equidistribution)

STEVE PINCUS*[†] AND RUDOLF E. KALMAN[‡§]

*990 Moose Hill Road, Guilford, CT 06437; [‡]Mathematical System Theory, Swiss Federal Institute of Technology, CH-8092 Zurich, Switzerland; and [§]Istituto "Asinius Pollio"

Contributed by Rudolf E. Kalman, February 10, 1997

ABSTRACT **The need to assess the randomness of a single sequence, especially a finite sequence, is ubiquitous, yet is unaddressed by axiomatic probability theory. Here, we assess randomness via approximate entropy (ApEn), a computable measure of sequential irregularity, applicable to single sequences of both (even very short) finite and infinite length. We indicate the novelty and facility of the multidimensional viewpoint taken by ApEn, in contrast to classical measures. Furthermore and notably, for finite length, finite state sequences, one can identify maximally irregular sequences, and then apply ApEn to quantify the extent to which given sequences differ from maximal irregularity, via a set of deficit (def_m) functions. The utility of these def_m functions which we show allows one to considerably refine the notions of probabilistic independence and normality, is featured in several studies, including (*i*) digits of e, π, $\sqrt{2}$, and $\sqrt{3}$, both in base 2 and in base 10, and (*ii*) sequences given by fractional parts of multiples of irrationals. We prove companion analytic results, which also feature in a discussion of the role and validity of the almost sure properties from axiomatic probability theory insofar as they apply to specified sequences and sets of sequences (in the physical world). We conclude by relating the present results and perspective to both previous and subsequent studies.**

Suppose one were asked, "Are the digits in the decimal expansion of $\sqrt{2}$ random?" We consider such a question problematic, as discussed in refs. 1–4, and even epistemologically ill-posed, since randomness as addressed by the study of axiomatic probability theory (5, 6) is concerned with ensemble process behavior, rather than the assessment of a specific sequence. In principle, we can instead ask "Is $\sqrt{2}$ a normal number?," since normality is a well-defined sequence notion. Indeed, normality is an expected characteristic of a real number, from a measure–theoretic perspective, since Borel showed that almost all numbers are normal (7). However, Geiringer (ref. 8, p. 311) states the issue poignantly: "The fact that a set of nonnormal numbers is of measure zero does not help in any way in the extremely difficult problem of deciding whether a given number is normal or not."

Moreover, and most importantly, in applications we use finite segments of putatively random sequences, and hence we require computable techniques to assess the "randomness" of such segments, to which neither axiomatic probability theory or normality apply.

The purpose of this paper is to demonstrate, in several distinct settings, the utility of a recently introduced notion of sequential irregularity, approximate entropy (ApEn) (4, 9), with which we can evaluate, e.g., the extent of irregularity of decimal digits of $\sqrt{2}$. As indicated in ref. 4, ApEn addresses and actually refines both the classical probabilistic notion of randomness, and normality, from a fundamentally different vantage point than either of these notions. Specifically, the development of ApEn has the following properties.

(*i*) It is combinatorial, rather than oriented toward almost sure laws, which as discussed below, fail in a number of settings for specified sequences.

(*ii*) It applies to single sequences of both (even very short) finite and infinite length.

(*iii*) It is explicitly computable, in counterpoint to the developments of algorithmic complexity (10–13), and axiomatic probability theory.

(*iv*) In particular, in assigning an explicit measure of irregularity to a sequence u: = $(u(1), u(2), \ldots u(N))$ via ApEn, it avoids the needs (a) to guess as to an underlying set of rules or process used to generate the sequence, and (b) to identify and evaluate the remainder of the sequence, i.e., $\{u(m), m > N\}$. The focus is to evaluate the sequence "at hand." For instance, a sequence of length $N = 100$ could equally well represent either an algorithmically simple block of 100 contiguous digits of π, or an algorithmically complicated output from a 99th degree polynomial, and the need for describing quantitative characteristics of this 100 point sequence exists apart from the disclosure of which of these two (or alternative) means were used to generate the sequence.

(*v*) ApEn(m, .) provides a family of functions that for $m \geq 1$ assesses multidimensional dynamics of contiguous blocks (of run length $m + 1$).

(*vi*) It allows one to identify and quantify maximally irregular finite sequences (which we prove below coincide with maximally equidistributed sequences), for sequences with a finite state space (e.g., binary and decimal digits). From this useful capability,

(*vii*) It allows one to quantify the extent to which nonrandom sequences differ from maximal irregularity, i.e., to provide a formulation of "closer to random," via a set of deficit (def_m) functions.

(*viii*) Both process independence in classical probability theory and normality reduce to a binary, YES/NO determination of whether all of these def_m functions converge to 0, implying an asymptotic convergence to frequency equidistribution.

(*ix*) From *viii*, all study of the def_m functions, beyond answering the YES/NO question of convergence to 0, explicitly characterizes the asymptotic behavior of sequential variation about (possible) maximal equidistribution, and thus allows one to considerably refine the notion of limiting equidistribution, or normality.

Maximal Irregularity

We recall several definitions from ref. 4. Note that we separately develop quantifications of irregularity for both finite sequences and for infinite sequences via approximate entropy, ApEn.

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 $x(i) = (u(i), u(i+1), \ldots u(i+m-1))$,

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.

0027-8424/97/943513-6$2.00/0
PNAS is available online at **http://www.pnas.org**.

Abbreviations: ApEn, approximate entropy; LIL, law of iterated logarithm; CLT, central limit theorem.

[†]To whom reprint requests should be addressed.

be defined by $d(x(i), x(j)) = \max_{p=1,2,\ldots,m} (|u(i+p-1) - u(j+p-1)|)$. Then let $C_i^m(r) =$ (number of $j \le N-m+1$ such that $d(x(i), x(j)) \le r)/(N-m+1)$. Now define

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

and

$$\mathrm{ApEn}(m, r, N)(u) = \Phi^m(r) - \Phi^{m+1}(r), m \ge 1;$$

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

$\mathrm{ApEn}(m, r, N)(u)$ measures the logarithmic frequency with which blocks of length m that are close together remain close together for blocks augmented by one position, with larger values of ApEn implying greater irregularity in u. Alternatively (9, 14),

$$\mathrm{ApEn}(m, r, N)(u) \approx \text{average over } i \text{ of log [conditional frequency that } |u(j+m) - u(i+m)| \le r, \text{given that } |u(j+p) - u(i+p)| \le r \text{ for } p = 0,1,\ldots, m-1], \quad [1]$$

with equality (for fixed m and r) in the limit as $N \to \infty$.

Herein, we consider sequences of base k integers, and as in ref. 4, set $r < 1$ as our measure of resolution. For this choice of r, we can suppress the dependence of ApEn on r and make

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

Definition 3: $u_*^{(N)}$ is said to be *N-irregular* (*N-random*) if it is $\{m, N\}$-irregular for $m = 0, 1, 2, \ldots, m_{crit}(N)$, with $m_{crit}(N)$ defined by: $m_{crit}(N) = \max(m: k^{2^m} \le N)$.

The specification of $m_{crit}(N)$ is discussed in ref. 4. Next, the following gives a useful equivalence of maximally irregular ApEn sequences, expressing that approximate stability of frequencies alternatively characterizes N-random sequences.

THEOREM 1. *A sequence u is N-*random *if and only if for each $1 \le m \le m_{crit}(N) + 1$, the expression*

$$max_{\{\mathrm{v1, v2, \ldots, vm}\}} \left| \frac{1}{N-m+1} \textit{(number of } \{\mathrm{v1, v2, \ldots, vm}\} \textit{ blocks in the sequence } \mathrm{u}) - 1/k^m \right| \quad [2]$$

is a minimum (among length N *sequences), where the max is evaluated over all blocks* $\{\mathrm{v1, v2, \ldots, vm}\}$ *where* $\mathrm{vi} \in \{0, 1, \ldots, \mathrm{k}\text{-}1\}$ *for all* $1 \le \mathrm{i} \le \mathrm{m}$.

Proof: We observe that $\mathrm{ApEn}(0) = -\Phi^1(r)$ must be maximized (among all length N sequences), then recursively that $\mathrm{ApEn}(m) = \Phi^m(r) - \Phi^{m+1}(r)$ must be maximized, hence $-\Phi^{m+1}(r)$ must be maximized for each $m \le m_{crit}(\mathrm{N})$. The proof now follows at once, upon recognition that

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

is the (discrete) entropy of the m-block empirical frequency distribution, maximized when most equidistributed on m-blocks, i.e., when Eq. **2** is satisfied.

Thus, maximal ApEn agrees with intuition for maximally equidistributed sequences, while allowing us to grade the remaining sequences in terms of proximity to maximality. From Theorem 1 and Eq. **1** it follows readily that for the k-state alphabet, asymptotic ApEn values converge to log k for maximally random sequences.

Remark: One can produce sets of maximally equidistributed (length $p - 1$) sequences via sequential digits formed by the base k expansion of q/p, for any integer $0 < q < p$, when p is a k-ergodic prime, as discussed in ref. 3. Recall the formulation of k-ergodic primes. By Fermat's little theorem (15), for p prime and not a divisor of k, it follows that $k^{p-1} = 1$ (*mod* p). Let d be the order of k (mod p), i.e., the smallest positive integer for which $k^d = 1$ (mod p). By Theorem 88, ref. 15, d is a divisor of $p - 1$. If as a special case, $d = p - 1$, we denote p as a k-ergodic prime. The formulation of k-ergodic primes thus leads to a useful set of finite "most random" sequences. Nonetheless, the aforementioned procedure only applies to a relatively sparse collection of sequence lengths—it is not even known if there are infinitely many k-ergodic primes for any given integer $k > 1$. As well, for general sequence lengths, an important open problem is to determine efficient procedures to generate all maximally irregular sequences.

We next recall the technology to quantify proximity of a finite sequence to maximal irregularity.

Definition 4: For a length N sequence $u^{(N)}$, define $\mathrm{def}_m[u^{(N)}]: = \max_{|v|=N} \mathrm{ApEn}(m, N)(v) - \mathrm{ApEn}(m, N)(u^{(\mathrm{N})})$.

Finally, for infinite sequences $u = (u(1), u(2), \ldots)$, $u(i) \in \{0, 1, \ldots, k-1\}$ for all i, and $r < 1$, define $u^{(N)} = (u(1), u(2), \ldots, u(N))$, and define $\mathrm{ApEn}(m, N)(u): = \mathrm{ApEn}(m, N)(u^{(\mathrm{N})})$. Then define $\mathrm{ApEn}(m)(u): = \lim_{N\to\infty} \mathrm{ApEn}(m, N)(u^{(N)})$, assuming this limit exists. Then

Definition 5: An infinite sequence u is called *C-random* if and only if $\mathrm{ApEn}(m)(u) = \log k$ for all $m \ge 0$.

Notably, for an infinite sequence of random variables $\{X_i\}$, $i \ge 1$, with "probability" $p = 1/k$ each of $0, 1, \ldots, k-1$, an assumption of joint independence as defined by classical probability theory reduces to C-randomness of realizations with probability one. Similarly, the normality of a number reduces to the condition that $\mathrm{ApEn}(m)(u) = \log k$, i.e., $\mathrm{def}_m[u^{(N)}] \to 0$ as $\mathrm{N} \to \infty$ for all $m \ge 0$. Thus, both independence and normality are limit statements, *without rates of convergence*, which further study via the def_m functions refine.

Chaitin Example

The following provides insight into the potential utility of assessing irregularity via the analysis of blocks of contiguous points. Chaitin (16) motivates the need for the development of algorithmic complexity (10–13) by contrasting two binary sequences of length $N = 20$, (A) denoted "with an obvious pattern," (B) "that seems to be random":

(A) 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1
(B) 0 1 1 0 1 1 0 0 1 1 0 1 1 1 1 0 0 0 1 0

He comments that in considering A and B, "each represents an event with probability of 2^{-20}"; he then notes that "The conclusion is singularly unhelpful in distinguishing the random from the orderly." From the present perspective, if we think in terms of aggregating 2-blocks, i.e., if we take a 2-dimensional view, ApEn(1) provides an immediate, computable difference: $\mathrm{ApEn}(1, 20) = 0$ for A, whereas $\mathrm{ApEn}(1, 20) = 0.6774$ for B. In particular, the $\mathrm{ApEn}(1) = 0$ calculation for sequence A reflects the observation that there are no length 2-blocks $\{0, 0\}$ or $\{1, 1\}$ anywhere in A. [As an aside, Chaitin did rather well insofar as selecting a reasonably irregular sequence (B), recalling that max $\mathrm{ApEn}(1, 20) \approx \log 2 = 0.693$.]

Thus, ApEn allows a direct, computable alternative to the severely noncomputational approach that algorithmic complexity provides, insofar as identifying random sequences. Furthermore, short data lengths readily sufficed to distinguish A from B above.

Digits of Irrationals

We next study $\mathrm{ApEn}(m, N)$ and def_m for m = 0, 1, and 2 for (relatively) large values of N, for binary and decimal sequences (expansions) of e, π, $\sqrt{2}$, and $\sqrt{3}$. Each of these numbers have been hypothesized to be normal; thus, we anticipate approximate equidistribution for large N. In base 2, we evaluated sequences of length $N \le 300{,}000$, produced from *Mathematica*, while in base 10, we evaluated sequences of length $N \le 1{,}000{,}000$, produced from project *Gutenberg*. Figs. 1 and 2 display def_m as a function of N, Fig. 1 for base 2, Fig. 2 for base 10. To reiterate, the def_m functions here

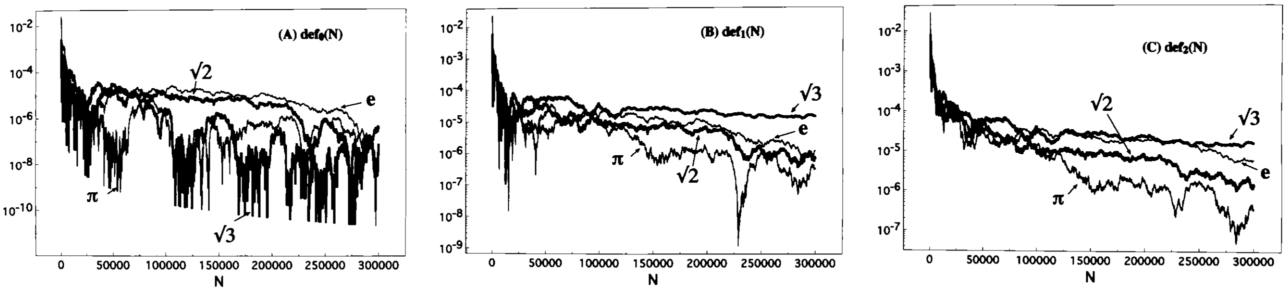

FIG. 1. Deficit from maximal irregularity for base 2 sequence expansions of e, π, $\sqrt{2}$, and $\sqrt{3}$: (*A*) def$_0(N)$, (*B*) def$_1(N)$, (*C*) def$_2(N)$.

for $m = 0$, 1, and 2 quantify the divergence from maximal equidistribution of singletons, pairs, and triples, respectively.[¶]

A number of points are evident from these figures. First, in base 2, differences among e, π, $\sqrt{2}$, and $\sqrt{3}$ are considerable, especially for triples (3-blocks, $m = 2$). Note, e.g., $\sqrt{3}$ is much less irregular than π, for larger N, for both $m = 1$ and $m = 2$, as indicated by the def$_1(N)$ and def$_2(N)$ functions—this difference is often nearly two orders of magnitude.

From another, albeit coarser perspective, for $N = 280{,}000$ for π, the most frequently occurring 3-block of contiguous points is $\{0, 0, 0\}$, with 35,035 occurrences, whereas the least frequently occurring 3-block of contiguous points is $\{1, 1, 1\}$, with 34,944 occurrences—a difference of 91. In comparison, for $N = 280{,}000$, for $\sqrt{3}$, the most frequently occurring 3-block of contiguous points is $\{0, 0, 0\}$, with 35,374 occurrences, whereas the least frequently occurring 3-block of contiguous points is $\{0, 1, 0\}$, with 34,615 occurrences—a difference of 759. Thus $\sqrt{3}$ is considerably further from maximal equidistribution than is π, for an extended range of N. As well, one can recast such calculations to establish greater conditional frequency "bias" for $\sqrt{3}$ than for π, based on pairs and triples, for $100{,}000 \leq N \leq 300{,}000$.

We emphasize that we do not have to validate the normality of e, π, $\sqrt{2}$, and $\sqrt{3}$ to derive meaningful utility from this analysis—def$_m(N)$ and ApEn(m, N) are well-defined functions associated with these numbers (sequences).

Interestingly, as seen in Fig. 2, base 10 differences among these four irrationals are much less pronounced, especially in dimensions 2 and 3; thus, base 10 and base 2 sequence properties are "incommensurate," insofar as persistence (across bases) of gradation by irregularity. Furthermore, in reconsidering Fig. 1, especially *B* and *C*, there is no separation along the lines of algebraic numbers versus transcendental numbers as one might have hypothesized. Namely, whereas π is consistently more irregular than $\sqrt{3}$ in the range $100{,}000 \leq N \leq 300{,}000$, $\sqrt{2}$ is intermediate, with both def$_1$ and def$_2$ for $\sqrt{2}$ between corresponding function values for π and e for the vast preponderance of this range. We must infer that the rapidity of rational approximations in the classical number theoretic sense does not appear to directly manifest itself in the degree to which sequential digits in a given base are irregular.

Additionally, distinct representational forms of a number can produce sequences of completely different character. For example, consider the continued fraction expansion representations $\sqrt{2} = [1, 2, 2, 2, \ldots]$; $\sqrt{3} = [1, 1, 2, 1, 2, 1, 2, \ldots]$; and $\frac{1+\sqrt{5}}{2} = [1, 1, 1, 1, \ldots]$. (We denote the *continued fraction* $a_0 + \frac{1}{a_1 +}\,\frac{1}{a_2 +}\cdots\frac{1}{a_n}$ *by* $[a_0, a_1, \ldots, a_n]$, calling $a_0, a_1, \ldots, a_n$ the partial quotients of the continued fraction.) Ignoring the first digit of each of these three sequences, and applying ApEn to the remaining terms (i.e., to the sequence $\{a_1, \ldots, a_n, \ldots\}$ of the partial quotients), we conclude that the continued fraction expansions for each of these three quadratic surds are quite regular, with ApEn $(m) = 0$ for all $m \geq 0$ for $\sqrt{2}$ and $\frac{1+\sqrt{5}}{2}$, and with ApEn $(m) = 0$ for all $m \geq 1$ for $\sqrt{3}$. So the irregularity of one representation of a number says little about the irregularity of the number in another representation. The point is to evaluate the sequence at hand, rather than the simplicity of a (typically unknown) underlying generation technique.

The overall message in the above, that cannot be overstated, is that we prefer to assess the randomness of a sequence, either finite or infinite, by the behavior of a countable sequence of computable functions def$_m$. These deficit functions provide much richer detail than does a simple YES or NO to the question "Do all these functions def$_m$ necessarily converge to 0?," which is all that normality reduces to, even ignoring the virtually nil set of techniques available to establish (possible) normality.

Multiples of Irrationals

One-Dimensional Deficit. We next consider sequences given by fractional parts of multiples of irrationals. For θ real, let $\omega(\theta) = \{u(1), u(2), \ldots\}$, with $u(n) := \{n\theta\}$, where $\{\ \}$ denotes fractional part, i.e., $n\theta$ mod 1. Such sequences have received considerable study since the beginning of the twentieth century. A famous theorem concerning $\omega(\theta)$, that some take as an alternative characterization of irrational numbers, is the following, discussed and reconsidered in refs. 1 and 2.

GLEICHVERTEILUNGSSATZ Weyl (17). *Let θ be a real number and consider the family S_θ of points on the unit interval given by*

[¶] As a mechanical, yet notable aside, ApEn(m, N) and def$_m(N)$ calculations were made via a linear-time (in N) algorithm, which consumed about 2 min for 1,000,000 points on a Macintosh Power PC. The discreteness of the state space affords the possibility of such linear-time calculations, in contrast with inherently quadratic-time (in N) ApEn algorithms for continuous state space.

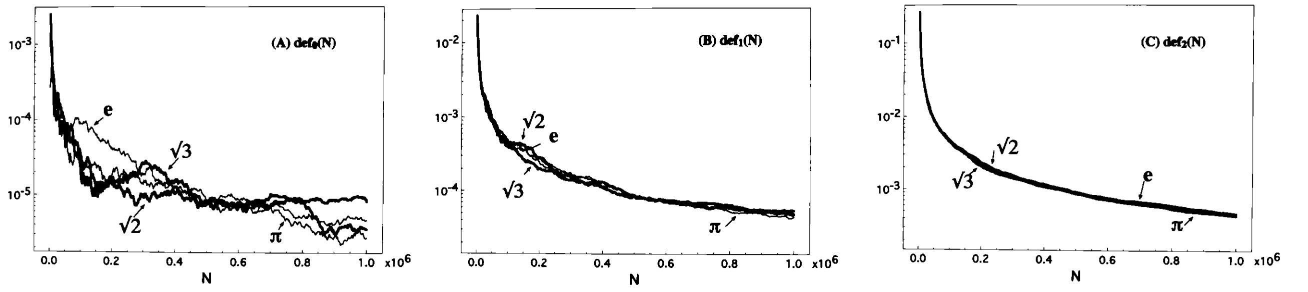

FIG. 2. Deficit from maximal irregularity for base 10 sequence expansions of e, π, $\sqrt{2}$, and $\sqrt{3}$: (*A*) def$_0(N)$, (*B*) def$_1(N)$, (*C*) def$_2(N)$.

S_θ: = {nθ *mod* 1: n = 1, 2, . . . , N}. *The points in* S_θ *are equidistributed (uniformly distributed) on* [0, 1] *in the limit* N → ∞ *if and only if* θ *is irrational*.

Of course, sequences $\omega(\theta)$ are inadequate candidates for "random" output, since contiguous points $u(i)$ in the sequence differ by θ, manifested in 2-dimensional correlation. In the next subsection, we apply ApEn(1) as a direct means to reject randomness of $\omega(\theta)$. Nonetheless, Weyl's Theorem is useful in response to a fundamental need in statistics: how to generate a collection of points uniformly distributed on [0, 1].

Now Weyl's Theorem says nothing about the rate of convergence of the points in S_θ to uniformity. We consider one aspect of this convergence, by study of the binary sequence $B(\theta) = \{\beta(1), \beta(2), \ldots\}$ derived from $\omega(\theta)$ by the following rule: $\beta(i) = 0$ if $u(i) < 1/2$, $\beta(i) = 1$ if $u(i) \geq 1/2$, for $\theta = \sqrt{2}$. From Weyl's Theorem, $\lim_{N\to\infty}$ ApEn$(0, N)$ $\{B(\theta)\} = \log 2$, i.e., $\lim_{N\to\infty}$ def$_0(N)\{B(\theta)\} = 0$, since in the limit, $\beta(i) = 1$ with asymptotic frequency 1/2. But consider Fig. 3, which compares def$_0(N)\{B(\sqrt{2})\}$ to def$_0$ for the binary digits e and $\sqrt{2}$. For $\{B(\sqrt{2})\}$, the def$_0$ function is generally several orders of magnitude smaller than for the binary digit expansions of e and $\sqrt{2}$. Alternatively, for $N = 150{,}000$, $\{B(\sqrt{2})\}$, there are 75,002 occurrences of {0}, and 74,998 occurrences of {1}, a difference of 4, whereas among the first 150,000 binary digits of e, there are 74,618 occurrences of {0}, and 75,382 occurrences of {1}, a difference of 764. Thus, def$_0(N)$ delineates $\{B(\sqrt{2})\}$ as decidedly more 1-dimensionally equidistributed than the sequences of binary digits of e and $\sqrt{2}$ for nearly the entire range of $N \leq 300{,}000$. Below, we supply theory to guarantee the very rapid convergence of def$_0(N)$ to 0 for$\{B(\sqrt{2})\}$ and related sequences, which raises resultant fundamental issues.

Correlation in Two Dimensions. Here we reject randomness of $\omega(\sqrt{2})$ by considering the associated binary sequence $B(\sqrt{2})$ specified above. If $\omega(\sqrt{2})$ were random, $B(\sqrt{2})$ would necessarily be C-random, with ApEn(1)$\{B(\sqrt{2})\} = \log 2 \approx 0.693$, and with the limiting frequencies of the four 2-blocks of contiguous observations (0, 1), (1, 0), (1, 1), and (0, 0) each = 1/4. Now denote the limiting frequencies of {0} and {1} in $B(\sqrt{2})$ by f_0 and f_1, and the limiting frequencies of the 2-blocks {0, 0}, {0, 1}, {1, 0}, and {1, 1} by $f_{0,\,0}$, $f_{0,\,1}$, $f_{1,\,0}$, and $f_{1,\,1}$, respectively. Recalling the notation of *Definition* 1, it follows that $\lim_{N\to\infty}$ ApEn$(0, N)$ $\{B(\sqrt{2})\} = \lim_{N\to\infty} -\Phi^1 = -\{f_0 \log f_0 + f_1 \log f_1\}$, and $\lim_{N\to\infty}$ ApEn$(1, N)$ $\{B(\sqrt{2})\} = \lim_{N\to\infty} \Phi^1 - \Phi^2 = \{f_0 \log f_0 + f_1 \log f_1\} - \{f_{0,\,0} \log f_{0,\,0} + f_{0,\,1} \log f_{0,\,1} + f_{1,\,0} \log f_{1,\,0} + f_{1,\,1} \log f_{1,\,1}\}$. As indicated above, $f_0 = f_1 = 1/2$, hence ApEn(0)$\{B(\sqrt{2})\} = \log 2$. To calculate $f_{0,\,0}$, we note that $\beta(n) = 0$ and $\beta(n + 1) = 0$ if and only if $u(n) = \{n\sqrt{2}\}$ satisfies $0 \leq u(n) < 1.5 - \sqrt{2}$. We immediately deduce that $f_{0,\,0} = 1.5 - \sqrt{2} \approx 0.086$, by the uniformity of the limiting distribution of u on [0, 1]. Since $f_{0,\,0} + f_{0,\,1} = f_0 = 1/2$, it follows that $f_{0,\,1} = \sqrt{2} - 1 \approx 0.414$. A symmetric argument establishes that $f_{1,\,1} = 1.5 - \sqrt{2}$, and $f_{1,\,0} = \sqrt{2} - 1$. Direct evaluation now gives that ApEn(1)$\{B(\sqrt{2})\} = \lim_{N\to\infty}$ ApEn$(1, N)$ $\{B(\sqrt{2})\} \approx 0.458$.

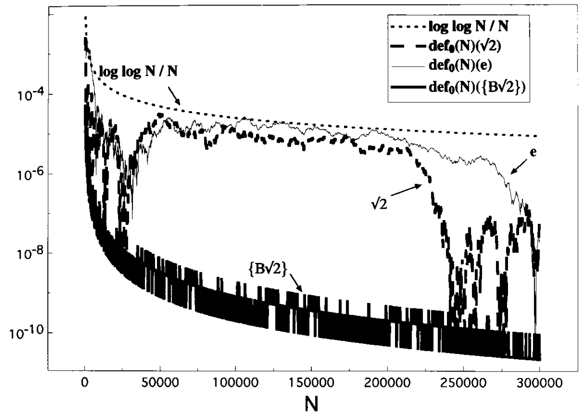

FIG. 3. One-dimensional deficit from maximal irregularity def$_0(N)$ for base 2 sequence expansions of e, $\sqrt{2}$, and for the binary sequence $\{B\sqrt{2}\}$, derived from fractional parts of multiples of $\sqrt{2}$, all compared with loglog N/N, where this last function is the asymptotic convergence rate of def$_0$ for sequences satisfying the law of the iterated logarithm.

Similarly, we determine that $\lim_{N\to\infty}$ ApEn$(2, N) = 0.423$, rather than log 2, manifesting the correlation among triples or 3-blocks of $B(\sqrt{2})$. Most vividly, for three contiguous measurements, asymptotic equidistribution is seen to be impossible—neither the triple {0, 0, 0} nor {1, 1, 1} can ever occur in $B(\sqrt{2})$, as seen by elementary arithmetic case analyses.

Finally, we comment that virtually the identical technique can be used to reject C-randomness of $\omega(\theta)$ for any other choice of θ.

Analytics: Asymptotic Variation of $B(\sqrt{2})$. To address the asymptotic variation of the binary sequence $B(\sqrt{2})$ associated with $\omega(\sqrt{2})$, we consider the notion of *discrepancy* (18–20). Let U be the unit interval [0, 1], and u: = $\{u(1), u(2), \ldots\}$ be any sequence of numbers in this interval. Given an a in U and a positive integer N, we define $Z(N, a)$ as the number of integers i with $1 \leq i \leq N$ and $0 \leq u(i) < a$, and we put $D(N, a)$: = $|\ Z(N, a) - Na\ |$. The discrepancy $D(N)$ is defined by $D(N)$: = $\sup_{a\in U} D(N, a)$. The sequence u is called uniformly distributed (on [0, 1]) if $D(N)$ is $o(N)$.

Van Aardenne-Ehrenfest first showed that $D(N)$ cannot remain bounded for any sequence, and subsequently she proved that there are infinitely many integers N with $D(N) > c$ loglog N/logloglog N, where $c > 0$ is an absolute constant (21). Schmidt (20) improved on this and showed that for any sequence, there is some constant k such that $D(N) > k \log N$ for infinitely many values of N (a best possible rate, given Eq. **3** below).

For sequences of fractional parts of multiples of an irrational θ, the following asymptotic bound has been known for some time: Ostrowski (22) and Hardy and Littlewood (23) showed that for such sequences $\omega(\theta) = \{\{\theta\}, \{2\theta\}, \ldots\}$, with $u(n)$: = $\{n\theta\}$, the function

$$S(N) := \sum_{i=1}^{N} (u(i) - 1/2)$$

satisfies $|S(N)| > c \log N$ for infinitely many N. Most importantly, for sequences $\omega(\theta)$, Ostrowski (22) also showed that

$$D(N) \leq 36\, A \log N, \qquad \textbf{[3]}$$

if the partial quotients in the continued fraction expansion of θ do not exceed A.

For the sequence $\omega(\sqrt{2})$, we now apply Eq. **3** with $A = 2$, since the continued fraction expansion of $\sqrt{2} = [1, 2, 2, 2, \ldots]$, and deduce that $D(N) \leq 72 \log N$. Observe that $|Z(N, a) - Na| = D(N, a) \leq D(N)$ for all a, in particular for $a = 1/2$ thus, for all N, $|Z(N, 1/2) - N/2| \leq 72 \log N$.

Now note that $Z(N, 1/2)$ = number of integers i with $1 \leq i \leq N$ and $\beta(i) = 0$. Setting

$$T_N := \sum_{i=1}^{N} (\beta(i) - 1/2),$$

we then infer that for all N,

$$|T_N| \leq 72 \log N. \qquad \textbf{[4]}$$

We complete this analysis by relating Eq. **4** to def$_0$ $(N)\{B(\sqrt{2})\}$. Recall from ref. 4 the following definition of *excess*, for a binary sequence u: {excess of "0" over "1"}$_N$ (u) = max(0, #0s in $u^{(N)}$ − #1s in $u^{(N)}$), and symmetrically for {excess of "1" over "0"}$_N$ (u). Let $EXC_N(u)$ = max ({excess of "0" over "1"}$_N$ (u)}, {excess of "1" over "0"}$_N$ (u)}). Observe that $|T_N| = 1/2\, EXC_N(\{B(\sqrt{2})\})$. Thus, from Eq. **4**, we deduce that $EXC_N(\{B(\sqrt{2})\}) \leq 144 \log N$ for all N. Now, as in ref. 4, p. 2086, there is an easily derived relationship between def$_0$ and EXC (for small values of def$_0$) applied here as def$_0(N)\{B(\sqrt{2})\} \approx$

$$\frac{1}{2}\left(\frac{EXC_N(\{B(\sqrt{2})\})}{N}\right)^2.$$

Therefore as $N \to \infty$, def$_0(N)\{B(\sqrt{2})\}$ is

$$O\left(\frac{\log N}{N}\right)^2.$$

Almost Sure and Distributional Properties of Random Variables

In the context of the present analysis, we raise an important question: How generic are the almost sure properties of independent identically distributed random variables, e.g., the law of the iterated logarithm (LIL) and the central limit theorem (CLT), for specified (sets of) sequences? Recall that for standardized binary sequences $\{X_i\}$, the LIL requires that almost surely $\lim\sup_{N\to\infty} S_N$ is asymptotic to $(2N \text{ loglog } N)^{1/2}$, where $S_N := X_1 + \ldots + X_N$. We consider the following:

(*i*) In ref. 4 it was shown that for a binary alphabet the LIL mandate is equivalent to requiring that $\lim\sup_{N\to\infty} \text{def}_0[u^{(N)}] = (\text{loglog } N)/N$, which is simply one subclass from the set of all C-random sequences. From Theorem 3, ref. 4, large classes of normal, C-random numbers violating the LIL were constructed from a single normal number, with rates of convergence of $\lim\sup_{N\to\infty} \text{def}_0$ at least as slow as $g_\beta(N) = N^{-\beta}$, for arbitrarily small $\beta > 0$. For these normal numbers, def_0 is much larger than $(\text{loglog } N)/N$, manifested in binary sequences with a remarkably slow convergence of the frequency of $\{0\}$s to 1/2.

(*ii*) The LIL does not describe the limiting asymptotic single digit variation for the binary sequence $B(\surd 2)$ associated with $\omega(\surd 2)$, fractional parts of multiples of $\surd 2$ (suggested by Fig. 3). To satisfy an LIL, the quotient function $Q(N) := \text{def}_0(N)\{B(\surd 2)\}/\{(\text{loglog } N)/N\}$ would need to approach an upper bound bound of 1 infinitely often for large N. We analytically establish that $Q(N) \to 0$ as $N \to \infty$: since $\text{def}_0(N)\{B(\surd 2)\}$ is

$$O\left(\frac{\log N}{N}\right)^2,$$

shown above, $Q(N)$is

$$O\left\{\left(\frac{\log N}{N}\right)^2 / \{(\text{loglog } N)/N\}\right\},$$

which is

$$O\left(\frac{1}{N}\right)\left(\frac{\log^2 N}{\text{loglog } N}\right).$$

Thus, $B(\surd 2)$ provides a counterpoint to the classes of non-LIL sequences indicated in *i* above. Specifically, for $B(\surd 2)$, the one-dimensional LIL is not satisfied because all initial sequence segments are remarkably nearly maximally equidistributed, much more so than LIL allows, whereas for the sequences indicated in *i*, the deviations of initial segments from equidistribution are much greater than those allowed by LIL, even though the correct limiting frequencies of 1/2 for both $\{0\}$ and $\{1\}$ are satisfied in all cases.

Notably, by a nearly identical argument, the corresponding $Q(N) \to 0$ as $N \to \infty$ for all quadratic surds θ, since the resultant continued fraction sequence $[a_0, a_1, \ldots, a_n, \ldots]$ is periodic (15), hence necessarily bounded, thus implying that Eq. **3** applies (for some $A < \infty$). Hence, the LIL is also qualitatively invalid insofar as describing single-digit deviations from centrality for $B(\theta)$ for θ any irrational root of a quadratic equation with integral coefficients, a very "nice" class of 1-dimensionally irregular sequences.

(*iii*) In ref. 3, sequences of 0s and 1s in the binary expansion of k/q were studied, for $0 < k < q$, for q the 2-ergodic prime 4093. An empirically natural state space, which we denote by $\Omega_{q,N}$, was then formed as the collection of expansions (sequences) of length N of k/q for *all* $0 < k < q$. As a consequence of the selection of q as an 2-ergodic prime, for large N, any such sequence in $\Omega_{q,N}$ is nearly maximally irregular. In the language of the above analysis, def_0 is small for *all* members of this state space. Thus, for $\Omega_{q,N}$, in stark contrast to the Bernoulli process, there are no occurrences or tails of rare events (e.g., of a sequence of all 0s or of all 1s); instead, there are strict cut-offs in the distributional characteristics of sequences, shown graphically in figures 5.4 and 5.5 of ref. 3. Furthermore, these same two figures strongly suggest that the distribution function of the frequency of 1s, while increasingly tightly centered about 0.5 as $N \to \infty$, not only is decidedly nonnormal, but indeed may be singular (to Lebesgue measure), i.e., there may be no density function for a limit law. We conclude that for the set of sequences given by the state space $\Omega_{q,N}$, the CLT does not hold for large N.

Given these examples, we infer that whereas the almost sure laws and distributional properties are verifiable within axiomatic probability theory, the validity of these laws as they apply to specified sequences, or sets of sequences, must be determined *ex nihilo* on a case-by-case basis. Of course, these laws remain useful in that they allow one to pose reasonable, quantitative hypotheses about sequential characteristics that often are valid, e.g., the possibility that base 2 (and undisplayed base 10) digit expansions of e and $\surd 2$ satisfy the LIL, as suggested by Fig. 3. As well, the important Kac–Erdos Gaussian law of errors for additive functions allows one to prove that asymptotically, the renormalized density of $\omega(n)$, the number of distinct prime divisors of the integer n, satisfies the CLT (24). This theorem is foundational in what has come to be known as probabilistic number theory, for which CLTs have now been established in a range of thematically similar settings to that considered by Kac and Erdos (25, 26). Nonetheless, the point remains that any prescribed collection of almost sure properties will hold for certain (sets of) sequences and fail for others.

Perspective and Future Direction

(*i*) We now clarify the punctuation in the title. While we do not aim to elucidate a vague notion of a specific "Random" infinite sequence, we do have an explicit, computable, frequency-based formulation of C-randomness. Moreover, we do not need to determine whether such numbers as e, π, $\surd 2$, and $\surd 3$ are indeed C-random, hence the "(possibly)" of the title, to provide considerable information, via ApEn(m, N) and $\text{def}_m(N)$, on (large-scale) proximity of finite initial segments to maximal frequency equidistribution and C-randomness.

(*ii*) It is remarkable that while relative frequencies play a fundamental role in the *intuitive* justification of theories of both probability and statistics, the evolution of formal theories explicitly derived from a frequency-based foundation diverged dramatically from those given by the axiomatic theory, and at least as expounded by von Mises (27), have basically disappeared from (advanced) mathematical research. History should remind us that the formulation of a framework to study randomness remained controversial for a long time after the introduction of axiomatic probability theory by Kolmogorov (28).

De facto, we are taking ("reviving"?), a frequentist approach to randomness, albeit from a considerably different perspective from that of von Mises, insofar as we feature (*a*) an entropy-like concept (in the aggregation of block data to form a single measure of irregularity); (*b*) explicit multi-dimensional or m-block evaluations; (*c*) applicability to small length datasets, e.g., sequences of length $N \geq 5$ (4), also seen as in the Chaitin example above, and as in a number of clinical, medical applications (29–31), for length $N \geq 60$ datasets. Pragmatically, *c* may be especially important—quite possibly, von Mises' distaste for small sample theory (ref. 27, pp. 158–159) alienated a large potential group of otherwise supportive end-users.

(*iii*) Historically, the randomness of a long finite alphabet sequence has often been assessed by whether or not the sequence passed a collection of, e.g., the following tests: χ^2, Kolmogorov–Smirnov, serial, poker, gap, run, as theoretically discussed in ref. 32 and applied, e.g., by Stoneham (33). However, in essence, the aforementioned tests presume the almost sure laws, and in particular, underlying binomial or normal distributions. As discussed above, since neither the almost sure laws nor a specified limiting distribution need be satisfied for specific C-random sequences, interpretation of these tests insofar as establishing a notion of randomness is problematic.

Furthermore, these tests are binary—"possibly random" or "nonrandom"—rather than providing a linear relation (as in, e.g., proximity to maximal irregularity). The utility of grading several nonrandom sequences in order of increasing irregularity is apparent in considering the myriad claims of chaos in time-series data, as discussed in refs. 9 and 34.

(*iv*) To highlight the concern with possible over interpretation of such tests, one need look no further than classical studies of both von Neumann *et al.* (35) and Fisher and Yates (36). In the former, the first 2,000 decimal digits of *e* were assessed by a χ^2 test, with $\chi^2 = 1.11$. This was remarked to be "very conspicuous," with "a significance level of about 1:1250." The comment was then made "thus something number-theoretically significant may be occurring at about $n = 2{,}000$." But Stoneham (table 1 in ref. 33) established that this very low χ^2 value was singular among the first 60,000 digits of *e*—von Neumann's observation simply indicates that at the precise cut-point $N = 2{,}000$, there is nearly one-dimensional maximal equidistribution of the digits 0, 1, . . . 9. And upon reconsideration of Figs. 1 and 2, it is clear that *e* is not consistently especially better equidistributed, as a function of sequence length *N*, than the other irrationals studied, either 1-, 2-, or 3-dimensionally, in either base 10 or base 2. Indeed, in base 10, for the range $75{,}000 \leq N \leq 250{,}000$, *e* has the poorest single-digit equidistribution among $\sqrt{3}$, $\sqrt{2}$, *e*, π, as seen in Fig. 2*A*. As well, in base 2, short initial segments of *e* are not remarkably equidistributed, with a relative ranking from most to least 1-dimensionally equidistributed for $N = 400$ of $\sqrt{3}$, $\sqrt{2}$, *e*, π, and for $N = 2{,}000$ of $\sqrt{3}$, $\sqrt{2}$, π, *e*.

Fisher and Yates (ref. 36, pp. 18–19) observed that there was an "excess of sixes" in their attempts to construct "random" numbers by selecting digits from the 15–19 places of a table of 20 place logarithms to the base 10; they quantified this via a χ^2 value of 15.63, "which corresponds to a probability of 0.075." They then went on "to reduce the number of sixes so as to give a more normal sample" (for a resultant standard random number table), which was done by "picking out 50 of the sixes strictly *at random* and replacing each of them by one of the other 9 digits selected *at random*." Disregarding the obvious objection to the means of the *at random* procedures of the last sentence, the more serious objection concerns the need to meddle with a well-defined initial table of reasonably, yet not maximally irregular numbers, simply to achieve a resultant more typical value of χ^2 .

(*v*) Within algorithmic complexity, there is considerable concern that both a sequence and all properly chosen subsequences should all be "random," for the appellation of randomness to be conferred (12). We believe that it is imperative, particularly for finite sequences, to separate two very distinct issues: **(A)** how does one quantify the regularity of a presented sequence?; **(B)** what are the properly chosen subsequences ?, to each of which one can then ask **A**. Evidently, the development of ApEn is directed at **A**. There is no consensus on **B**, since a response to **B** is typically application-specific. Nonetheless, we propose the following response to the aforementioned concern: A length *N* sequence $u^{(N)}$ is denoted $\underline{\varepsilon -}$ random w.r.t. {Appl}, if for all subsequences u_{sub} of $u^{(N)}$ in a specified collection {Appl}, $\text{De}[\text{u}_{\text{sub}}] < \varepsilon$, recalling Definition 8 in ref. 4. Thus, all flagged subsequences would be nearly maximally irregular.

(*vi*) An evaluation of cryptosystems via ApEn could prove productive. Specifically, relationships between the size of ApEn(*m*) values and each of (*a*) predictability of sequential output and (*b*) reconstructability of a key are important, yet unaddressed topics. Studies should include reevaluation of the pseudorandom number generators specified in sections 2 and 7 of ref. 37 (*en passant*, assessing the topicality of, e.g., the factoring and discrete logarithm problems), and especially of Shannon's classic treatise on this topic (38). As well, heavily studied bit generators such as RSA, modified Rabin, and discrete exponential methods (ref. 37, pp. 130–136) all require a source of *uniform samples*, a vagueness and deficiency given the above analysis. Indeed, a presumption of the availability of (a source of) "truly random bits" is central to a vast array of cryptosystems (39).

(*vii*) As developed above, the notions of both maximal irregularity of ApEn(*m*, *N*), and of $\text{def}_m(N)$ require a finite state space. A corresponding treatment for the reals is forthcoming, featuring maximal irregularity at a prescribed resolution level *r*, i.e., maximal ApEn(*m*, *r*, *N*). Consideration of the flip-flop pair of processes (40) indicates that such relative (to resolution level) maximality is the best that one can do. For the reals, a related notion to normality, denoted ∞-distributed, was introduced by Franklin in an interesting paper (41), which obtains results about the distribution properties of many special sequences. However, ref. 41 is concerned exclusively with infinite length sequences, and importantly, as for normality, there is a nearly nonexistent collection of explicitly known ∞-distributed reals, the (relatively) easiest and first construction given by Knuth (42), related to Champernowne's proof that .1234567891011. . . is normal (43).

We thank Margherita Pierantoni for her assistance in creating and providing the files for the base 2 and base 10 digits of *e*, π, $\sqrt{2}$, and $\sqrt{3}$, analyzed above.

1. Kalman, R. E. (1994) *Model. Identif. Control* **15,** 141–151.
2. Kalman, R. E. (1995) *Math. Jpn.* **41,** 41–58.
3. Kalman, R. E. (1996) *CWI Quarterly on Control and System Theory*, in press.
4. Pincus, S. & Singer, B. H. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 2083–2088.
5. Kolmogorov, A. N. (1933) *Grundbegriffe der Wahrscheinlichkeitsrechnung* (Springer, Berlin).
6. Feller, W. (1968) *An Introduction to Probability Theory and Its Applications* (Wiley, New York), 3rd Ed, Vol. 1.
7. Borel, E. (1909) *Rend. Circ. Mat. Palermo* **27,** 247–271.
8. Geiringer, H. (1954) *Studies in Mathematics and Mechanics* (Academic, New York), pp. 310–322.
9. Pincus, S. M. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 2297–2301.
10. Chaitin, G. J. (1966) *J. ACM* **13,** 547–569.
11. Chaitin, G. J. (1975) *J. ACM* **22,** 329–340.
12. Kolmogorov, A. N. & Uspenskii, V. A. (1987) *Theory Probab. Appl.* **32,** 389–412.
13. Martin-Lof, P. (1966) *Inf. Control* **9,** 602–619.
14. Pincus, S. M. & Goldberger, A. L. (1994) *Am. J. Physiol.* **266,** H1643–H1656.
15. Hardy, G. H. & Wright, E. M. (1979) *An Introduction to the Theory of Numbers* (Oxford Univ. Press, Oxford), 5th Ed., pp. 63–71, 125–128.
16. Chaitin, G. J. (1975) *Sci. Am.* **232,** 47–52.
17. Weyl, H. (1916) *Math. Ann.* **77,** 313–352.
18. Hua, L. K. & Yuan, W. (1981) *Applications of Number Theory to Numerical Analysis* (Springer, Berlin/Science Press, Beijing), Chapt. 3–4.
19. Koksma, J. F. (1950) *Math. Cent. Amsterdam Scriptum* **5,** 1–51.
20. Schmidt, W. M. (1972) *Acta Arith.* **21,** 45–50.
21. Van Aardenne-Ehrenfest, T. (1949) *Indagationes Math.* **11,** 264–269.
22. Ostrowski, A. (1922) *Abh. Math. Semin. Univ. Hamb.* **1,** 77–98.
23. Hardy, G. H. & Littlewood, J. E. (1922) *Abh. Math. Semin. Univ. Hamb.* **1,** 212–249.
24. Erdos, P. & Kac, M. (1939) *Proc. Natl. Acad. Sci. USA* **25,** 206–207.
25. Elliott, P. D. T. A. (1980) *Probabilistic Number Theory* (Springer, Berlin), Vols. 1 and 2.
26. Kubilius, J. (1964) *Probabilistic Methods in the Theory of Numbers: Translations of Mathematical Monographs, Vol. 11* (Am. Math. Soc., Providence, RI).
27. von Mises, R. (1957) *Probability, Statistics, and Truth* (Macmillan, New York), 2nd Ed.
28. Fine, T. (1973) *Theories of Probability* (Academic, New York).
29. Pincus, S. M., Cummins, T. R. & Haddad, G. G. (1993) *Am. J. Physiol.* **264,** R638–R646.
30. Pincus, S. M., Gevers, E. F., Robinson, I. C. A. F., van den Berg, G., Roelfsema, F., Hartman, M. L. & Veldhuis, J. D. (1996) *Am. J. Physiol.* **270,** E107–E115.
31. Pincus, S. M., Mulligan, T., Iranmanesh, A., Gheorghiu, S., Godschalk, M. & Veldhuis, J. D. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 14100–14105.
32. Knuth, D. E. (1981) *Seminumerical Algorithms: The Art of Computer Programming, Vol. 2* (Addison–Wesley, Reading, MA), 2nd Ed., Chapt. 3.
33. Stoneham, R. G. (1965) *Am. Math. Mon.* **72,** 483–500.
34. Pincus, S. M. (1994) *Math. Biosci.* **122,** 161–181.
35. Metropolis, N., Reitwiesner, G. & von Neumann, J. (1950) *Math. Tables Other Aids Computation* **4,** 11–15.
36. Fisher, R. & Yates, F. (1938) *Statistical Tables for Biological, Agricultural, and Medical Research* (Oliver & Boyd, London), 3rd Ed., pp. 18–19.
37. Lagarias, J.C. (1990) *Proc. Symp. Appl. Math.* **42,** 115–143.
38. Shannon, C.E. (1949) *Bell Syst. Tech. J.* **28,** 656–715.
39. Luby, M. (1996) *Pseudorandomness and Cryptographic Applications* (Princeton Univ. Press, Princeton).
40. Pincus, S. M. & Huang, W.-M. (1992) *Commun. Statist.-Theory Methods* **21,** 3061–3077.
41. Franklin, J. N. (1963) *Math. Comput.* **17,** 28–59.
42. Knuth, D. E. (1965) *BIT* **5,** 246–250.
43. Champernowne, D. G. (1933) *J. London Math. Soc.* **8,** 254–260.