CWI Quarterly Volume 9 (3) 1996, pp. 181 – 204

Probability in the Real World as a System Attribute

R.E. Kalman
Swiss Federal Institute of Technology, Zürich, Switzerland
Istituto "Asinius Pollio"

1. Overview

Randomness, in some intuitive sense, is all around us in the real world. Hence it is the proper object of *scientific* study.

In two recent publications (Kalman [8, 9]) we have argued that a *scientific* study of randomness (never really undertaken before) should begin by separating the intuitive concept of "random" from any immanent or implied relation with "probability", whatever that word may mean. This is just the opposite of the usual procedure; for example, in axiomatic probability theory "random" is given a meaning only after the probabilities have been defined.

In this paper we take a position, the *scientific* one, on the opposite side from that of Dennis Lindley, a well-known Bayesian, whose conviction is that

> *Probability is the only satisfactory description of uncertainty.*
>
> Lindley [10, p. 17].

Lindley's position might be labeled – without in any way questioning its honesty – as a *religious*[1] one, or a *dogmatic* one, or even a *fanatical* one.

Indeed, "mainstream" probability theory in its various forms (abstract axiomatic, naïve, Bayesian, etc.) has come largely under the influence of the *religious* point of view, but there is surely no consensus; see, for example the introductory comments of Pincus and Singer [12].

Granted that there is no rigid connection (in the scientific sense) between randomness and probability and that randomness, as a basic phenomenon is now somewhat better understood in the light of Kalman [8, 9], we may venture to proceed, as the next step, to the *scientific* study of probability. We should ask:

- does probability exist in the real world?
- if so how is it created?
- is it a physical concept?

The history of probability is *not* a chapter in the history of science. Probability theory – or more properly, abstract axiomatic probability as it is conceived today – is the result of a long epistemological development, mainly since about 1700, and it was influenced in turn by problems motivated by games of chance, lotteries, insurance, opinion surveys, prediction of economic events, etc., all of which are man-made phenomena. In this evolution, the questions asked were: what is chance? what is a random choice? what is probability? – always questions in the abstract and not primarily aimed at a better understanding of the real world. Research in probability resulted in a nice axiomatic formatting of abstract answers to the abstract questions, but there was no concern for contact with the physical world. Yet the net result, perhaps regrettably, was that "probability" acquired a meaning that made it seem, in the eyes of the man of the street and many intellectuals as well, at least as real and as physical as, say, mass or gravitation or advertising.

The main object of this paper is to initiate a study of probability from the scientific point of view. We are quickly led to two major results:

(i) The old-fashioned, pre-abstract notion of probability, known as relative *frequency* (the number of interesting events divided by the number of all events), is a perfectly natural mathematical starting point; it hasn't been adequately researched until now; it could lead to a rich theory in the near future.
(ii) Frequency (the old, revived notion of probability) is not universal; it does not exist in the abstract; it cannot be made to exist by axiom; it must be deduced from "interactions" within a system; it is a system attribute.

2. Why is probability NOT a satisfactory way of looking at randomness?

There are two aspects:

(a) *technical:* rigorous probability theory lives in world $\mathcal{A}$, a world, which, seemingly, is incommunicado with concrete mathematics that lives in world $\mathcal{R}$ that, for us, is the real world;
(b) *operational:* probability theory says almost nothing about different "kinds" of randomness.

See Figure 1.

Both difficulties are old and very well known. As an illustration, let us mention two typical unsolved problems:

(i) *The infinite sequence of decimal digits of an irrational number.*

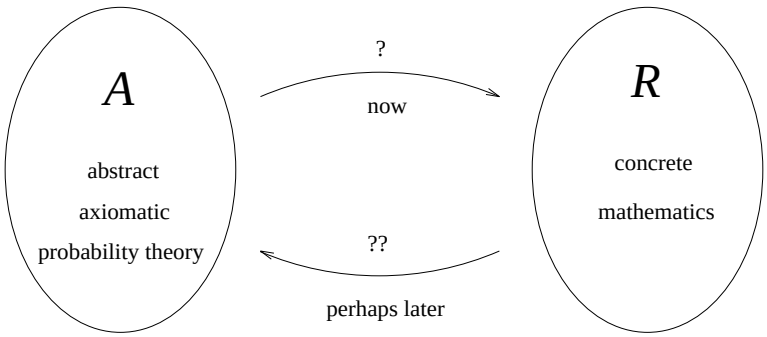

FIGURE 1

Consider

$$\sqrt{2} - 1 = 0.4142135623\ 7309504880\ 1688724209\ldots \tag{2.1}$$

Examination of (even a few of) these digits indicates that they are, to the naked eye and to formal statistical analysis, random, in some sense. More precisely, it can be verified, by using lots of digits (if needed, millions), that each single digit occurs approximately $\frac{1}{10}$ of the time, each pair of adjacent digits occurs approximately $\frac{1}{100}$ of the time, each triple $\frac{1}{1000}$ of the time and so forth. This is the empirical evidence. It should be possible to formulate a theorem in the $\mathcal{R}$-world which expresses these known facts in precise language, but no such theorem is presently known.

However, there is a classical theorem of BOREL [2] (see HARDY and WRIGHT [6 p. 124]), in the $\mathcal{A}$-world, which asserts that almost every number on the unit interval is *normal*, that is, the above property concerning the frequency of digits holds not only for decimal expansions but for all expansions to an integer base $= 2$ (binary expansion), 3 (ternary expansion), ... This is an early, pioneering example of defining randomness (here made precise as "normality") via abstract probability. The reason why Borel's theorem is not applicable to the $\mathcal{R}$-world is that the command "take a point *at random* on the unit interval" cannot be defined (as far as is known) in the $\mathcal{R}$-world.

Since the arrow $\mathcal{A} \to \mathcal{R}$ in Figure 1 is not known (and probably does not exist) there is no way of transferring a definition of randomness from $\mathcal{A}$ to $\mathcal{R}$. And if we want to prove the randomness of the decimal digits of a *particular* number such as $\sqrt{2} - 1$, it seems that we must somehow do it without help from Borel's theorem, work exclusively in the $\mathcal{R}$-world, and give up hope of being guided by probability theory.

(ii) *Finite sequences of zeros and ones.*

Consider the following sequence of sixty zeros and ones.

$$0111100110\ 1101000111\ 0101100000\ 1000011001\ 0010111000\ \ 1010011111 \tag{2.2}$$

This sequence could be regarded as part of a coin-tossing (or Bernoulli) sequence (see FELLER [5, p. 104]). But is it really that? Does it contain, perhaps, some typing errors?[2]

Let's look at the sequence (2.2) from the point of view of digit statistics. The result is highly suggestive:

digit sequence	0	1	00	01	10	11	000	001	010	011	...
number of occurrences	30	30	15	15	15	15	7	8	8	8	...

Each consecutive digit sequence of length $r \leq 3$ seems to have (almost exactly) the same frequency – is this a case analogous to (2.1)?

For both examples, the basic question, relevant to the $\mathcal{R}$-world, is: are the sequences (2.1) and (2.2) "random"? In what sense?

The usual answer of epistemologists (ATLAN [1, p. 118], CHATFIELD [3]), slightly paraphrased, is the following:

> *if you know* that the sequence was generated by tossing a (fair) coin (an event in the $\mathcal{A}$-world), then the sequence is *random* (i.e., it is a classical Bernoulli sequence, with equal probabilities of $\frac{1}{2}$ for 0 and 1);

but

> *if you know* that the sequence was generated by a known rule (an event in the $\mathcal{R}$-world), then the sequence is *not random*.

The dilemma is: if we are given the sequences (2.1) or (2.2), but no other information, what do we know about them? Can we determine how they were generated?

It seems that an element of subjectivity cannot be excluded; therefore the concept "random" cannot be well defined mathematically in a concrete case like (2.1) or (2.2). The phrase "a finite sequence is random" in such a context has no *scientific* meaning, it cannot be either confirmed or contradicted.

Probability is not much help in dealing with randomness in the $\mathcal{R}$-world. If (2.2) *is* Bernoulli[3] its probability is 2^{-60}, as is the probability of *any* Bernoulli sequence of sixty digits. Probability theory in this situation merely describes the collection of all 0/1 sequences of sixty digits and states what the probability of each individual sequence is.[4]

The difficulty in not being able to define randomness via conventional probability theory in the $\mathcal{R}$-world can only be avoided by changing the very definition of randomness (or nonrandomness) which means going back to the most primitive level of the conceptual discussion. Such considerations led to a new (primordial) definition of randomness adopted in KALMAN [8, 9].

Accordingly, in this paper, any sequence is called RANDOM[5] if it is (rather vaguely speaking) not "unique". More precisely, for the digit sequences of the two examples, nonRANDOMness means "consisting of a single digit". Thus our

RANDOMness is the opposite of (strong) regularity, much like "nonlinear" is the opposite of "linear". RANDOMness is not the opposite of "deterministic".

Starting with the new notion of RANDOMness we shall show that it leads naturally to a well-defined notion of "probability"; this is not the abstract probability of the mainstream theory, nor the naïve probability of dice playing or card games – but the classical idea of relative frequency.[6]

3. Classical results about randomness in the real world

The examples in the preceding sections are frequently viewed intuitively as the $\mathcal{R}$-world projection, or approximation, of i.i.d processes in the $\mathcal{A}$-world. ("i.i.d" means *independent, identically distributed.*) But it seems that there are in fact no i.i.d. processes in the $\mathcal{R}$-world – at least not in the strict sense. The trouble is the requirement of "independence".

Because we are interested in randomness in the $\mathcal{R}$-world, we need to modify the problem setting to get away from the constraints imposed by the $\mathcal{A}$-world notion of "independence". This suggests two questions:

(I) How does randomness arise in the $\mathcal{R}$-world?
(II) Why do we often observe "equidistribution", that is, equal frequencies, of random events in the $\mathcal{R}$-world?

Relevant to both of these questions is a famous theorem (situated in the $\mathcal{R}$-world, period 1909–1916), usually known as the *Gleichverteilungssatz* (equidistribution theorem), reviewed at length in KALMAN [8, p. 148 and 9, p. 48].

We first define *equidistribution.* Consider a finite set of points $X_T = \{x_t : t = 1, \ldots, T\}$ on the open unit interval and let $J = (a, b) \subset (0, 1)$. The points in X_T *are exactly equidistributed* if the following is true: for all J

$$\frac{\text{number } \{x_t \in J\}}{T} = b - a = \text{ length } J. \tag{3.1}$$

This condition is rather hard to satisfy for *every* interval J since the ratio on the left side of (3.1) is rational by definition while $b - a =$ length J may be irrational. So it is natural to claim equidistribution holding only in the limit $T \to \infty$.

This provides some rationale for the classical formulation of the

GLEICHVERTEILUNGSSATZ (WEYL [14]). *Consider the set* $X_T = \{\alpha t \pmod 1 : t = 1, \ldots, T\}$. *Exact equidistribution on the open unit interval holds in the limit* $T \to \infty$ *if and only if* $\alpha =$ *irrational.*

If we view $(\alpha \mod 1) \cdot 2\pi$ as a fixed rotation (less than a full turn), then given any angle $\theta < 2\pi$ the angle $(\alpha t \mod 1) \cdot 2\pi$ will be arbitrarily close to θ for some t; and for large T the rotations $\{(\alpha t \mod 1) \cdot 2\pi,\ t \le T\}$ will be spread out roughly evenly over the circle, not omitting any subinterval. This is the operational meaning of the technical term "ergodic". In current pure-mathematical language, we would say: *irrational rotations are ergodic.*

Mathematicians educated in the early 1900's tended to view the theorem as a new characterization of irrational numbers.

In KALMAN [8, 9] we pointed out that the theorem can be viewed, alternatively, as responding to a fundamental problem of statistics: how to take a finite random sample of (the population of) points uniformly distributed on the unit interval – but *without* requiring that the sampling process, $\ldots$, $\alpha t \pmod 1$, $\alpha(t+1) \pmod 1$, $\ldots$ is independent from sample to sample.

In view of these well-known aspects of the Gleichverteilungssatz, it is perhaps surprising that it also answers, implicitly, questions (I) and (II) posed at the beginning. To make this clear, we reformulate the theorem. The condition "α = irrational" is dropped because it turns out to be irrelevant for our purposes; "exact equidistribution" is replaced by "as equidistributed as possible", in order to avoid technicalities stemming from the format rather than the content of formula (3.1); finally, after these changes, it is possible to relax the "unphysical" requirement $T \to \infty$. Roughly speaking we arrive at the conclusion: *all rotations are ergodic.*

The new formulation is:

EXTENDED EQUIDISTRIBUTION THEOREM. *Consider the set* $X_T = \{\alpha t \pmod 1 : \alpha > 0,\ t = 1,\ 2,\ \ldots,\ T\}$. *For suitably large* T *these points on the open unit interval are as equidistributed as possible.*

SKETCH OF PROOF. First consider a rational $\alpha = k/q$, k, q coprime positive integers, $k < q$. The definition of the set X_{q-1} suggests examining the map

$$\rho : t \mapsto kt \pmod q \tag{3.2}$$

which sends the positive integers Z^+ into the finite set

$$Z^+_{q-1} := \{1,\ 2,\ \ldots,\ q-1\}. \tag{3.3}$$

It is trivial but essential to note that

$$\rho \textit{ is an isomorphism on } Z^+_{q-1}. \tag{3.4}$$

Indeed, $k(t_2 - t_1) = 0 \pmod q$ implies $q|(t_2 - t_1)$ which is impossible for $t_1,\ t_2$ in Z^+_{q-1}.

As t is enumerated $1,\ 2,\ \ldots,\ q-1$ the points

$$u = kt \pmod q, \qquad t \in Z^+_{q-1}$$

exhaust the set Z^+_{q-1}, in some RANDOM sequence. As t is enumerated, the corresponding points

$$\left\{ \frac{kt \pmod q}{q},\ t = 1,\ \ldots,\ q-1 \right\} \tag{3.5}$$

are placed RANDOMLY on the unit interval; by (3.4) the points (3.5) are just

$$Z^+_{q-1}/q = \left\{ \frac{l}{q},\ l \in Z^+_{q-1} \right\} \tag{3.6}$$

and so these points are uniformly distributed on the unit interval. (The end points of the interval are *not* counted, in accordance with the condition $\alpha > 0$; so the indices $t = lq = 0 \pmod q$ are omitted in the enumeration of points for T large, and the number of T of all "events" in the denominator of (3.1) is correspondingly modified.)

This description of the set X_{q-1} makes it clear how "as equidistributed as possible" should be interpreted. For example, for $q = 2$, we consider the sequence $t = 1,\ 3,\ 5,\ \ldots$ (the other values of t yielding 0 (mod 2) which is disregarded), so we have a set consisting of copies of the midpoint $\frac{1}{2}$ of the unit interval. Given that this point is the only one allowed in the interior of the unit interval, the set is "as equidistributed as possible". The same picture holds for $\alpha = \frac{2}{3},\ \frac{1}{4},\ \frac{3}{5}$, etc.[7] "Suitably large T" means $T = q,\ 2q,\ \ldots$, and it is understood that the "events" $t = q,\ 2q,\ \ldots$ are excluded from the count.

The case of $\alpha =$ irrational is handled in the same way as in the proof of Gleichverteilungssatz given in HUA [7, Theorem 10.11.1, p. 269]; in this case we get "exact equidistribution" as in the classical version. □

Precisely, how does all this analysis answer the two questions posed at the beginning?

QUESTION (I). RANDOMness arises from the fact that:

$$\rho = \text{ isomorphism of a finite set} = \text{permutation.}$$

A permutation is an example of RANDOMness; a permutation is a basic randomizing operation, as in shuffling cards. A single permutation is of course not expected to produce "perfect" randomness, just as a single shuffle of cards will not produce a "perfectly" shuffled deck, whatever that means. (See the work of DIACONIS [12] and of his collaborators.) In the $\mathcal{R}$-world there seems no way to avoid viewing any permutation as RANDOM.[8]

QUESTION (II). Equidistribution arises from the fact that the permutation ρ "lives" on the largest possible set in the unit interval: Z^+_{q-1}/q for rational α, the whole open interval for irrational α. This situation is peculiar to the special definition of the set X_T in the Gleichverteilungssatz and to the special map ρ given by (3.2).

It is worthwhile to recall here a remark of THOM [1986, p. 25]. Consider an irrational (transcendental) of the Liouville type, for example,

$$\alpha = \sum_{t>0} 10^{-t!}.$$

It is well known that such irrationals are extremely well approximable by rationals; for any given r there is a (large) q so that we have

$$\left|\alpha - \frac{k}{q}\right| < \frac{1}{q^r}.$$

The set X_T has its points tightly clustered around $\{1/q,\ \ldots,\ (q-1)/q\}$ and these points are "as equidistributed as possible", as long as $T < q^{r-2}$ (say);

but as $q^r < T \to \infty$, the new points slowly drift away and in the limit become exactly equidistributed. Thus there is a kind of transition between different kinds of RANDOMness but not a violation of the claim "as equidistributed as possible". In other words, conventional statistics (histograms) would not reveal any unusual behavior. Approximate equidistribution holds for all large T but there is something happening which is not yet well understood.

4. Randomness of the digits representing rational numbers

The sequences $\{\alpha t \pmod 1\}$ analyzed by the Gleichverteilungssatz may be viewed as a "model" of certain kind of RANDOMness in the $\mathcal{R}$-world. Unfortunately, the behavior of these sequences does not seem to be relevant to understanding the examples of Section 2. Perhaps the Weyl model is too simple.

It is surprising that a deeper and probably more "natural" study of randomness and probability can be initiated by looking at the very old, very elementary, but not completely understood problem of the "fractional part" of a rational number as displayed by its expansion with respect to an arbitrary base b (= integer > 1).

For k, q positive integers with $(k,\ q) = 1$ consider

$$\begin{aligned} \frac{k}{q} \pmod 1 \quad &= \sum_{t \geq 1} \beta_t b^{-t}, \qquad 0 \leq \beta_t < b, \\ &= . \text{ concatenated sequence of } b\text{-ary digits.} \end{aligned} \tag{4.1}$$

Elementary number theory, of the type discussed in secondary school (see HARDY and WRIGHT [6, Chapter IX]), tells us that, for every integer $b > 1$,

- *The expansion defines of a unique infinite sequence of integers β_t.*
- *The expansion is ultimately periodic.*
- *If $(b, q) = 1$ the expansion is periodic.*

All this is well known. What is seldom or never discussed is the explicit determination of the period of the expansion as a function of $b,\ q$, and k.

To determine the period, we first note that

$$\frac{b^{t-1}k \pmod q}{q} = \sum_{s \geq t} \beta_s b^{-s+(t-1)}, \qquad t = 1,\ 2,\ \ldots; \tag{4.2}$$

or, as a concatenated sequence,

$$= .\beta_t \beta_{t+1} \ \ldots;$$

in other words multiplying k/q by $b^{t-1} \pmod 1$ shifts the "b-ary point" (analogous to the decimal point) to just before the digit β_t, discarding all digits before the b-ary point. This pinpoints the digit β_t.

This suggests introducing a new map σ which is to play a role similar to ρ given by (3.2). Define by

$$\sigma : t \mapsto b^{t-1}k \pmod{q} \tag{4.3}$$

a map from the positive integers Z^+ to the set Z^+_{q-1}. Let $t_2 = s$ be the first integer t in the sequence 1, 2, ... such that some point in Z^+_{q-1} is reached for a second time, and let $t_1 = r < s$ be the first time that the same point was reached . Then

$$b^s - b^r = b^r(b^{s-r} - 1) = 0 \pmod{q}.$$

Assuming that $b,\ q$ are coprime (denoted as $(b,\ q) = 1$) there exists a positive integer $b^{-1} \pmod{q}$; multiplying the preceding relation by b^{-r} gives

$$b^{s-r} = 1 \pmod{q}. \tag{4.4}$$

Comparing this result with the definition of the "shift" map σ shows that *the least period of the expansion of q in base b* may be defined as

$$\pi_b(q) = s - r. \tag{4.5}$$

Clearly this definition of $\pi_b(q)$ is independent of k. The abstract construction used in (4.5) gives no information about $\pi_b(q)$ as a function of q. However, elementary number theory, namely the "little Fermat theorem" and the Euler function φ (see HARDY and WRIGHT [6, Theorems 71, 72, and 88], HUA [7, Theorem 7.4, p. 48]), tells us much more, namely

$$\pi_b(q) | \varphi(q) \tag{4.6}$$

for all $b,\ q$ with $(b,\ q) = 1$. If $q =$ prime $= p$ we have in particular, recalling that $\varphi(p) = p - 1$,

$$\pi_b(p) | p - 1. \tag{4.7}$$

If, as a special case, $\pi_b(p) = p - 1$, we call the prime p *ergodic relative to b* or simply a *b-ergodic prime*.

To stress the importance of this definition, we recall that in elementary number theory b is known as a *primitive root* of a prime p if and only if $p - 1$ is the *least* (integer) exponent l for which

$$b^l = 1 \pmod{p}. \tag{4.8}$$

By the "little Fermat theorem" (4.8) is true with exponent $l = p - 1$ for all $b < p$; conversely, if (4.7) is true for $l < p - 1$ then $l = 0 \pmod{p - 1}$.

Thus, a fixed prime p_0 *is ergodic relative to all b which are primitive roots of p_0; and, dually, for a fixed b_0 the b_0-ergodic primes are those primes for which b_0 is a primitive root.*

$\pi_b(61)$	$\varphi(\pi_b(61))$	b
60	16	2, 6, 7, 10, 17, 18, 26, 30, 31, 35, 43, 44, 51, 54, 55, 59
30	8	4, 5, 19, 36, 39, 45, 46, 49
20	8	8, 23, 24, 28, 33, 37, 38, 53
15	8	12, 15, 16, 22, 25, 42, 56, 57
12	4	21, 29, 32, 40
10	4	3, 27, 41, 52
6	2	14, 48
5	4	9, 20, 34, 58
4	2	11, 50
3	2	13, 47
2	1	60
1	1	1

$\pi_b(61)$	$\varphi(\pi_b(61))$	b
60	16	2, 6, 7, 10, 17, 18, 26, 30, 31, 35, 43, 44, 51, 54, 55, 59
30	8	4, 5, 19, 36, 39, 45, 46, 49
20	8	8, 23, 24, 28, 33, 37, 38, 53
15	8	12, 15, 16, 22, 25, 42, 56, 57
12	4	21, 29, 32, 40
10	4	3, 27, 41, 52
6	2	14, 48
5	4	9, 20, 34, 58
4	2	11, 50
3	2	13, 47
2	1	60
1	1	1

TABLE 1. Primitive roots of $p = 61$ relative to periods $\pi_b(61)|60$.

Every odd prime has primitive roots; a standard theorem (HUA [7, Theorem 7.5, p. 48] and HARDY and WRIGHT [6, Theorem 110]) assures that p has $\varphi(p-1)$ primitive roots less than p. Moreover, for any $l|p-1$ there will be exactly $\varphi(l)$ numbers in Z_{p-1}^+ which satisfy (4.7) with minimum exponent l.

Table 1 shows the primitive roots of $p = 61$ relative to all divisors of $p-1 = 60$. Aside from the trivial fact that $\pi_{p-1}(p) = 2$ for all primes p, the entries for b in Table 1 are clearly RANDOM (there is no obvious pattern).

It is not known if there are infinitely many b-ergodic primes for any given integer $b > 1$.

Our interest in ergodic primes is solely due to the following easy observation:

(4.9) *Independently of $k < q$ the map σ given by (4.3) is an isomorphism on Z_{p-1}^+ if and only if $q = p$ is a prime which is ergodic relative to the base b.*

From this we get equally easily:

EQUIDISTRIBUTION OF DIGIT SEQUENCES. *For any prime p and any $0 < k < p$ the sequence of digits given by the b-ary expansion of $k/p = .\beta_1\beta_2\ldots$ induce a finite point set $X_T(k,p) = \{x_t = .\beta_t\beta_{t+1}\ldots : t = 1, 2, \ldots, T\}$.*

Claim: *$X_T(k,p)$ is as equidistributed as possible for large T if and only if p = b-ergodic prime and, if so, $X_{l(p-1)}(k,p)$ is independent of k* (for any positive integer l) *and isomorphic with Z^+_{p-1}/p.*

The statistics of the periodic sequence $k/p = .\beta_1\beta_2\ldots$ are also independent of k. To see this, define $t(k)$ by $\sigma(t) = b^{t-1}k = 1 \pmod p$, $0 < t < p$. Then $1/p = .\beta_t\beta_{t+1}\ldots$, that is, the decimal point for the sequence k/p is shifted $t-1$ steps to the right. Since the sequences are periodic independently of k, the digits of k/p are simply shifted digits of $1/p$ and conversely. Note also that $t(k)$ is a RANDOM map $k \mapsto t$.

"As equidistributed as possible" implies that each of the b possible values $h = 0, \ldots, b-1$ of the digits β_t occur as equally often as possible. To see this, simply partition the unit interval into the segments $J_h = [hb^{-1}, (h+1)b^{-1})$, $h = 0, 1, \ldots, b-1$. Then $\beta_t = h$ iff $x_t \in J_h$. By equidistribution, the number of points x_t in each J_h must be as equal as is possible, given the "granularity" of the ratios k/p. To look for adjacent pairs of digits, let $J_h = [hb^{-2}, (h+1)b^{-2})$, and so forth.

For example, $p = 9931$ is ergodic with respect to 10. Therefore the sequence of digits will have a period 9930 for all k/p; since 9930 is divisible by 10, there will be exactly 993 decimal digits of each type $0, \ldots, 9$ within a period. On the other hand, for adjacent digit-pairs the average number 99.3 is not an integer; among the 100 possible pairs $00, 01, \ldots, 99$ some will occur 99 times and others will occur 100 times. The former case covers 70 pairs, the other 30 pairs, so as to produce the exact average 99.3. Moreover, in the sequence of digit-pairs $00, 01, \ldots, 99$ those thirty pairs occurring 100 times will be distributed "as uniformly as possible" among those seventy pairs occurring 99 times. (The reader should check this out for himself.)

To illustrate the practical implications of the equidistribution theorem, we use the prime $p = 61$. We have already investigated the ergodicity of this prime with respect to the base $b < p$. (See Table 1.) Table 2 shows sixty digits of the expansion of $1/61$ to various bases. Each line is a period, but it is not a minimal period except in the ergodic cases $b = 2, 6, 7, 10$ and 30.

In the ergodic cases $b = 2, 6, 10, 30$ we see from Table 2 that each single digit $0, 1, \ldots, b-1$ occurs *exactly* the same number of times, that is, we have *exact* equidistribution of single digits. This follows from the equidistribution theorem and the (accidental) fact that $b|p-1$ for the cases chosen. For $b = 7$, another ergodic case, the digits $0, 1, \ldots, 6$ occur either 8 or 9 times (see Table 2); namely $0, 3, 6$ occur eight times, and $1, 2, 4, 5$ occur nine times; all this follows rigorously from "as equidistributed as possible".

If 61 is not ergodic relative to the given base, there is no equidistribution

Expansion of $\frac{1}{61}$ to different bases b

b	π	1	11	21	31	41	51
2	60	0000010000	1100100101	1100010100	1111101111	0011011010	0011101011
3	10	0001022212	0001022212	0001022212	0001022212	0001022212	0001022212
5	30	0020110332	1013044243	3411234314	0020110332	1013044243	3411234314
6	60	0033125040	4415445301	4342320220	5522430515	1140110254	1213235335
7	60	0054234463	5336211556	5251644062	6612432203	1330455110	1415022604
8	20	0103113424	7674664353	0103113424	7674664353	0103113424	7674664353
9	5	0128501285	0128501285	0128501285	0128501285	0128501285	0128501285
10	60	0163934426	2295081967	2131147540	9836065573	7704918032	7868852459
11	4	01a901a901	a901a901a9	01a901a901	a901a901a9	01a901a901	a901a901a9
13	3	02a02a02a0	2a02a02a02	a02a02a02a	02a02a02a0	2a02a02a02	a02a02a02a
14	6	032dab032d	ab032dab03	2dab032dab	032dab032d	ab032dab03	2dab032dab
15	15	03a4db8562	32e3e03a4d	b856232e3e	03a4db8562	32e3e03a4d	b856232e3e
16	15	04325c53ef	368eb04325	c53ef368eb	04325c53ef	368eb04325	c53ef368eb
21	12	074h49kdg3	gb074h49kd	g3gb074h49	kdg3gb074h	49kdg3gb07	4h49kdg3gb
30	60	0emikjk4rg	6qglj5c8ph	6bo2sfm3s0	tf7b9a9p2d	n3d8aohl4c	ni5r1e7q1t

Expansion of $\frac{1}{31}$ to base 3

b	π	1	11	21	31	41	51
3	30	0002121112	2102022201	0111001202	0002121112	2102022201	0111001202

TABLE 2. Behavior of the expansion of the prime 61 relative to different bases, compared with prime 31 ergodic to base 3.

and the distribution depends on k. For example, for $b = 15$ the period is 15 and therefore in case of equidistribution each digit should occur just once, but this is far from true. For $b = 5$ the period is 30 so that each digit should occur exactly 6 times; however, the digit statistics are

digit	0	1	2	3	4
occurrences	6	7	4	7	6

which is certainly not equidistribution. (The "dip" in the middle at the digit 2 is often observed, see also next section. The symmetry of the distribution follows from the theory of k/p, $p =$ ergodic, but will not be discussed here.)

The binary sequence exhibited in Section 2 is the shifted binary expansion of $1/61$ in Table 2, in fact, it is the binary expansion of $2^{34}/61 \pmod 1 = 29/61$.

Irrespective of whether 61 is ergodic or not, all sequences in Table 2 exhibit certain intuitively acceptable features of randomness; of course, they also conform with our definition of RANDOMness.

This new class of random sequences of digits has one immediate practical application: it enables precise statements to be made in a debate with the "fanatical school" of probability theory.

The simplest problem of this sort concerns the existence of i.i.d. sequences in the $\mathcal{R}$-world. Although perhaps not a random example (I learned probability from his book), the views expressed by FELLER [5, p. 105] seem typical of the prejudices prevailing in the 1950's:

> "The Bernoulli scheme of trials is a theoretical model and only experience can show whether it is suitable for the description of specified physical experiments. *Our knowledge that successive tossings of a coin conform to the Bernoulli scheme is derived from experimental evidence*[9] The *man in the street ... believes* that *after a run of seventeen heads tail becomes more probable*. This argument has nothing to do with imperfections of physical coins; it *endows nature with memory*, or, in our terminology, *it denies the statistical independence of successive trials;* ... [this] cannot be refuted by logic but has been rejected because of lack of *empirical support*." [My italics.]

It is amazing that Feller, a subtle thinker (and superb mathematician), should allow such confusion to arise in his own mind over the relationship between the $\mathcal{A}$-world and the $\mathcal{R}$-world. Independence is certainly well defined (via probabilities) in the $\mathcal{A}$-world, but it is inaccessible to rigorous experimental study in the $\mathcal{R}$-world. Our examples give, however, plenty of "empirical support" to the contention that seventeen 1's (i.e., heads) may not in some (and may in some other) circumstances be followed by another 1 with the prescribed probability of $\frac{1}{2}$.

To be specific: for a 2-ergodic prime p there are four possibilities for the binary expansion of $1/p$ (hence any k/p):

(1) If $p < 2^{17}$ there will be no sequence of seventeen 1's;

(2) if $2^{18} > p > 2^{17}$ (say), then there will be one sequence of seventeen 1's but none of eighteen 1's so the conditional frequency (= analogous to conditional probability here and hereafter) of another 1 immediately after seventeen 1's is zero;

(3) If $p \gtrsim 2^{18}$ there will be no sequence of nineteen 1's, one sequence of eighteen 1's and two sequences of seventeen 1's (which are, both, subsequences of the single sequence with eighteen 1's) but none with nineteen 1's. For short sequences of 1's the conditional frequency of the next digit being either 0 or 1 will be very nearly $\frac{1}{2}$. For the sequence (0)(seventeen 1's) the conditional

frequency of the next digit being either 0 or 1 will be exactly 0 and 1 (the next digit must be 1 to give a sequence of eighteen 1's); and for the sequence (1)(seventeen 1's) the same frequency will be exactly 1 and 0 (the next digit must be 0 since there is no sequence of nineteen ones). Thus the conditional frequency of either 0 or 1 following seventeen 1's will be exactly $\frac{1}{2}$.

(4) if $p \gg 2^{17}$ then there will be many sequences of seventeen 1's, half as many sequences of eighteen 1's, one quarter as many sequences of nineteen 1's, and so on ...; hence the conditional probability of either 0 or 1 following seventeen 1's is very close to $\frac{1}{2}$.

When Feller arrogantly accused the man in the street of "endowing Nature with memory"[10] , he was probably not aware of how the size of an ergodic prime (or more generally, the size of $\pi_2(q)$, see below) affects the "degree of independence", that is, why "independence" may continue to hold or to fail after seventeen successive 1's.

This discussion should not create the impression that the behavior of digits of k/q is ergodic (with controlled properties) only for q = ergodic prime. It appears, but it is not yet a theorem, that (q = very large) almost always implies ($\pi_b(q)$ = very large) and hence the property "very nearly ergodic" of the digit sequence.

To make these vague remarks inoffensive, we note an interesting special case[11]

$$\pi_{10}(3^r) = 3^{r-2}, \qquad r > 2. \tag{4.10}$$

Checking the decimal-digit behavior of the fraction $k/27$, $k \neq 0 \pmod 3$ by a hand calculator shows rather erratic behavior, certainly not ergodic and very much dependent on k (there are only 3 digits!). However, the situation is entirely different for large r:

(4.11) *The decimal digit sequence of the expansion of* $k/3^r$, $k \neq 0 \pmod 3$, *is very nearly "as equidistributed as possible" for large* r; σ *given by* (4.3) *is an isomorphism* $Z^+_{3^{r-2}} \mapsto \{k \pmod 9 + 9Z^+_{3^{r-2}}\}$ *and thus* NOT *independent of* k.

Already for $r = 4$, this theorem predicts that the period of nine will contain every digit except one (dependent on k).[12] For $r = 9$, $3^9 = 19683$ and $\pi_{10}(3^9) = 2187$. Here the sets Z^+_{2187} and $\{\frac{k \pmod 9}{9} + Z^+_{2187}\}$ are "almost" the same for all k. The digits are shown in Table 3. By displaying the digits in groups of five Table 3 gives the appearance of a piece of a table of 10^5 random numbers (i.e., a random sample of 437 from the set $0 < n < 10^5$) but here with strictly controlled properties.

Decimal Expansion of $\frac{1}{19683}$

	1	6	11	16	21	26	31	36	41	46
1	00005	08052	63425	29086	01331	09790	17426	20535	48747	65025
51	65665	80297	71884	36722	04440	38002	33704	21175	63379	56612
101	30503	48016	05446	32423	91911	80206	26936	95066	80892	14042
151	57481	07503	93740	79154	60041	66031	60087	38505	30915	00279
201	42894	88390	99730	73210	38459	58441	29451	81120	76411	11619
251	16374	53640	19712	44220	90128	53731	64659	85876	13676	77691
301	40882	99547	83315	55149	11344	81532	28674	49067	72341	61459
351	12716	55743	53503	02291	31738	04806	17792	00325	15368	59218
401	61504	85190	26571	15277	14271	19849	61642	02611	39054	00599
451	50210	84184	32149	57069	55240	56292	23187	52222	73027	48564
501	75130	82355	33201	23964	84275	77096	98724	78788	80251	99410
551	65894	42666	26022	45592	64339	78560	17883	45272	57023	82766
601	85464	61413	40242	84915	91728	90311	43626	47970	32972	61596
651	30137	68226	38825	38230	96072	75313	72250	16511	71061	32195
701	29543	26068	18066	35167	40334	29863	33384	13859	67586	24193
751	46644	31235	07595	38688	20809	83589	89991	36310	52177	00553
801	77737	13356	70375	45089	67128	99456	38368	13493	87796	57572
851	52451	35396	02702	84001	42254	73759	08144	08372	70741	24879
901	33749	93649	34207	18386	42483	36127	62282	17243	30640	65437
951	17929	17746	27851	44540	97444	49524	97078	69735	30457	75542
1001	34618	70649	79931	92094	70101	10247	42163	28811	66488	84824
1051	46781	48656	20078	24010	56749	47924	60498	90768	68363	56246
1101	50713	81395	11253	36584	87019	25519	48381	85235	99044	86104
1151	76045	31829	49753	59447	23873	39328	35441	75176	54829	04028
1201	85738	96255	65208	55560	63608	18980	84641	56886	65345	72981
1251	76091	04303	20581	21221	35853	27439	92277	59995	93557	89259
1301	76731	18935	12167	86059	03571	61001	87979	47467	35761	82492
1351	50622	36447	69598	13036	63059	49296	34710	15597	21587	15642
1401	94060	86470	55834	98450	43946	55286	28765	94015	13996	85007
1451	36676	31966	67174	71930	09195	75267	99776	45684	09287	20215
1501	41431	69232	33246	96438	55103	38871	10704	66900	37087	84230
1551	04623	27897	17014	68272	11299	09058	57846	87293	60361	73347
1601	55880	70924	14774	17060	40745	82126	70832	69826	75405	17197
1651	58166	94609	56155	05766	39739	87705	12625	10796	11847	78743
1701	07778	28583	04120	30686	37910	88756	79520	39831	32652	54280
1751	34344	35807	54966	21449	98221	81578	01148	19895	34115	73439
1801	00828	12579	38322	41020	16968	95798	40471	47284	45866	99182
1851	03525	88528	17151	85693	23781	94380	93786	51628	30869	27805
1901	72067	26616	87750	85098	81623	73621	90722	95889	85418	88939
1951	69415	23141	79749	02199	86790	63150	94243	76365	39145	45546
2001	91866	07732	56109	33292	68912	25931	00645	22684	55011	93923
2051	69049	43352	13128	08006	90951	58258	39556	97810	29314	63699
2101	63928	26296	80434	89305	49204	89762	73941	98038	91683	17837
2151	72798	86196	20992	73484	73301	83407	00096	53000	05080	52634

TABLE 3.

5. THE LAW OF LARGE NUMBERS IS NEITHER A CHILD NOR THE FATHER OF PROBABILITY

Many people have thought about infrequent (or nonexistent) communication between the $\mathcal{A}$-world of probability and the $\mathcal{R}$-world where the action is. This includes even those who, like Feller, have taken a fanatical position on "independence" (see the quote in the previous section). Feller writes [5, p. 141], as a preamble to motivating the "law of large numbers",

> "On several occasions we have mentioned that our *intuitive notion* of [abstract] *probability* is based on the following assumption. If in n identical trials A occurs ν times, and if n is very large, then ν/n should be near the

probability p of A. Clearly, a formal mathematical theory can never refer directly to real life[1]3 but it should at least provide theoretical counterparts [that is, theorems in the $\mathcal{A}$-world] to the phenomena [in the $\mathcal{R}$-world] which it tries to explain. Accordingly, we require that the vague introductory remark [about the *intuitive notion of probability*] be made precise in the form of a theorem [in the $\mathcal{A}$-world]." (Brackets added by the writer for clarity.)

Feller then goes on to formulate the "law of large numbers" which, from his point of view, makes more precise the intuitive notion of probability – as regards events in the $\mathcal{A}$-world.

The credibility of relating the $\mathcal{A}$-world to the $\mathcal{R}$-world in this way rests largely on an unstated assumption:

> *the* "law of large numbers" *is an inseparable consequence of intuitive* (or abstract) *probability*; its truth is derived from probability ($\mathcal{A}$-world definition) and only from probability; because it is true in the $\mathcal{A}$-world, it should be, more or less, true in the $\mathcal{R}$-world as well.

Following the lead of R. A. Fisher, almost the whole field of statistics (after about 1920 but not before) has been reorganized on the basis of this belief, as articulated by Lindley in the remark quoted in the introduction. And it is this unbending belief in abstract probability which among other, scientifically objectionable features, forces the use of the so-called confidence intervals (see discussion below).

In contrast to this line of arguments (which a scientist would qualify as prejudiced), we wish to investigate other kinds of questions:

- Does the "law of large numbers" – note that the traditional label does not include the word "probability" – exist (as a theorem) only in the $\mathcal{A}$-world of probability?
- Or does it exist (as an empirical fact or as a theorem) also in the $\mathcal{R}$-world?
- If so, is it a theorem that is logically independent of the abstract (for Feller, intuitive) notion of probability?

By a statistical analysis of a generic example like those discussed in the last section, we shall demonstrate:

> *The "law of large numbers" is independent of (abstract) probability. The scientific content of this "law" is simply that it says something interesting about "large numbers".*

And THAT IS ALL THERE IS TO IT!

No probability, as contrasted with frequency. To mention an obvious "application": from our point of view the success of classical statistical mechanics could be explained simply by the fact that Avogadro's number ($A \sim 6 \cdot 10^{23}$) is a very large number indeed; moreover, this is an objective fact about the

physical world, unrelated to any metaphysical or epistemological speculation about probability.

The technical details of our argumentation are quite simple. Let us first review the standard (FELLER [1550, p. 141]) formulation of the "law of large numbers".

Consider the (i.i.d.) Bernoulli process with two symbols, $x_t = 0$ (tail) or 1 (head) ($t = 1, 2, \ldots, T$), with Pr (1) $= p$ and Pr (0) $= 1 - p$. By the independence assumption, the probability of exactly S occurrences of 1's in a string of T symbols is equal to the frequency of 1's among the 2^T possible sequences, weighted by the probability of each sequence which (by independence) is determined solely by the total number S of 1's and $T - S$ of 0's:

$$\text{Pr } \{S \text{ occurrences of 1 in } T \text{ cases}\} = \binom{T}{S} p^S (1-p)^{T-S}. \tag{5.1}$$

The theorem called "law of large numbers" is concerned with the behavior of the probability Q of the event "the mean of a sequence of length T is far from p"; this probability is denoted as

$$\text{Pr} \left\{ \left| \frac{1}{T} \sum_1^T x_t - p \right| \geq r \right\} = Q(r, T). \tag{5.2}$$

The theorem asserts (for all fixed $0 < r < 0.5$) that the probability $Q \to 0$ as $T \to \infty$. Notice that the formulation and claim of the theorem is in terms of probabilities; this is natural and unavoidable because the entire discussion takes place in the $\mathcal{A}$-world.

We are interested in the symmetrical case where $p = \frac{1}{2}$. To see how (5.2) behaves, the discrete probability distributions of S given by (5.1) are shown in Figure 5.1 from moderate to large values of T. (For better visualization, adjacent discrete point of the distribution are connected by a straight line.) As $T \to$ large the figures show the convergence of the distributions to a "narrow pulse" having the shape of the Gaussian distribution centered at $p = 0.5$.

Convergence to "equidistribution" is obvious; the existence of the limit is guaranteed by $p =$ ergodic prime.

In short, *the* "law of large numbers" *exists as a fact of Nature and not only as a mathematical deduction from formalized intuitive probability.*

A definitive mathematical formulation of this observation, however, is unavailable at this time.

The proof of the theorem requires two steps: (1) to show that (5.1) converges with very good approximation to a suitably scaled Gaussian distribution; (2) to estimate the "tails" of the distribution, i.e., the probability of the term in { } in (5.2), namely, the probability of the event $|\frac{S}{T} - p| \geq r$.

The first step was an important and highly nontrivial result around 1720, due to de Moivre. With computers, the pictures in Figure 5.1 tell practically the whole story, so this step is of no further interest here.

The second step is the explicit calculation of the probability of the "bad" cases arising from the tails of the distributions. The discussion began by viewing the Bernoulli process as a model for "random" 0/1 sequences in the real world (see Feller's comments). But now sequences corresponding to the tail of the binomial distribution (which always include the sequence containing only 0's and the sequence containing only 1's) are declared as "bad", that is, "not sufficiently random", because they violate the intuitive test for randomness, that $S/T \sim p$.

The number $0 < r < 0.5$ is "adjusted" to express the probabilist's view of what sequences should be exorcised in a "good" model of randomness. For example, if $r = 0.01$ and $T = 100$ then $|2S - T| \geq 2$ defines "bad": the only "good" sequences are those for which $S = 49,\ 50$, or 51. The joint probability of these events is

$$Pr\ \{\text{"good"},\ T = 100\} = 2^{-100}\left[\binom{100}{49} + \binom{100}{50} + \binom{100}{51}\right],$$

which is quite small (~ 0.23565). As T increases, this number increases (monotonically) so that (for fixed r) Pr {"good"} attains any value less than 1 for suitably large T. If Pr {"good"} = 0.99 statisticians would say that there is 99% confidence that a bad sequence violating $|S/T - p| < r$ will not have "occurred". (Remember that this confidence is in the minds of the inhabitants of the $\mathcal{A}$-world, not of $\mathcal{R}$-world.)

The operational meaning of the theorem is that a *Bernoulli process is a pretty good model of* (intuitive) *randomness in the $\mathcal{A}$-world*; not a perfectly good one, however, especially not for short sequences, because of the need of stating the criterion for a "good model" in terms of a T large enough to produce the desired confidence level.

These statements are rigorously correct in the $\mathcal{A}$-world but in the $\mathcal{R}$-world their plausibility rests solely on the belief that actual sequences have the same probabilities and the same independence properties as was assumed in the $\mathcal{A}$-world. Since RANDOM sequences constructed in the previous section certainly do not satisfy the i.i.d. assumption beyond subsequences of length r (where $b^r \sim T$), the standard version of the "law of large numbers" as just discussed does not apply to such sequences.

But what does Nature actually do?

Consider the sequence of discrete probability distributions for the binary digits of k/q, any $0 < k < q$, q the 2-ergodic prime $p = 4093$ ($\sim 4098 = 2^{12}$). (In each graph, the discrete points of the distribution have been connected by straight lines for better visualization.) The graph in Figure 5.2 looks quite smooth for $T = 10$ (because the binary subsequences of length $r \leq 12$ are almost exactly independent). But for large T the graphs Figure 5.3-5 develop a "fractal" appearance, while those for the i.i.d. Bernoulli sequence are smooth for all T. This shows clearly that the smoothness of the distributions as a function of T is directly linked to the abstract i.i.d. assumption and it is not to be expected in the $\mathcal{R}$-world.

The awkwardness of confidence intervals disappears; these distributions have no tails (because sequences of length T having only 0's or only 1's must satisfy $2^T < p$).

6. Conclusions

Whatever (intuitive, formalized, abstract) probability might be, and however the axioms might be phrased to hide the underlying assumptions, physical probability has meaning only in relation to specific systems. It is NOT a universal concept.

This does not mean that mainstream applications of probability – which are based on finding conditions under which happenings in the $\mathcal{R}$-world are believed to mimic the state of affairs in the $\mathcal{A}$-world – should be abandoned. It does mean that we should take a new look at all the probabilistic and statistical analysis to see what they really amount to in the $\mathcal{R}$-world.

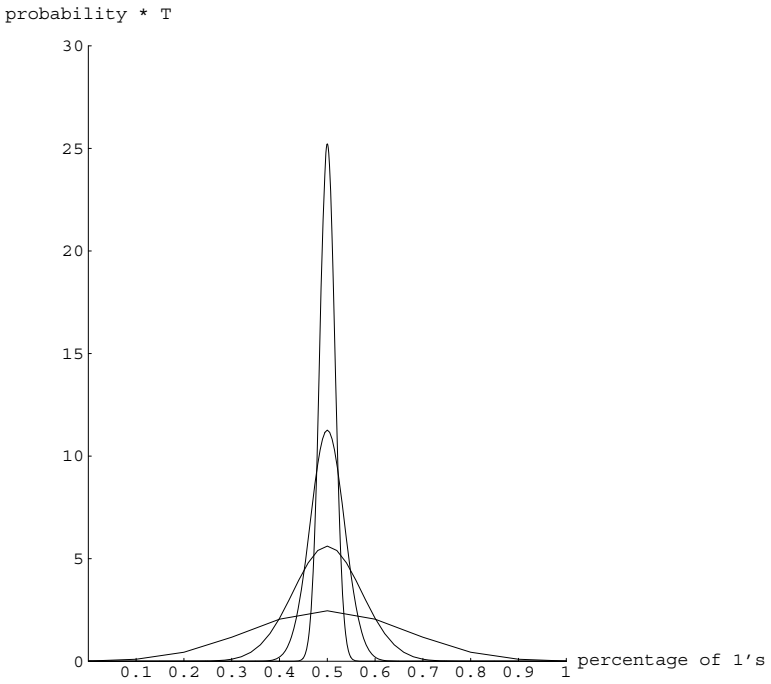

FIGURE 5.1
The (discrete) binomial distribution for $T = 10,\ 50,\ 200,\ 1000$.

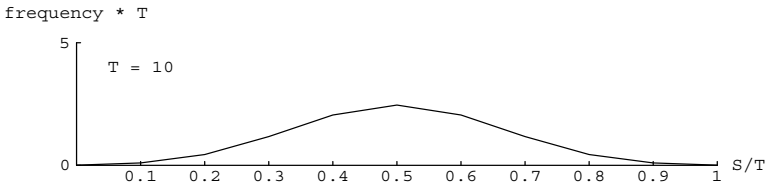

FIGURE 5.2
(Discrete) frequency of sequences of length $T = 10$ in the binary expansion of $\frac{1}{4093}$ (2-ergodic prime) for a fixed number S of 1's, according to S/T = percentage of 1's

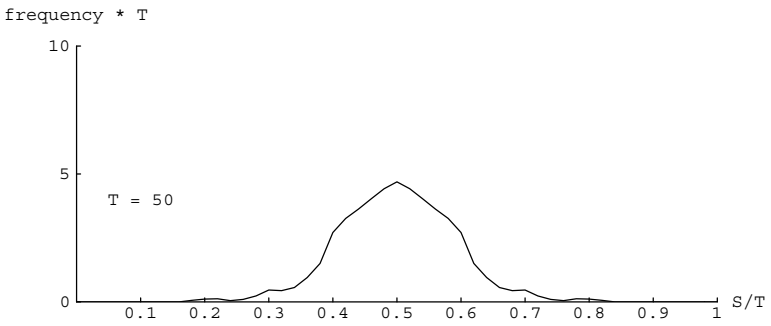

FIGURE 5.3
(Discrete) frequency of sequences of length $T = 50$ in the binary expansion of $\frac{1}{4093}$ (2-ergodic prime) for a fixed number S of 1's, according to S/T = percentage of 1's

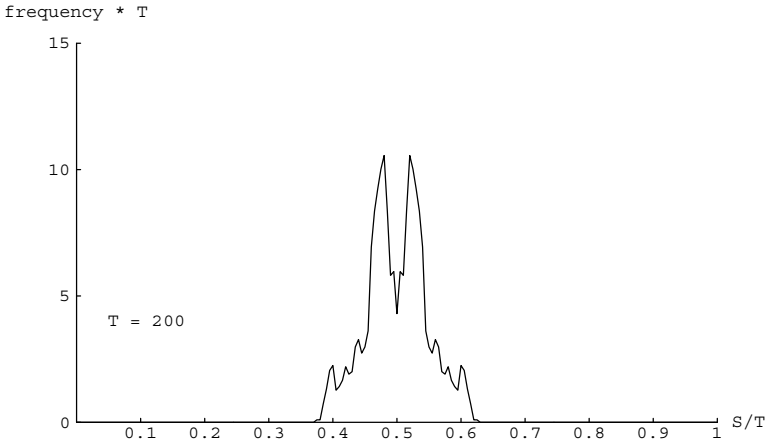

FIGURE 5.4
(Discrete) frequency of sequences of length $T = 200$ in the binary expansion of $\frac{1}{4093}$ (2-ergodic prime) for a fixed number S of 1's, according to S/T = percentage of 1's

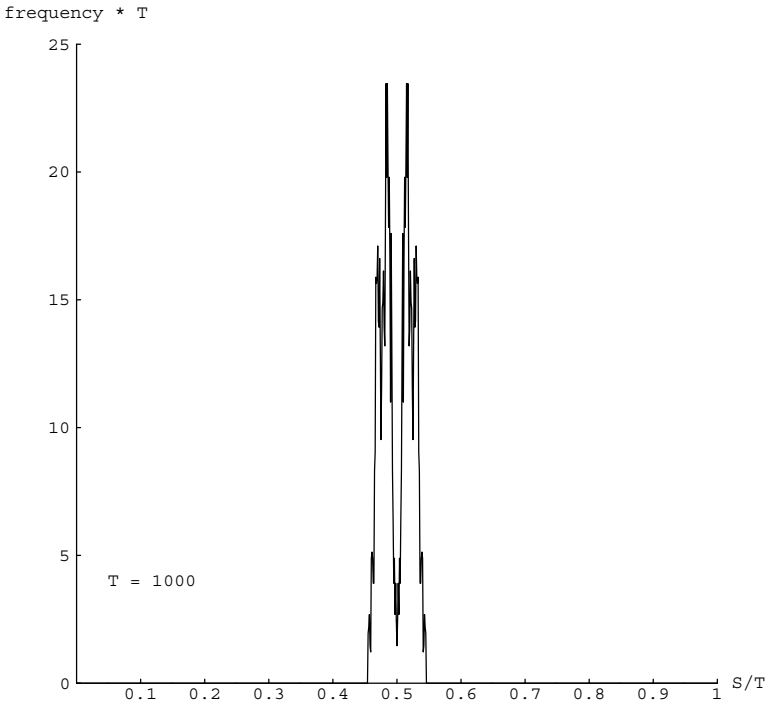

FIGURE 5.5
(Discrete) frequency of sequences of length $T = 1000$ in the binary expansion of $\frac{1}{4093}$ (2-ergodic prime) for a fixed number S of 1's, according to S/T = percentage of 1's

References

1. H. Atlan (1979). Postulats métaphysiques et méthodes de recherche, in *La Querelle du Déterminisme* (book), organized by Krysztof Pomian, Gallimard, Paris 1990.
2. E. Borel (1909). Les probabilités dénombrables et leurs applications arithmétiques, *Rendiconti del circolo matematico di Palermo*, **27**: 247–271.
3. C. Chatfield (1995). Personal communication to the author (letter of June 29, 1995).
4. P. Diaconis (1996). The cutoff phenomenon in finite Markov chains, *Proceedings of the National Academy of Sciences (USA),* **93**: 1659–1664.
5. W. Feller (1950). *An Introduction to Probability Theory and its Applications*, volume 1 (first edition), Wiley, 419 pages.
6. G. H. Hardy and E. M. Wright (1979). *An Introduction to the Theory of Numbers* (5th edition),Clarendon Press, Oxford, 426 pages.
7. L. K. Hua (1982). *Introduction to Number Theory*, Springer, 572 pages.
8. R. E. Kalman (1994). Randomness reexamined, *Modeling, Identification and Control* **15** : 141–151.
9. R. E. Kalman (1995). Randomness and probability, *Mathematica Japonica* **41** : 41–58.
10. D. V. Lindley (1987). The probability approach to the treatment of uncertainty in artificial intelligence and expert systems, *Statistical Science* **2**: 17–24.
11. S. Pincus (1991). Approximate entropy as a measure of system complexity, *Proceedings of the National Academy of Sciences (USA),* **88**: 2297–2301.
12. S. Pincus and B. H. Singer (1996). Randomness and degrees of irregularity, *Proceedings of the National Academy of Sciences (USA)* **93**: 2083–2088.
13. R. Thom (1986). Préface à Pierre-Simon (de) Laplace, *Essai Philosophique sur les Probabilités*, réedition de la texte de la 5^e édition (1825), Christian Bourgois, Paris, 1986.
14. H. Weyl (1916). Über die Gleichverteilung von Zahlen mod. 1, *Mathematische Annalen* **77**: 313–352.

Remarks

1. Indeed, Lindley's advertisement in praise of probability has a formal similarity with "there is only one God and X is its prophet". This seems to be relevant to distinguishing between "religious" and "scientific" intellectual attitudes but, of course, should not be considered as derogatory in regard to any religion or person.
2. *No* typing error; these are the first 60 digits of the binary expansion of the fraction 29/61. See Table 2.
3. The phrase "Bernoulli sequence" as used here has nothing to do with an objective property of a sequence of 0's and 1's (because any such sequence is allowed) but with the hypothesis that the sequence was generated in the $\mathcal{A}$-world, by tossing a fair coin. ,

4. Probability does not help much with uncertainty, either. Unless, like, apparently, Lindley, we wish to live, in the $\mathcal{A}$-world, incommunicado with the $\mathcal{R}$-world. By Borel's theorem we are quite "sure" (if we are in the $\mathcal{A}$-world, then sure with probability 1) that the one millionth digit of $\sqrt{2}-1$ is a 5 with probability $\frac{1}{10}$ or a 0 also with a probability $\frac{1}{10}$. But this does not provide much "information", simply because Borel's theorem has nothing to do with the given number whose one millionth digit we wish to know. On the other hand, it is possible to compute the one millionth digit of $\sqrt{2}-1$ (this is an important subspecialty of computational science), so there is no uncertainty about the value of this digit in the $\mathcal{R}$-world.
5. When "random" is used in the technical sense, corresponding to our definition, we write RANDOM for special emphasis.
6. To see what can be done about randomness without any direct appeal to probability, please consult PINCUS [11].
7. It is quite possible that some number $kt \pmod 1$ *has* a common factor with q but this is irrelevant to the problem formulation which works for all $\alpha =$ rational. So, the requirement $q =$ prime is irrelevant; "equidistribution", is assured by the ergodic property valid for rational as well as irrational α. In the rational case "ergodic" means that ρ acts on the *whole* set Z_{q-1}^{+} and is not invariant on some proper subset; in the irrational case "ergodic" means that there is no invariant interval J under the action of t other than the whole interval $(0,1)$.
8. Please note also that when $\rho =$ identity permutation, that is, $k=1$ and $\alpha = 1/q$, the sequence of points $1/q,\ 2/q,\ \ldots,\ (q-1)/q$ in the unit interval corresponding to $t = 1,\ 2,\ \ldots,\ q-1$ very much satisfies the claim "as equidistributed as possible", even though these points "arrive regularly" in X_{q-1}.
9. We may be permitted to wonder if Feller had experimented in binary arithmetic, for example, dividing 1 by a large 2-ergodic prime. .
10. And we may surmise that the man in the street will have reciprocated by accusing (an abstract mathematician like) Feller of not noticing a loaded coin when he sees it.
11. This theorem shall not be proved here; the proof is similar to the proof of $\pi_5(2^l) = 2^{l-2}, l \geq 2$, given in HUA [7, Theorem 3.9.2, p. 50]. Note that $\pi_5(2) = \pi_5(4) = 1$; this fact is analogous to $\pi_{10}(3) = \pi_{10}(9) = 1$.)
12. If $0 \leq m_{k/81} < 10$ is the missing digit in the decimal expansion of $k/81$ (period $= 9$) then $m_{k/81} + k = 0 \pmod 9$.
13. No, *not* clear. Not even true. If we were to take Feller's opinion literally, there would never have been any Western science, no Copernicus, no Galileo, no Newton, perhaps no Feller.