Statistical Independence and Normal Numbers: An Aftermath to Mark Kac's Carus Monograph

Gerald S. Goodman

1. INTRODUCTION. In 1958, Mark Kac delivered the prestigious Philips Lectures at Haverford College, which formed the basis for his Carus Monograph on Statistical Independence [**2**]. As a student, I had the privilege of attending those lectures. By foraging in their *aftermath*, meaning literally, "that which grows after the harvest," I have been led to the following ideas. They may be regarded as comprising a long footnote, or possibly a missing chapter, of that Monograph, composed in memory of its author by one of his erstwhile "guinea pigs."

In [**2**], Kac presents an arithmetical model for coin-tossing, due to Borel [**1**], which culminates in an analytical proof of the Strong Law of Large Numbers for Bernoulli trials. The idea is to identify the digits occurring in the binary expansion of a point $\omega \in \Omega = [0,1)$,

$$\omega = \sum_{j=1}^{\infty} \frac{a_j(\omega)}{2^j}, \tag{1.1}$$

with the outcomes of a fair coin-tossing experiment. The j-th digit, $a_j(\omega)$, equals 1 if *tails* occurs on the j-th toss, and equals 0 if the outcome is *heads*. Then

$$S_n(\omega) = a_1(\omega) + a_2(\omega) + \cdots + a_n(\omega) \tag{1.2}$$

is the total number of tails in the first n tosses, and Borel's Strong Law of Large Numbers asserts that, when Ω is endowed with Lebesgue measure,

$$\frac{S_n(\omega)}{n} \to \frac{1}{2} \text{ a.e., as } n \to \infty. \tag{1.3}$$

Kac's proof of Borel's theorem makes use of the *Rademacher functions*. These are defined, for each $j \in \mathbb{N}$ and $\omega \in \Omega$, as

$$r_j(\omega) = 1 - 2a_j(\omega), \tag{1.4}$$

and they represent the net gain of a gambler who bets one florin on the outcome *heads* at the j-th toss. The orthonormality of the Rademacher functions on the unit interval Ω, and, indeed, of their finite products, is a consequence of the fact that the binary digits are statistically independent. It is this orthonormality that Kac exploits to give a simple proof of the Law of Large Numbers.

Kac goes on to interpret this result in terms of the existence of *simply normal numbers*. For such a number, the relative frequency of the 1's, among the first n digits in the binary expansion of its fractional part, tends to $1/2$ as $n \to \infty$. It follows from Borel's theorem that *almost all real numbers* have this property. Kac points out that a similar result holds for any integral base $b > 1$, sketches a proof, and concludes, with Borel, that almost all numbers are simply normal to every such base.

Now, it is known [**8**] that any number that is simply normal to all bases that are positive powers of an integral base b has the following property: Every string of finitely many b-adic digits occurs in the base b expansion of its fractional part with asymptotic relative frequency $1/b^m$, where m is the length of the string. Partial overlaps of the string with itself are counted as distinct occurrences. Numbers exhibiting this property are termed *normal to base b*, and the fact that almost every real number has this property is known as *The Normal Number Theorem* for b-adic digits.

Since Kac's use of orthonormality leads to such a simple proof of the case $m = 1, b = 2$, it is natural to ask whether his method can be adapted to give a direct proof of the Normal Number Theorem, at least for strings of binary digits. It turns out that this can be done by replacing Rademacher functions by their finite products, which themselves form an orthonormal system known as the *Walsh functions*. Kac presents them in an exercise, without any mention of their possible connection with normal numbers [**2**, p. 11].

That discovery was made by Mendès-France [**6**], who showed how base 2 normality can be characterized in terms of the Walsh functions. He then used this characterization to prove the Normal Number Theorem for binary digits. His approach, which, as he showed, generalizes to arbitrary integral bases $b > 1$, makes use of Haar measure, group characters, a generalized Weyl Criterion for asymptotically equidistributed sequences, and Birkhoff's Ergodic Theorem, applied to the dyadic map.

We shall show, instead, how the orthonormality of the Walsh functions leads directly to a simple proof, *a là* Kac, of the Normal Number Theorem for binary digits. The only additional idea required is the well-known observation, due to Wall [**10**] and used by Mendès-France, that a number is normal to base 2 if and only if the iterates, under the dyadic map, of its fractional part are uniformly distributed in the unit interval. Using this idea, our proof proceeds in the same spirit as one suggested by Kac—again, in the form of an exercise—for proving the classical theorem of Weyl [**11**] on the equidistribution of the fractional parts of multiples of irrational numbers [**2**, p. 41].

We can extend this approach to integral bases $b > 1$ by defining, with Mendès-France [**6**], the *b-adic Rademacher functions* as

$$r_j(\omega) = \exp\left(\frac{2\pi i\, b_j(\omega)}{b}\right), \text{ for all } \omega \in \Omega,\ j \in \mathbb{N},$$

where the b_j are the b-adic coefficients of the point ω,

$$\omega = \sum_{j=1}^{\infty} \frac{b_j(\omega)}{b^j},$$

and defining the b-adic Walsh functions as their power products. When $b = 2$, $b_j \in \{0, 1\}$ and Euler's formula ensures that the new definition of r_j agrees with the old one.

These b-adic Rademacher functions have mean zero, and all that is needed to establish their orthogonality (over the complex field) and further multiplicativity properties is the formula

$$\int_0^1 \prod_{j=1}^{k} \exp\big(i\mu_j b_j(\omega)\big)\, d\omega = \prod_{j=1}^{k} \int_0^1 \exp\big(i\mu_j b_j(\omega)\big)\, d\omega,$$

for suitable values of the real parameters μ_j and arbitrary values of $k \in \mathbb{N}$. But the validity of such a formula is guaranteed by the statistical independence of the

b_j's, as Kac well knew [**3**], and its proof is virtually the same as one found in the opening pages of [**2**].

In Section 8, we give the details, and show how the reasoning used to establish the base 2 normality of almost every real number goes through in the case of general integral bases b. Once done, it remains only to collect the exceptional sets in order to arrive at a new proof of The Normal Number Theorem, in its full force: *almost every real number is normal to every base*.

We then go on to examine more carefully the connection between the multiplicativity of the b-adic Rademacher functions, understood as the vanishing of their mixed moments, and the statistical independence of the b-adic coefficients. Using a device of Rényi's [**9**] we are able to draw a remarkable conclusion—the two notions are entirely equivalent! Since the multiplicativity property can be established easily by elementary analysis, this yields a new proof of the independence of the b-adic coefficients.

The same ideas can be applied to b-adic Walsh functions. While they themselves are not statistically independent, we find that the ones whose indices form a geometrical progression made up of fixed multiples of powers of b are. The statistical independence of the b-adic Rademacher functions, and thus of the b-adic coefficients, follows as a special case.

2. THE RADEMACHER FUNCTIONS. The functions r_j are defined by (1.4). Since the binary coefficients a_j satisfy

$$\int_0^1 a_j(\omega)\, d\omega = \frac{1}{2} \text{ for } j \in \mathbb{N},$$

it follows that

$$\int_0^1 r_j(\omega)\, d\omega = 0 \text{ for } j \in \mathbb{N}.$$

The statistical independence of the a_j implies that the r_j are also independent, and, since they each have mean zero, they satisfy the *Multiplicativity Formula*

$$\int_0^1 r_{j_1}(\omega) r_{j_2}(\omega) \cdots r_{j_k}(\omega)\, d\omega = 0,\ k \in \mathbb{N}, \tag{2.1}$$

whenever the subscripts are distinct. Since $r_j^2 = 1$ for every j, it follows from (2.1) that the Rademacher functions form an *orthonormal system* on Ω, that is,

$$\int_0^1 r_j(\omega) r_k(\omega)\, d\omega = \delta_{jk} \text{ for } j, k \in \mathbb{N}. \tag{2.2}$$

3. KAC'S PROOF [2], [3]. In view of (1.4), the limiting relation (1.3) is equivalent to the assertion that

$$\frac{R_n(\omega)}{n} \to 0 \text{ a.e., as } n \to \infty, \tag{3.1}$$

where

$$R_n(\omega) = r_1(\omega) + r_2(\omega) + \cdots + r_n(\omega) \tag{3.2}$$

is the difference between the number of heads and the number of tails in the first n tosses.

A direct computation using (2.1) and (2.2) shows that

$$\int_0^1 [R_n(\omega)]^4 \, d\omega = n + \frac{4!}{2!2!}\binom{n}{2} = n + 3n(n-1). \tag{3.3}$$

Consequently,

$$\sum_{n=1}^{\infty} \int_0^1 \left[\frac{R_n(\omega)}{n}\right]^4 d\omega < \infty,$$

and it follows from Beppo Levi's Theorem that

$$\sum_{n=1}^{\infty} \left[\frac{R_n(\omega)}{n}\right]^4 d\omega < \infty \text{ a.e.}$$

This implies that

$$\left[\frac{R_n(\omega)}{n}\right]^4 \to 0 \text{ a.e., as } n \to \infty,$$

which is clearly equivalent to (3.1).

4. A VARIANT OF KAC'S PROOF. The preceding proof uses (2.1) up to fourfold products, which is evidently a stronger property than the mere orthonormality of the Rademacher functions expressed by (2.2). However, a proof of (3.1) that uses only the orthonormality of the r_j (along with the uniform boundedness of their absolute values) can be based upon an argument employed by H. Weyl [**11**] in a similar context.

The orthonormality (2.2) implies that, for each $n \in \mathbb{N}$,

$$\int_0^1 [R_n(\omega)]^2 \, d\omega = n.$$

It follows that

$$\sum_{n=1}^{\infty} \int_0^1 \left[\frac{R_{n^2}(\omega)}{n^2}\right]^2 d\omega < \infty,$$

and, therefore, reasoning as before,

$$\frac{R_{n^2}(\omega)}{n^2} \to 0 \text{ a.e., as } n \to \infty. \tag{4.1}$$

Now, to each value of n, not a perfect square, there is a unique positive integer m_n such that $m_n^2 < n < (m_n + 1)^2$. Clearly, $m_n \to \infty$ as $n \to \infty$. In view of (3.2) and the fact that $|r_j| = 1$ for all j,

$$|R_n| \le |R_{m_n^2}| + |r_{m_n^2+1}| + \cdots + |r_n| \le |R_{m_n^2}| + 2m_n.$$

Dividing by m_n^2 and using (4.1) with n^2 replaced by m_n^2 yields

$$\frac{|R_n(\omega)|}{m_n^2} \to 0 \text{ a.e., as } n \to \infty.$$

Since $m_n^2 < n$, (3.1) follows.

5. SHIFT DYNAMICS. Define the *binary shift* (or *dyadic map*) T on Ω by the formula [**2**, p. 19 and p. 93]

$$T(\omega) = 2\omega \pmod 1. \tag{5.1}$$

Then, when ω is expressed in terms of its binary expansion (1.1), $T(\omega)$ takes the form

$$T(\omega) = \sum_{j=1}^{\infty} \frac{a_j \circ T(\omega)}{2^j} = \sum_{j=1}^{\infty} \frac{a_{j+1}(\omega)}{2^j}.$$

Making the convention that binary rationals have finite expansions, the uniqueness of the binary coefficients gives $a_{j+1}(\omega) = a_j \circ T(\omega)$ for $j \in \mathbb{N}$. Iteration yields

$$a_{j+k}(\omega) = a_j \circ T^k(\omega) \text{ for } j \in \mathbb{N},\, k \in \mathbb{N}_0, \tag{5.2}$$

and, in particular,

$$a_{1+k}(\omega) = a_1 \circ T^k(\omega) \text{ for } k \in \mathbb{N}_0. \tag{5.3}$$

Now, a_1 is the indicator function of the interval $[\frac{1}{2}, 1)$. In view of (5.3) and the definition (1.2) of S_n,

$$S_n(\omega) = a_1(\omega) + a_1 \circ T(\omega) + \cdots + a_1 \circ T^{n-1}(\omega).$$

Thus, $S_n(\omega)$ counts the number of times that the *orbit* of ω under T, that is,

$$\omega, T(\omega), \dots, T^k(\omega), \dots,$$

is found in $[\frac{1}{2}, 1)$ during the first n steps, starting from step 0. The Strong Law of Large Numbers can therefore be interpreted as saying that *the relative time that the orbit of* ω *occupies* $[1/2, 1)$ *tends to* $1/2$ *as* $n \to \infty$, for almost all starting points $\omega \in \Omega$, in accordance with [**2**, p. 93].

This dynamical interpretation of the Law of Large Numbers leads to a formulation of the Normal Number Theorem in an analogous way. Let

$$\alpha_1, \alpha_2, \dots, \alpha_m \tag{5.4}$$

denote a string of binary digits of length m. Set

$$l = \sum_{j=1}^{m} \frac{\alpha_j}{2^{j-m}},$$

and let

$$I_l = \left[\frac{l}{2^m}, \frac{l+1}{2^m}\right). \tag{5.5}$$

Then I_l is the l-th *binary interval of order* m, and it consists of those points in Ω whose binary expansion starts out with (5.4). With this enumeration, the lexicographic order of the strings (5.4) is expressed by the natural order of the intervals (5.5), in terms of increasing values of l.

If we now denote by χ_{I_l} the indicator function of the interval I_l,

$$\frac{1}{n} \sum_{k=0}^{n-1} \chi_{I_l} \circ T^k(\omega)$$

is the average number of times the string (5.4) occurs among the first $n + m - 1$ binary coefficients of ω, where overlappings count as multiple occurrences, and an occurrence is marked at the moment when the string starts to appear. Any real number whose fractional part is ω will then be normal if, for each $m \in \mathbb{N}$,

$$\frac{1}{n} \sum_{k=0}^{n-1} \chi_{I_l} \circ T^k(\omega) \to \frac{1}{2^m}, \text{ as } n \to \infty, \tag{5.6}$$

where χ_{I_l} is the indicator function of a generic binary interval I_l of order m, and $1/2^m$ is its length. The Normal Number Theorem will, accordingly, be a consequence of the assertion that (5.6) holds a.e. on Ω for each such binary interval I_l.

6. THE WALSH FUNCTIONS. Kac introduces the Walsh functions, as follows. For each $j \in \mathbb{N}$, let

$$2j = 2^{j_1} + 2^{j_2} + \cdots + 2^{j_s}, \tag{6.1}$$

where $j_1 < j_2 < \cdots < j_s$ are in $\mathbb{N}$, be the unique expansion of the integer $2j$ in base 2. Then, for each $j \in \mathbb{N}$, define the *Walsh functions* by means of the formula

$$\omega_j(\omega) = r_{j_1}(\omega) r_{j_2}(\omega) \cdots r_{j_s}(\omega) \text{ for } \omega \in \Omega, \tag{6.2}$$

and set $w_0 \equiv 1$. The Walsh functions thus constitute an enumeration of all the finite products of Rademacher functions.

The Walsh functions, like the Rademacher functions, have a probabilistic interpretation in terms of coin tossing. For each $j \in \mathbb{N}$, w_j represents the net gain of the gambler who bets one florin on the outcome that the total number of *tails* occurring on the tosses $j_1, j_2, \ldots, j_s$ is *even.*

Because of (2.1), the Walsh functions are orthogonal on Ω, while the identity $|w_j^2| \equiv 1$ implies that they are orthonormal. For each $m \in \mathbb{N}$ the 2^m functions $w_0, w_1, \ldots, w_{2^m - 1}$ are constant on binary intervals of order m. The range of any such function thus corresponds to a vector whose l-th component is the value taken by the function on the interval I_l. The orthogonality of the functions makes these vectors orthogonal and therefore linearly independent. It follows that the functions themselves are linearly independent. Consequently, any real-valued function f that is constant on binary intervals of order m can be written as a linear combination of the first 2^m Walsh functions:

$$f(\omega) = \sum_{j=0}^{2^m - 1} \lambda_j w_j(\omega), \ \omega \in \Omega, \tag{6.3}$$

where the weights λ_j are the *Fourier-Walsh* coefficients of f:

$$\lambda_j = \int_0^1 f(\omega)\, \omega_j(\omega)\, d\omega, \ j = 0, \ldots, 2^m - 1.$$

In particular, the function χ_{I_l} can be written in such a form. In this case,

$$\lambda_0 = \int_0^1 \chi_{I_l}(\omega) w_0(\omega)\, d\omega = \frac{1}{2^m}, \tag{6.4}$$

while

$$\lambda_j = \int_0^1 \chi_{I_l}(\omega) w_j(\omega)\, d\omega = \pm \frac{1}{2^m}, \ j = 1, \ldots, 2^m - 1,$$

where the value of the $\pm$ sign is such as to make $\operatorname{sgn}[\lambda_j w_j] = +1$ on I_l.

7. PROOF OF THE NORMAL NUMBER THEOREM FOR BINARY DIGITS. In order to prove that (5.6) holds a.e. for any binary interval of the form (5.5) with $m \in \mathbb{N}$ arbitrary, observe that (1.4) and (5.2) imply that

$$r_{j+k}(\omega) = r_j \circ T^k(\omega) \text{ for all } \omega \in \Omega, j \in \mathbb{N}, k \in \mathbb{N}_0, \tag{7.1}$$

so that, in view of (6.1) and (6.2),

$$w_{2^k j}(\omega) = w_j \circ T^k(\omega) \text{ for all } \omega \in \Omega,\ j, k \in \mathbb{N}_0. \tag{7.2}$$

It is easy to verify that, for each $j \in \mathbb{N}$, the subset $w_j, w_{2j}, \ldots, w_{2^k j}, \ldots$ forms a multiplicative, orthonormal system, so the reasoning of the previous sections, applied to the sums $w_j(\omega) + w_{2j}(\omega) + \cdots + w_{2^{n-1}j}(\omega)$, instead of to the R_n, shows that, for each fixed $j \in \mathbb{N}$,

$$\frac{1}{n} \sum_{k=0}^{n-1} w_j \circ T^k(\omega) \to 0 \text{ a.e. as } n \to \infty,$$

while, trivially,

$$\frac{1}{n} \sum_{k=0}^{n-1} w_0 \circ T^k(\omega) \to 1 \text{ as } n \to \infty.$$

Consequently, using (6.3) with f replaced by χ_{I_l},

$$\frac{1}{n} \sum_{k=0}^{n-1} \chi_{I_l} \circ T^k(\omega) = \frac{1}{n} \sum_{k=0}^{n-1} \sum_{j=0}^{2^m-1} \lambda_j w_j \circ T^k(\omega)$$

$$= \sum_{j=0}^{2^m-1} \lambda_j \left(\frac{1}{n} \sum_{k=0}^{n-1} w_j \circ T^k(\omega) \right) \to \lambda_0 \text{ a.e. as } n \to \infty.$$

Since, by (6.4), $\lambda_0 = 1/2^m$, (5.6) holds a.e., and the theorem is proved.

8. GENERALIZATION TO *b*-ADIC DIGITS. Suppose that $b > 1$ is any integral base. Let the b-adic expansion of a generic point $\omega \in \Omega$ be

$$\omega = \sum_{j=1}^{\infty} \frac{b_j(\omega)}{b^j}, \tag{8.1}$$

with the convention that b-adic rationals have finite expansions. Following **[6]**, define the *b-adic Rademacher functions*

$$r_j(\omega) = \exp\left(\frac{2\pi i\, b_j(\omega)}{b} \right) \text{ for all } \omega \in \Omega,\ j \in \mathbb{N}. \tag{8.2}$$

Thus, the r_j assume values in the cyclotomic group of b-th roots of unity, and each root is taken on a subset of Ω having measure $1/b$.

It follows at once that, for each $j \in \mathbb{N}$ and any integer d,

$$\int_0^1 r_j^d(\omega)\, d\omega = 0, \tag{8.3}$$

unless $d \equiv 0 \bmod b$. It is also evident that, when d is in the range from 0 to $b - 1$,

$$\overline{r_j^d(\omega)} = r_j^{b-d}(\omega) \text{ for all } \omega \in \Omega,\ j \in \mathbb{N},$$

so the complex conjugates of powers of the r_j are expressible as positive powers of r_j. Accordingly, (8.3) holds also when r_j is replaced by its complex conjugate, as could have been seen directly.

As stated in the Introduction, we shall make use of the formula

$$\int_0^1 \prod_{j=1}^{k} \exp\big(i\mu_j b_j(\omega)\big)\, d\omega = \prod_{j=1}^{k} \int_0^1 \exp\big(i\mu_j b_j(\omega)\big)\, d\omega \tag{8.4}$$

for certain values of the real parameters μ_j and arbitrary $k \in \mathbb{N}$. The left-hand integral is one form of the multidimensional characteristic function (Fourier transform) of the first k b-adic coefficients, considered as random variables on Ω [**2**, p. 42], and it is their statistical independence that assures the equality of the two expressions. An elementary proof can be based upon Kac's demonstration of an analogous formula [**2**, p. 7].

Indeed, let $\mathbf{P}$ denote Lebesgue measure on Ω. Then

$$\int_0^1 \prod_{j=1}^{k} \exp\left(i\mu_j b_j(\omega)\right) d\omega$$

$$= \sum_{\beta_1,\ldots,\beta_k} \prod_{j=1}^{k} \exp\left(i\mu_j \beta_j\right) \mathbf{P}\{b_1(\omega) = \beta_1, \ldots, b_k(\omega) = \beta_k\},$$

where the β_j range independently over the set $0, 1, \ldots, b-1$. Using the statistical independence of the b_j's, the last expression becomes

$$\sum_{\beta_1,\ldots,\beta_k} \prod_{j=1}^{k} \exp\left(i\mu_j \beta_j\right) \prod_{j=1}^{k} \mathbf{P}\{b_j(\omega) = \beta_j\} = \sum_{\beta_1,\ldots,\beta_k} \prod_{j=1}^{k} \exp\left(i\mu_j \beta_j\right) \mathbf{P}\{b_j(\omega) = \beta_j\},$$

which, in turn, reduces to

$$\prod_{j=1}^{k} \sum_{\beta_j=0}^{b-1} \exp\left(i\mu_j \beta_j\right) \mathbf{P}\{b_j(\omega) = \beta_j\} = \prod_{j=1}^{n} \int_0^1 \exp\left(i\mu_j b_j(\omega)\right) d\omega,$$

as required.

Setting $\mu_j = 2\pi d_j/b$ in (8.4), where the d_j are integers between 0 and $b-1$, yields, in view of (8.3) and the defining equation (8.2), the *Multiplicativity Formula for the b-adic Rademacher functions*,

$$\int_0^1 r_1^{d_1}(\omega) r_2^{d_2}(\omega) \cdots r_k^{d_k}(\omega)\, d\omega = 0 \text{ for all } k \in \mathbb{N}, \tag{8.5}$$

unless all of the d's vanish, in which case the integral has the value 1. Recalling our comments about complex conjugates, we see that (8.5) continues to hold when any number of factors in the integrand are replaced by their conjugates. When $b = 2$, (2.1) can also be expressed in the present form, when k is suitably chosen and the exponents of the selected factors are set equal to 1, while the other exponents all vanish.

To generalize the Walsh functions, take any $j \in \mathbb{N}$ and make the partition

$$bj = d_{j_1} b^{j_1} + d_{j_2} b^{j_2} + \cdots + d_{j_s} b^{j_s}, \tag{8.6}$$

where the d's are integers in the range $1, \ldots, b-1$. Then set $w_0 \equiv 1$ and define the *b-adic Walsh functions*

$$w_j(\omega) = r_1^{d_{j_1}}(\omega) r_{j_2}^{d_{j_2}}(\omega) \cdots r_{j_s}^{d_{j_s}}(\omega), \quad \omega \in \Omega,\ j \in \mathbb{N}. \tag{8.7}$$

This definition agrees with that of Mendès-France [**6**, p. 44].

By (8.2) and (8.7), $|w_j|^2 \equiv 1$ for all $j \in \mathbb{N}_0$. It then follows that the b-adic Walsh functions satisfy the orthonormality relations

$$\int_0^1 w_j(\omega) \overline{\omega_k(\omega)}\, d\omega = \delta_{jk} \text{ for } j, k \in \mathbb{N}_0, \tag{8.8}$$

for, by (8.5) and (8.7), the value of the integral can be 1 only if w_j and w_k are products of the same b-adic Rademacher functions, and their corresponding powers agree.

Instead of the dyadic map (5.1), consider the b-adic map $T(\omega) = b\omega \pmod 1$ and note that the uniqueness of the expansion (8.1) implies that $b_{j+k}(\omega) = b_j \circ T^k(\omega)$ for $j \in \mathbb{N}$, $k \in \mathbb{N}_0$. Thus, in view of (8.2), we also have

$$r_{j+k}(\omega) = r_j \circ T^k(\omega) \text{ for all } \omega \in \Omega,\ j \in \mathbb{N},\ k \in \mathbb{N}_0,$$

which is formally the same as (7.1). Consequently, by (8.7),

$$w_j \circ T^k(\omega) = r_{j_1+k}^{d_1}(\omega) r_{j_2+k}^{d_2}(\omega) \cdots r_{j_s+k}^{d_s}(\omega),\ \omega \in \Omega, \tag{8.9}$$

and so, in view of (8.7) and (8.6),

$$w_{b^k j}(\omega) = w_j \circ T^k(\omega) \text{ for all } \omega \in \Omega,\ j, k \in \mathbb{N}_0, \tag{8.10}$$

which generalizes (7.2).

Let m be any natural number, and take any string

$$\beta_1, \beta_2, \ldots, \beta_m \tag{8.11}$$

of m b-adic digits. Let I denote the b-adic interval of order m made up of those points in Ω whose b-adic expansion starts out with (8.11). Clearly, I has length $1/b^m$.

In view of (8.8), the first b^m b-adic Walsh functions, $w_0, w_1, \ldots, w_{b^m-1}$, are linearly independent over the complex numbers, and their span consists of all complex-valued functions that are constant on b-adic intervals of order m. In particular, χ_I can be written as

$$\chi_I(\omega) = \sum_{j=0}^{b^m-1} \lambda_j w_j(\omega),\ \omega \in \Omega, \tag{8.12}$$

where the weights λ_j are given by the formulas

$$\lambda_j = \int_0^1 \chi_I(\omega) \overline{w_j(\omega)}\, d\omega = \frac{\eta_j}{b^m},\ j = 1, \ldots, b^m - 1.$$

Here, η_j is a b-th root of unity for each j, and

$$\lambda_0 = \int_0^1 \chi_I(\omega) \overline{w_0(\omega)}\, d\omega = \frac{1}{b^m}. \tag{8.13}$$

It follows, again from (8.8), that, for each $j \in \mathbb{N}$, the functions

$$w_j, w_{bj}, \ldots, w_{b^k j}, \ldots \tag{8.14}$$

themselves form an orthonormal system on Ω. Accordingly, in view of (8.10), we can establish that, for each fixed $j \in \mathbb{N}$,

$$\frac{1}{n} \sum_{k=0}^{n-1} w_j \circ T^k(\omega) \to 0 \text{ a.e. as } n \to \infty, \tag{8.15}$$

by appealing to the reasoning of Section 4, provided we replace the second moments of the $R_n(\omega)$ there by the second moments of the *absolute values* of the sums

$$w_j(\omega) + w_{bj}(\omega) + \cdots + w_{b^{n-1}j}(\omega)$$

for $j = 1, \ldots, b^{m+1}$ and $n \in \mathbb{N}$. A more involved calculation, using the fact that, for each $k \in \mathbb{N}$,

$$w_{b^k j}(\omega) = r_{j_1+k}^{d_1}(\omega) r_{j_2+k}^{d_2}(\omega) \cdots r_{j_s+k}^{d_s}(\omega),$$

which follows from (8.9) and (8.10), and using, *mutatis mutandis*, the Multiplicativity Formula (8.5), shows that the functions (8.14) are multiplicative and that

$$\int_0^1 \left| w_j(\omega) + w_{bj}(\omega) + \cdots + w_{b^{n-1}j}(\omega) \right|^4 d\omega \le n + 3n(n-1),$$

so Kac's original argument of Section 3 can still be used.

Once again,

$$\frac{1}{n}\sum_{k=0}^{n-1} w_0 \circ T^k(\omega) \to 1 \text{ as } n \to \infty.$$

Consequently, because of (8.12) and (8.15),

$$\frac{1}{n}\sum_{k=0}^{n-1} \chi_I \circ T^k(\omega) \to \lambda_0 \text{ a.e., as } n \to \infty,$$

just as in Section 7. But, in view of (8.13), $\lambda_0 = 1/b^m$. This proves that the relative frequency with which the string (8.11) occurs in the b-adic expansion of the fractional part of almost every real number tends to $1/b^m$ as $n \to \infty$, and that is enough to establish the Normal Number Theorem.

9. INDEPENDENCE AND MULTIPLICATIVITY IN BASE 2. Returning now to base 2, observe that

$$\frac{1+\gamma\eta}{2} = \delta_{\gamma\eta}, \tag{9.1}$$

whenever $\gamma, \eta = \pm 1$, as noted by Rènyi [**9**, p. 130]. Consequently, for any $j \in \mathbb{N}$,

$$\frac{1+\gamma_j r_j}{2} = \chi_{\{r_j = \gamma_j\}}, \tag{9.2}$$

where r_j is the j-th Rademacher function and $\gamma_j = \pm 1$. Since

$$\prod_{j=1}^{m} \chi_{\{r_j=\gamma_j\}} = \chi_{\{r_1=\gamma_1,\ldots,r_m=\gamma_m\}},$$

it follows that for any $m \in \mathbb{N}$

$$\chi_{\{r_1=\gamma_1,\ldots,r_m=\gamma_m\}} = \prod_{j=1}^{m}\left(\frac{1+\gamma_j r_j}{2}\right). \tag{9.3}$$

Rényi uses this formula in the following way. He expands the product in (9.3) to get

$$\prod_{j=1}^{m}\left(\frac{1+\gamma_j r_j}{2}\right) = \frac{1}{2^m}\left(1 + \sum_{j=1}^{m}\gamma_j r_j + \sum_{1\le j<l\le m}\gamma_j\gamma_l r_j r_l + \cdots + \prod_{j=1}^{m}\gamma_j r_j\right). \tag{9.4}$$

Setting

$$\gamma_j = 1 - 2\alpha_j \text{ for } j = 1,\ldots,m, \text{ where } \alpha_j = 0 \text{ or } 1, \tag{9.5}$$

then gives, in view of (1.2) and (9.3), an expansion of the indicator function χ_{I_l} of a generic binary interval (5.5) of order m in terms of the first 2^m Walsh functions, as in our (6.3) *et seq*., but without any recourse to orthonormality.

Rényi's approach can be used to motivate the introduction of finite products of Rademacher functions, and, therefore, of the Walsh functions, to deal with problems involving strings. However, it is also possible to employ these formulas in another way to establish an unexpected connection between the multiplicativity of the Rademacher functions and their statistical independence.

Recall that, in Section 2, we used the statistical independence of the Rademacher functions, together with the vanishing of their means, to justify the Multiplicativity Formula (2.1). However, as noted by Kaczmarz and Steinhaus, this formula can also be established directly by calculus, without appealing to the notion of statistical independence [**4**, p. 125*f*].

To do so, note first that (7.1) implies that

$$r_j = r_1 \circ T^{j-1} \text{ for } j \in \mathbb{N},$$

where T is the binary shift given by (5.1). For fixed j, let l vary in the range from 0 to $2^{j-1} - 1$, and denote by I_l the l-th dyadic interval of order $j - 1$. Since T^{j-1} is linear on each such I_l and maps it onto Ω, while the derivative of T^{j-1} on I_l is 2^{j-1}, the change-of-variable formula gives

$$\int_{I_l} r_j(\omega)\, d\omega = \int_{I_l} r_1 \circ T^{j-1}(\omega)\, d\omega = \frac{1}{2^{j-1}} \int_0^1 r_1(\omega)\, d\omega = 0 \tag{9.6}$$

for each I_l. Summing on l then yields

$$\int_0^1 r_j(\omega)\, d\omega = 0 \text{ for } j \in \mathbb{N}, \tag{9.7}$$

as asserted at the start of Section 2.

To arrive at the Multiplicativity Formula (2.1), we can now employ an argument used by Kac in a different connection [**2**, p. 27]. Order the subscripts so that j_k is the largest, and then write

$$\int_0^1 r_{j_1}(\omega) r_{j_2}(\omega) \cdots r_{j_k}(\omega)\, d\omega = \sum_{l=0}^{2^{j_k-1}-1} \int_{I_l} r_{j_1}(\omega) r_{j_2}(\omega) \cdots r_{j_k}(\omega)\, d\omega,$$

where the I_l are the dyadic intervals of order $j_k - 1$. Since the first $k - 1$ functions $r_{j_1}, r_{j_2}, \ldots, r_{j_{k-1}}$ are *constant* on each such interval, while, by (9.6) with $j = j_k$, the integral of r_{j_k} over each one is zero, the integrals in the sum at right all vanish, and the Multiplicativity Formula follows.

With this result in hand, we can establish the statistical independence of the Rademacher functions by proceeding in the following way. We integrate (9.2) over Ω and use (9.7) to find that for every $j \in \mathbb{N}$

$$\mathbf{P}\{r_j = \gamma_j\} = \frac{1}{2} \text{ whenever } \gamma_j = \pm 1. \tag{9.8}$$

Integrating (9.3) over Ω gives

$$\mathbf{P}\{r_1 = \gamma_1, \ldots, r_m = \gamma_m\} = \int_0^1 \prod_{j=1}^m \left(\frac{1 + \gamma_j r_j}{2} \right).$$

Now integrating (9.4) and making use of the Multiplicativity Formula along with (9.8) gives

$$\mathbf{P}\{r_1 = \gamma_1, \ldots, r_m = \gamma_m\} = \frac{1}{2^m} = \prod_{j=1}^m \mathbf{P}\{r_j = \gamma_j\}. \tag{9.9}$$

Since $m \in \mathbb{N}$ is arbitrary, it follows that *multiplicativity of the Rademacher functions implies their statistical independence*.

The preceding reasoning is not limited to the case of Rademacher functions. As a further application, recall from Section 7 that the Walsh functions

$$w_j, w_{2j}, \dots, w_{2^k j}, \dots \tag{9.10}$$

are, for fixed $j \in \mathbb{N}$ and varying $k \in \mathbb{N}$, multiplicative. When $j = 1$, these functions reduce to the Rademacher functions, as is evident from (6.1) and (6.2). Noting that they have mean zero, we can repeat our argument and conclude that the functions (9.10) are, in fact, statistically independent for *every* fixed $j \in \mathbb{N}$.

If, finally, we make the substitution (9.5) in (9.9), and do the same in (9.8), then (9.9) becomes

$$\mathbf{P}\{a_1 = \alpha_1, \dots, a_m = \alpha_m\} = \frac{1}{2^m} = \prod_{j=1}^{m} \mathbf{P}\{a_j = \alpha_j\}.$$

Thus, *statistical independence of the binary digits is, itself, a direct consequence of multiplicativity of the Rademacher functions, and, therefore, the two notions are entirely equivalent.*

10. EXTENSION TO GENERAL BASES b. All the reasoning of the preceding section can be carried over to integral bases $b > 1$, once we find a suitable generalization of Rényi's identity (9.1). To this end, let γ and η be b-th roots of unity and consider the expression

$$\sum_{d=0}^{b-1} \gamma^{b-d}\eta^d = \sum_{d=0}^{b-1} \left(\frac{\eta}{\gamma}\right)^d.$$

Since $\gamma^b = \eta^b = 1$, the quotient η/γ is also a b-th root of unity, and, therefore, the right-hand sum vanishes unless $\gamma = \eta$. Consequently,

$$\frac{1 + \sum_{d=1}^{b-1} \gamma^{b-d}\eta^d}{b} = \gamma_{\gamma\eta},$$

and we have found the identity we need.

Proceeding as before, for each $j \in \mathbb{N}$ we get

$$\frac{1 + \sum_{d=1}^{b-1} \gamma_j^{b-d} r_j^d}{b} = \chi_{\{r_j = \gamma_j\}}, \tag{10.1}$$

and, since, by (8.3),

$$\int_0^1 r_j^d(\omega)\, d\omega = 0 \text{ for } j \in \mathbb{N} \text{ and } d = 1, \dots, b-1,$$

integrating (10.1) gives

$$\int_0^1 \chi_{\{r_j = \gamma_j\}} = \mathbf{P}\{r_j = \gamma_j\} = \frac{1}{b}. \tag{10.2}$$

This time,

$$\chi_{\{r_1 = \gamma_1, \dots, r_m = \gamma_m\}} = \prod_{j=1}^{m} \chi_{\{r_j = \gamma_j\}} = \prod_{j=1}^{m} \left(\frac{1 + \sum_{d=1}^{b-1} \gamma_j^{b-d} r_j^d}{b} \right). \tag{10.3}$$

Expanding the second product, we get an expression of the form

$$\frac{1}{b^m}(1 + \text{sums of weighted power-products of the } r_j), \tag{10.4}$$

where the weight assigned to each r_j^d is γ_j^{b-d}. Thus, integrating (10.3) over Ω and using the Multiplicativity Formula (8.5) for the b-adic Rademacher functions, yields, in view of (10.2),

$$\mathbf{P}\{r_1 = \gamma_1, \ldots, r_m = \gamma_m\} = \frac{1}{b^m} = \prod_{j=1}^{m} \mathbf{P}\{r_j = \gamma_j\}. \tag{10.5}$$

Accordingly, *multiplicativity of the b-adic Rademacher functions implies their statistical independence*, just as in the *case* $b = 2$.

In Section 8, we introduced subsets of b-adic Walsh functions of the form

$$w_j, w_{bj}, \ldots, w_{b^k j}, \ldots, \tag{10.6}$$

for fixed values of $j \in \mathbb{N}$ and varying $k \in \mathbb{N}$, and pointed out that they were multiplicative, in the sense that

$$\int_0^1 w_{b^{j_1} j}(\omega) w_{b^{j_2} j}(\omega) \cdots w_{b^{j_k} j}(\omega)\, d\omega = 0$$

whenever the exponents are distinct. We can now show that they are statistically independent for each $j \in \mathbb{N}$ by extending this formula in the following way. Let b' be the smallest positive integer such that $w_j^{b'} \equiv 1$. Then, as a consequence of (8.5) and the choice of b', the functions (10.6) are multiplicative in the extended sense that

$$\int_0^1 w_j^{d_1'}(\omega) w_{bj}^{d_2'}(\omega) \cdots w_{b^k j}^{b_k'}(\omega)\, d\omega = 0$$

for every $k \in \mathbb{N}$, where the d's are restricted to the values $0, 1, \ldots, b' - 1$ and do not all vanish. It is thus possible to apply the foregoing reasoning to them, by replacing b by b' in the formulas above. This yields the statistical independence of the functions (10.6), and it generalizes (10.5), which, in view of (8.6) and (8.7), corresponds to the case $j = 1$.

Returning now to (10.3) and setting

$$\gamma_j = \exp\left(\frac{2\pi i \beta_j}{b}\right) \tag{10.7}$$

for $j = 1, \ldots, m$, with $\beta_j \in \{0, 1, \ldots, b - 1\}$, we get, in view of (8.2) and the expansion (10.4), a formula for the indicator functions

$$\chi_{\{b_1 = \beta_1, \ldots, b_m = \beta_m\}}$$

of b-adic intervals of order m, in terms of weighted power-products of the r_j. The resulting formula is, clearly, comparable to our (8.5), *et seq.*, and it can be used to motivate the introduction of power-products of the b-adic Rademacher functions, and, thus, of the b-adic Walsh functions, in order to deal with problems involving b-adic strings.

It can also be employed to establish an equivalence between multiplicativity of the b-adic Rademacher functions and the statistical independence of the b-adic digits, generalizing the case $b = 2$ treated in the previous section. Indeed, it is enough to make the substitution (10.7) in (10.5) to get

$$\mathbf{P}\{b_1 = \beta_1, \ldots, b_m = \beta_m\} = \frac{1}{b^m} = \prod_{j=1}^{m} \mathbf{P}\{b_j = \beta_j\}.$$

Consequently, *multiplicativity of the b-adic Rademacher functions implies statistical independence of the b-adic digits.*

Since there is no difficulty in extending to a general base b the argument used in the previous section to establish the Multiplicativity Formula for Rademacher functions, without having recourse to their statistical independence, we conclude that *statistical independence of the b-adic digits is equivalent to the multiplicativity property* (8.5) *possessed by the b-adic Rademacher functions.*

One way to look at this result is as follows. Recall that, in Section 8, we derived the Multiplicativity Formula (8.5) from the formula

$$\int_0^1 \prod_{j=1}^{k} \exp\bigl(i\mu_j b_j(\omega)\bigr)\, d\omega = \prod_{j=1}^{k} \int_0^1 \exp\bigl(i\mu_j b_j(\omega)\bigr)\, d\omega, \tag{8.4}$$

by setting $\mu_j = 2\pi d_j/b$, where the d_j are integers in the range from 0 to $b - 1$ and j varies from 1 to k. As shown there, (8.4) is a direct consequence of the statistical independence of the functions b_j.

However, it is known that the validity of (8.4) for *all* real values of the parameters μ_j *implies* statistical independence of the otherwise quite arbitrary real functions b_j, and Kac himself was among the first to prove it, by use of Fourier analysis [**3**].

In effect, what we have found is a refinement of Kac's sufficiency result, in a more algebraic setting, where the additive group of the reals has been replaced by the group of integers mod b. For, in restricting the b_j's to take integral values in the range from 0 to $b - 1$, instead of arbitrary real values, we have established that the validity of (8.4) for the *narrow* range of values of μ_j indicated above is already enough to ensure statistical independence of the b_j's.

The argument we have used cannot be regarded as new, even though its application to Rademacher and Walsh functions is. It is a reduction to the case of the cyclotomic group of a method suggested by Rényi [**9**, p. 171] for proving a more general result about the statistical independence of families of random variables, attributed to Kantorovic. Rényi would, no doubt, have appreciated that the method he proposed could be applied, trivially, to the classical Rademacher functions in order to establish their statistical independence, but he seems to have been unaware, as was Kac, of Mendès-France's extension of the Rademacher functions to the b-adic case and, thus, missed the opportunity to apply his method there.

The reader interested in learning more about normal numbers can consult Niven's Carus Monograph [**7**] and the treatise by Kuipers and Niederreiter [**5**], which contains an extensive bibliography.

ACKNOWLEDGMENT. This work was supported in part by the Consiglio Nazionale delle Ricerche and the Italian Ministry of the Universities and Scientific and Technological Research.

REFERENCES

1. E. Borel, Sur les probabilités dénombrables e leurs applications arithmétiques, *Rend. Circ. Mat. Palermo* **27** (1909) 247–271.
2. M. Kac, *Statistical Independence in Probability, Analysis and Number Theory*, Carus Math. Monographs, no. 12, Wiley, New York 1959.

3. M. Kac, Sur les fonctions indépendentes, I, *Studia Math.* **6** (1936) 46–58.
4. S. Kaczmarz, and H. Steinhaus, *Theorie der Orthogonalreihen*, Monografie Matematyczne, vol. 6, Warsaw and Lwow, 1935.
5. L. Kuipers, and H. Niederreiter, *Uniform Distribution of Sequences*, Pure and Applied Math., no. 29, Wiley, New York, 1974.
6. M. Mendès-France, Nombres normaux. Applications aux fonctions pseudo-aléatoires, *J. d'Anal. Math.* **20** (1967) 1–56.
7. I. Niven, *Irrational Numbers*, Carus Math. Monographs, no. 11, Wiley, New York, 1956.
8. S. S. Pillai, On normal numbers, *Proc. Indian Acad. Sci.*, Sect. A **12** (1940) 179–184.
9. A. Rényi, *Foundations of Probability*, Holden-Day, San Francisco, 1970.
10. D. D. Wall, *On Normal Numbers*, Ph.D. Thesis, Univ. of California, Berkeley, 1949.
11. H. Weyl, Über die Gleichverteilung von Zahlen mod. Eins, *Math. Ann.* **77** (1916) 313–352.

GERALD S. GOODMAN took degrees in mathematics at Haverford, Bryn Mawr, and Stanford, where he was a pupil of Loewner. His doctoral thesis was devoted to Loewner's theory of transformation semigroups, applied to conformal mapping. A chance encounter with Rényi led him to use those ideas to obtain new results concerning Markov chains. Later on, he applied probabilistic tools to Loewnerian semigroups, such as totally positive matrices. Although he has published papers in real analysis, control theory, and fractal image generation, some consider his finest mathematical achievement was finding a permanent post in a Renaissance city.
Statistics Department, University of Florence, Viale Morgagni, 59, 50134 Florence, ITALY
goodman@ds.unifi.it

...it did not take the Miloslavskys long to find a pretext for arresting Matveev. That educated gentleman was foolish enough to be found with a book of algebra in his baggage, which was, naturally, taken to be a form of black magic.

Even Nikita, when he heard of the arrest of his mentor, could only shake his head and remark: "He was asking for trouble. What did he want with such stuff anyway?"

Edward Rutherfurd, *Russka, The Novel of Russia*, Crown Publishers, Inc., New York, 1991, pp. 347–348.

Contributed by George Dorner, William Rainey Harper College, Palatine, IL