A proof that $\sqrt{s}$ for $s$ not a perfect square is simply normal to base 2

Richard Isaac

Abstract

Since E. Borel proved in 1909 that almost all real numbers with respect to Lebesgue measure are normal to all bases, an open problem has been whether simple irrational numbers like $\sqrt{2}$ are normal to any base. This paper shows that each number of the form $\sqrt{s}$ for $s$ not a perfect square is simply normal to base 2, that is, the averages of the first $n$ digits of its dyadic expansion converge to $1/2$. Let $\omega$ be irrational with dyadic expansion $\ldots x_1 x_2 \cdots$. Put $f_n(\omega) = (x_1 + x_2 + \cdots + x_n)/n$ and $f(\omega) = \limsup_k f_{n_k}(\omega)$ for any subsequence. Then $f$ is a tail function with respect to the $x$'s, that is, $f$ is a function of $x_n, x_{n+1}, \cdots$ for all $n$. Let irrational $\nu = \omega^2$ with dyadic expansion $\ldots u_1 u_2 \cdots$. Then $f_n$ and $f$ can be written as functions $h_n$ and $h$ of the $u$ variables. It turns out that if $h$ is a tail function with respect to the $u$ variables (Condition (TU)), then it is easy to prove the simple normality (section 2). The proof of Condition (TU) in section 3 depends on developing discrete versions of the partial derivatives of $f_n$ and $h_n$ with respect to their arguments and identifying convergence to 0 of these with the desired tail property. In theorem 3 it is shown that such convergence to 0 for $f_n$ relative to the $x$'s (which are assumed independent) carries over to the $h_n$ relative to the $u$'s. For those interested, an Appendix describes a heuristic probability model providing a motivation for the solution.\footnote{AMS 2010 subject classifications. primary, 11K16; secondary 39A12. Keywords and phrases. normal number, tail function, difference calculus.}

1 Introduction

A number is simply normal to base $b$ if its base $b$ expansion has each digit appearing with average frequency tending to $b^{-1}$. It is normal to base $b$ if its base $b$ expansion has each block of $n$ digits appearing with average frequency tending to $b^{-n}$. A number is called normal if it is normal to base $b$ for every
base. For a more detailed introductory discussion we refer to chapter 8 of [7] or section 9.11 of [3]. The most important theorem about normal numbers is the celebrated result (1909) of E. Borel in which he proved the normality of almost all numbers with respect to Lebesgue measure (for a proof see section 9.13 of [3]; an elegant probability proof using the strong law of large numbers appears in [6], p. 43).

Borel’s theorem left open the question, however, of identifying specific numbers as normal, or even exhibiting a common irrational normal number. The difficulty in exhibiting normality for such common irrational numbers is not surprising since normality is a property depending on the tail of the base $b$ expansion, that is, on all but a finite number of digits. By contrast, we mostly “know” these numbers by finite approximations, the complement of the tail.

Identification of well-known irrational numbers as normal has interest for computer scientists. The $b$-adic expansion of a number normal to the base $b$ is a sequence of digits with the basic properties of a random number table. There is thus the possibility that such numbers could be used for the generation of pseudo-random numbers. This would only be possible if the digits of the normal expansion could be generated quickly enough or stored efficiently enough to make the method practical.

In this paper we exhibit a class of numbers simply normal to the base 2. More precisely, we prove

**Theorem 1** Let $s$ be a natural number which is not a perfect square. Then the dyadic (base 2) expansion of $\sqrt{s}$ is simply normal.

For earlier versions of results in this article see [4] and [5].

Consider numbers $\omega$ in the closed unit interval $\Omega$, and represent the dyadic expansion of $\omega$ as

$$\omega = .x_1x_2\cdots, \quad x_i = 0 \text{ or } 1. \quad (1)$$

Also of interest is the dyadic expansion of $\nu = \omega^2$:

$$\nu = \omega^2 = .u_1u_2\cdots, \quad u_i = 0 \text{ or } 1. \quad (2)$$

In order to have uniqueness we do not consider rational numbers with terminating expansions (it will be seen shortly that all rational numbers must be excluded from the discussion). It is convenient to refer to the expansion of $\omega$ as an $x$ sequence and the expansion of $\nu$ as a $u$ sequence. A point of the unit interval can also be denoted by its coordinate representation, that is, $\omega = (x_1, x_2, \cdots)$ or $\nu = (u_1, u_2, \cdots)$. The coordinate functions $x_n$ and $u_n$
give the $n$th coordinates of $\omega$ and $\nu$ respectively.

1.1 Tail functions, coordinate averages

Given any dyadic expansion $s_1 s_2 \cdots$ and any positive integer $n$, the sequence of digits $s_n, s_{n+1}, \cdots$ is called a tail of the expansion. Two expansions are said to have the same tail if there exists $n$ so large that the tails of the sequences from the $n$th digit are equal. We consider functions $g$ with domain the set of all the dyadic expansions defined above. The function $g$ is called a tail function with respect to the expansions if $g(s^{(1)}) = g(s^{(2)})$ whenever $s^{(1)}$ and $s^{(2)}$ have the same tail. Let $s$ be an expansion and let $s_k$ be the expansion for which the $k$th coordinate of $s$ and $s_k$ differ, but all other coordinates are the same. Then the tail function $g$ satisfies $g(s) = g(s_k)$, so that $g$ does not depend on the values of any individual coordinate or finite set of them.

The average

$$f_n(\omega) = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

is the relative frequency of 1’s in the first $n$ digits of the expansion of $\omega$. Simple normality for $\omega$ is the assertion that $f_n(\omega) \to 1/2$ as $n$ tends to infinity. Let $n_k$ be any fixed subsequence and define

$$f(\omega) = \limsup_{k \to \infty} f_{n_k}(\omega).$$

We note that the function $f$ is a tail function with respect to the $x$ sequence.

1.2 $f$ not permitted to be a function of a finite number of $x$ variables; exclusion of rationals

It is fundamental to our arguments that $f$, being a tail function, does not depend on any finite set of $x$ variables. But there is a problem if certain points are chosen. To see this let $T$ be the shift transformation on $\Omega: T(x_1, x_2, \cdots) = (x_2, x_3, \cdots)$. Suppose, for a fixed value $\omega_0$, there exist distinct non-negative integers $r$ and $s$ with $T^r(\omega_0) = T^s(\omega_0)$. Then it is easy to see that the sequence is periodic, that is, consists of repeated patterns. To fix ideas let $\omega_0 = .0100110011 \cdots$ be the dyadic expansion of the decimal .3 with the repeating pattern 0011. Then $T^2 \omega_0 = T^6 \omega_0$. Once the coordinates of a period $x_03, x_04, x_05, x_06$ are known, all coordinates of a tail are determined. Since $f$ is a tail function, $f(\omega_0)$ can be written as a function of only the four coordinates given. Thus there is a function $g(\omega) = g(x_3, x_4, x_5, x_6)$
such that $g(\omega_0) = f(\omega_0)$. Clearly, if $g$ is used to calculate $f$ at $\omega_0$, $f$ surely depends on a finite set of coordinates, contradicting our requirements. However, it is easy to eliminate this problem: the sequence $\omega$ consists of repeated patterns if and only if $\omega$ is rational so it will be assumed throughout that $\omega$ and $\nu = \omega^2$ are irrational.

1.3 $f_n$ written as a function $h_n$ of the $u$ variables; Condition (TU)

Observe that the average $f_n$, defined in terms of the $x$ sequence, can also be expressed as a function $h_n(\nu)$ of the $u$ sequence because the $x$ and $u$ sequences uniquely determine each other. This relationship has the simple form $f_n(\omega) = f_n(\sqrt{\nu}) = h_n(\nu)$. Define $h(\nu) = \limsup_k h_{n_k}(\nu)$; then clearly $f(\omega) = h(\nu)$. We say that $\omega$ and $\nu = \omega^2$ are points (or expansions) that correspond to one another.

The functions $h_n$ are functions of $u$ variables. We have seen that $f$ is a tail function with respect to the $x$ variables. Consider the possibility that $h$ is a tail function with respect to the $u$ variables; more precisely, consider the following condition.

**Definition:** Let $f$ be defined as in relation 4 for any fixed subsequence $n_k$. We say that Condition (TU) is satisfied if $f(\omega) = h(\nu)$ is a tail function with respect to the $u$ sequence whatever the subsequence $n_k$, that is, for any $\omega$ and any positive integer $n$, $h(\nu)$ (and therefore $f(\omega)$) only depends on $u_n, u_{n+1}, \cdots$, the tail of the expansion of $\nu = \omega^2$. (The notation “TU” is meant to suggest the phrase “tail with respect to the $u$ sequence”.)

An immediate consequence of Condition (TU) is:

**Proposition 1** Let $\eta$ be the dyadic expansion of an irrational number. Let $\eta_1$ be a dyadic expansion that agrees with $\eta$ at all but a finite number of indices. If Condition (TU) is satisfied then

$$\lim_n (f_n(\sqrt{\eta}) - f_n(\sqrt{\eta_1})) = 0.$$ 

**An Example:** let $\omega = \sum_{k=1}^{\infty} 2^{-k^2}$. The expansion of $\omega$ has larger and larger stretches of 0’s and $f_n(\omega)$ converges to 0. The first digit $u_1$ of $\nu = \omega^2$ is 0. Let $\nu_1 = 1/2 + \nu$ and define $\omega_1 = \sqrt{\nu_1}$. Then Condition (TU) implies that $f_n(\omega_1)$ also converges to 0. So although the digits of the expansion of $\omega_1$ differ from those of $\omega$ in an infinite number of places, the averages $f_n$ have the same limiting behavior there.
1.4 Informal roadmap to this work

Sufficiency of Condition (TU); Section 2

1. In section 2 it is shown that Condition (TU) implies theorem 1. This is done by constructing two numbers \( \omega_{s1} \) and \( \omega_{s2} \) between 0 and 1 with the following properties: (1) the squares of the numbers have the same tail, so asymptotic averages of the expansions of the numbers are equal by Condition (TU), (2) the sum of the asymptotic averages of the expansions is 1. This easily leads to the conclusion.

Sketch of ideas of argument that Condition (TU) holds; Section 3

2. The functions \( f_n \) do not have classical partial derivatives, but if the \( x \) variables were independent and real valued, defined on the unit interval, one could differentiate to get \( \partial f_n / \partial x_j = 1/n \to 0 = \partial f / \partial x_j \) for all \( j \). This relation implies the tail property and gives the rate at which the effect of \( x_j \) on \( f_n \) dies out in the limit. In this case clearly one does not need derivatives to see the tail property of \( f \) with respect to the \( x \)'s because of the form of \( f_n \) as an average. To prove Condition (TU), however, it must be shown that \( h \) has the tail property with respect to the \( u \)'s and, unlike with \( f_n \) and \( f \), there is no way to see this directly; some tool is needed to imply this tail property. If there were classical partial derivatives, a natural tool would be to try and show \( \partial h_n / \partial u_i \to 0 \) for all \( i \). This implies the desired tail property. Unfortunately these partials do not exist.

3. Since the classical partials don’t exist in our context, we develop a calculus for functions on the dyadic sequences (subsection 3.1). Analogs to the classical partial derivatives are defined, called partial differences. These can be used to get discrete versions of the total differential formulas of multivariable calculus (subsection 3.2).

4. Once the partial differences have been defined, it is easy to check that if the \( x \) variables are independent and \( \Delta x_r \neq 0 \), the partial differences of \( f_n \) with respect to \( x_r \) (written \( \Delta f_n / \Delta x_r \)) equal \( 1/n \to 0 \) (corollary 1). This is what we expect since the partial differences are the discrete analogs of \( \partial f_n / \partial x_j \). The idea now is to see whether the partial differences of \( h_n \) with respect to \( u_i \) (written \( \Delta h_n / \Delta u_i \)) also converge to 0, somehow “inheriting” this convergence property from \( f_n \) relative to the \( x \) sequences. To study this it is necessary to get a connection between the \( x \) and \( u \) sequence spaces.

5
5. Section 3 begins with a result connecting the $x$ and $u$ sequence spaces (lemma 2). The simplest form of the statement asserts that given any $\omega$ and $\nu = \omega^2$, consider a fixed $u_i$, say. Then one can find a finite initial segment $x_1, x_2, \ldots, x_N$ of $\omega$ such that if $\omega_1$ is any other point with the same initial segment, then the $i$th coordinate of $\omega_1^2$ is $u_i$. So $u_i$ is determined by (i.e., is a function of) this initial segment. There is a similar statement interchanging the roles of $x$ and $u$ sequences.

6. The fundamental theorem uses the foregoing ideas to prove the validity of Condition (TU). See the Appendix, subsection 4.2, for a sketch of the main steps of the proof.

Appendix; section 4

6. The Appendix has extra material not integral to the understanding of the paper. Subsection 4.1 describes a heuristic probability model that inspired the approach to the problem taken in this work. Subsection 4.2 outlines the main steps in the proof of theorem 3.

2 Condition (TU) implies simple normality

In this section we prove that Condition (TU) implies theorem 1.

**Theorem 2** If Condition (TU) is satisfied then theorem 1 is true.

**Lemma 1** Let $s$ be a natural number which is not a perfect square, and let $l$ be any integer such that $2^l > s$. Define the points

$$\omega_{s1} = 1 - (\sqrt{s}/2^l)$$

and

$$\omega_{s2} = (\sqrt{s} - 1)/2^l.$$  

Let $f = \lim \sup_{i \to \infty} f_{n_i}$ where $n_i$ is any fixed subsequence. Assume Condition (TU) is satisfied. Then $f(\omega_{s1}) = f(\omega_{s2})$.

Proof: The numbers $\omega_{si}$ are less than 1 for $i = 1, 2$ and their squares are both irrational and are respectively given by

$$1 + s(2^{-4l}) - (2^{-2l+1}\sqrt{s}) \text{ and } (s + 1)2^{-2l} - (2^{-2l+1}\sqrt{s}).$$  

(5)

The dyadic expansions of the rational terms $1 + s(2^{-4l})$ and $(s + 1)2^{-2l}$ in relation 5 have only a finite number of non-zero digits. Now consider the
dyadic expansion of the term $2^{-2l+1}\sqrt{s}$ (this is obtained from the expansion of $\sqrt{s}$ by shifting the “decimal” point $2^{2l-1}$ places to the left). To get each of the values in relation 5, this term must be subtracted from each of the larger rational terms which have terminating expansions; it is clear that the resulting numbers have expansions with the same tail, that is, the expansions of $\omega_s^2$ and $\omega_s^2$ have the same tail. Then Condition (TU) implies that $f(\omega_s^1) = f(\omega_s^2)$. This finishes the proof of lemma 1.

The proof of theorem 2 will now be completed. It is sufficient to prove simple normality for $\lambda = \sqrt{s} - [\sqrt{s}] < 1$ where $[t]$ is greatest integer $\leq t$. Define $g_n(\omega)$ to be the average number of 0’s in the first $n$ digits of the expansion of $\omega$. Let $n_i$ be any subsequence such that $f_{n_i}(\lambda)$ converges to some value $a$. Now consider the point $\lambda' = 1 - \lambda$, and notice that for all $j$ the $j$th digit of $\lambda$ and the $j$th digit of $\lambda'$ add to 1. It follows that $g_{n_i}(\lambda')$ also converges to $a$. Note that the point $\omega_s^1$ (as defined in lemma 1) would have the same tail as $\lambda'$ were we to shift a finite number of places, and therefore $\omega_s^1$ and $\lambda'$ have the same asymptotic relative frequency of 0’s and 1’s. The same can be asserted for $\omega_s^2$ and $\lambda$. Thus lemma 1 can be applied to conclude that the asymptotic averages based on $f_{n_i}$ evaluated at the points $\lambda$ and $\lambda'$ are equal, that is,

$$\limsup_{i \to \infty} f_{n_i}(\lambda') = \lim_{i \to \infty} f_{n_i}(\lambda) = a.$$ 

But the equation $f_n + g_n = 1$ holds for all $n$ at all points; apply it for $n = n_i$ at the point $\lambda'$, take the limit, and conclude that since $g_{n_i}(\lambda')$ converges to $a$, $f_{n_i}(\lambda')$ converges to $1 - a$. The preceding relation then shows $a = 1 - a$, or $a = 1/2$. Since we have obtained convergence to 1/2 for $f_n(\lambda)$ along the arbitrary convergent subsequence $n_i$, it follows that $f_n(\lambda)$ itself converges to 1/2. The proof of theorem 2 is complete.

3 Proof that Condition (TU) is satisfied

We begin with some elementary observations about the relationship between the digits in the expansion of $\omega$ and those in the expansion of $\nu = \omega^2$. The initial segment of length $n$ of an expansion refers to its first $n$ digits.

**Lemma 2** Let $\omega^2 = \nu$ be irrational.

(a) Let $\cdot u_1 u_2 \cdots u_r$ be the initial segment of length $r$ of $\nu$. Then there exists a positive integer $N = N(\omega, r)$ such that if $\omega_*$ is any point with the same initial segment of length $N$ as $\omega$, then $\omega_*^2$ has the same initial segment of
length $r$ as $\nu$. Therefore each $u_i$, $i \leq r$ is a function of $x_1, x_2, \ldots, x_N$. 
(b) Let $x_1x_2 \cdots x_n$ be the initial segment of length $n$ of $\omega$. Then there exists a positive integer $m = m(\nu, n)$ such that if $\nu_*$ is any point with the same initial segment of length $m$ as $\nu$, then $\sqrt{\nu_*}$ has the same initial segment of length $n$ as $\omega$. Therefore each $x_j$, $j \leq n$ is a function of $u_1, u_2, \ldots, u_m$.

Proof: We prove (a). For fixed $r > 1$, consider the decomposition of $\Omega - \{1\}$ by the intervals
\[
I_{k+1} = [k 2^{-r}, (k + 1) 2^{-r}) \quad 0 \leq k \leq 2^r - 1.
\]
Each of these intervals will be called an $r$ box. Note that the point defined by $\nu_r = \nu_1u_2 \cdots u_r$ is the left hand endpoint of an $r$ box. Points lie in the same $r$ box if and only if they have identical initial segments of length $r$. Since $\omega^2 = \nu$ is irrational, $\nu$ lies in the interior of the $r$ box for which $\nu_r$ is the left hand endpoint; let the minimum distance of $\nu$ from the ends of the $r$ box be $\varepsilon$. Let the point $\omega_j = x_1x_2 \cdots x_j$ be the initial segment of length $j$ of $\omega$. As $j \to \infty$, $\omega_j^2$ converges to $\nu$. Then there is a smallest integer $N$ so large that if $\omega_*$ is any point with the same initial segment of length $N$ as $\omega$
\[
|\omega_*^2 - \omega_N^2| \leq 2 |\omega_* - \omega_N| < 2^{-(N-1)} \frac{\varepsilon}{2} \quad \text{and} \quad |\omega_N^2 - \nu| < \frac{\varepsilon}{2}
\]
so that the distance between $\omega_*^2$ and $\nu$ is less than $\varepsilon$, that is, $\omega_*^2$ and $\nu$ lie in the same $r$ box. Thus $\omega_N$ determines the initial segment of length $r$ of $\nu$. The proof of (a) is complete. The proof of (b) is similar to that of (a). Details are omitted.

### 3.1 A calculus for functions on the dyadic sequences

To prove that Condition (TU) holds, we need to study how changing $u_i$ produces changes in the averages $h_n$. It is possible to examine these discrete changes by creating a difference calculus. Since there do not seem to be any useful references to what is needed we develop these simple results here.

A very elementary introduction to finite differences may be found, for example, in [1]. To review some of the notation, let $v(y_1, \ldots, y_l)$ be a function on the $l$-fold product space $S^l$ where the variables $y_i$ take values in $S$, a set of real numbers. Suppose that the variable $y_i$ is changed by the amount $\Delta y_i$ such that the $l$-tuple $y^{(1)} = (y_1, \ldots, y_l)$ is taken into $y^{(2)} = (y_1 + \Delta y_1, \ldots, y_l + \Delta y_l)$ in the domain of definition of $v$. 

---

8
Put $v(y^{(2)}) - v(y^{(1)}) = \Delta v$, and let

$$
\Delta v_i = v(y_1, \ldots, y_{i-1}, y_i + \Delta y_i, y_{i+1}, \ldots, y_l + \Delta y_l)
\quad - \quad v(y_1, \ldots, y_{i-1}, y_i, y_{i+1}, \ldots, y_l + \Delta y_l).
$$

Then $\Delta v = \sum_i \Delta v_i$ is the total change in $v$ induced by changing all of the $y_i$, where this total change is written as a sum of step-by-step changes in the individual $y_i$. Formally, by dividing, we can write

$$
\Delta v = \sum_i (\Delta v_i / \Delta y_i) \cdot \Delta y_i.
$$

If some $\Delta y_i = 0$, its coefficient in relation (7) has the form $0/0$, which is interpreted as 0. Let us then formally define the partial difference of $v$ with respect to $y_i$, evaluated at the pair $(y^{(1)}, y^{(2)})$ by

$$
\frac{\Delta v}{\Delta y_i} = \begin{cases} 
\Delta v_i / \Delta y_i = \Delta v_i, & \text{if } \Delta y_i = 1, \\
\Delta v_i / \Delta y_i = -\Delta v_i, & \text{if } \Delta y_i = -1, \\
0, & \text{if } \Delta y_i = 0.
\end{cases}
$$

Notice that the forward slash (/) in this relation expresses division and the horizontal slash on the left hand side is the partial difference operator.

The sum $\Delta v$ of relation (7) is called the total difference of $v$ evaluated at the given pair and can now be written

$$
\Delta v = \sum_i \frac{\Delta v}{\Delta y_i} \cdot \Delta y_i.
$$

The $i$th summand in relation (9) is called the $i$th partial difference of $v$ relative to the given pair. The partial and total differences are the discrete analogs of the partial and total differentials in the theory of differentiable functions of several real variables and the partial difference with respect to a given $y$ variable is the analog of the partial derivative (see, e.g., pp. 300-310, [2]).

The $i$th partial difference of $v$ at a given pair is a measure of the contribution of $\Delta y_i$ to $\Delta v$ when all the other $y$ variables are held constant.

Returning to our particular problem, we have seen in Section 1 that the average $f_n(\omega)$ of relation (3) can be written as a function $h_n(\nu)$, where $\omega$ and $\nu = \omega^2$ correspond to one another. With a slight abuse of notation we can write

$$
f_n(x_1, \ldots, x_n) = f_n(\omega) = h_n(\nu) = h_n(u_1, u_2, \ldots).
$$
For a pair $(\omega, \omega^{(1)})$, let $\Delta x_j$ be the $j$th coordinate difference of $\omega$ from $\omega^{(1)}$ ($\Delta x_j = 0, 1, \text{or} -1$). Let $\nu$ and $\nu^{(1)}$ correspond to $\omega$ and $\omega^{(1)}$ with $\Delta u_i$ the $i$th coordinate difference of $\nu$ from $\nu^{(1)}$. Then the changes $\Delta x_j$ in the $x$ coordinates which we can think of as having been chosen independently, have induced changes $\Delta u_i$ in the dependent $u$ coordinates. Of course our point of view could have been reversed. Starting with a pair $(\nu, \nu^{(1)})$, independent changes $\Delta u_i$ induce dependent changes $\Delta x_j$ in the $x$ coordinates. Note consistency requirements when writing terms like $x_j + \Delta x_j$: if $x_j = 0$, then $\Delta x_j = 0$ or $+1$; if $x_j = \pm 1$, then $\Delta x_j = 0$ or $\mp 1$.

3.2 Representations of total differences for $f_n$ and $h_n$

As with multivariable calculus where total differentials can be represented as a sum of partial differentials, total differences also have representations in terms of partial differences. The following results are basic for the arguments in this paper. The first lemma is fairly evident.

**Lemma 3** At the pair $(\omega, \omega^{(1)})$, $\Delta f_n$ can be represented as a total difference

$$
\Delta f_n = f_n(\omega^{(1)}) - f_n(\omega) = \frac{1}{n} \sum_{1 \leq j \leq n} \Delta x_j
$$

Proof: The pair $(\omega, \omega^{(1)})$ denotes a generic (variable) input. Decompose according to the recipe given in relations [8] to [9] to get relation [11].

The next result is the partial difference analog of the formal relation $\partial f_n / \partial x_j = 1/n$. It gives a rate of convergence for the dying out of the influence of $x_j$ on $f$.

**Corollary 1** Suppose the variables $x_j$ are independent and $\Delta x_r \neq 0$ for fixed $r$, $1 \leq r \leq n$. Let $f$ be defined according to relation [4]. Then

$$
\frac{\Delta f_n}{\Delta x_r} = \frac{1}{n} \to 0 = \frac{\Delta f}{\Delta x_r}, \quad \text{as} \ n \ \text{tends to} \ \infty.
$$

Proof: Take the partial difference of both sides of relation [11] with respect to $x_r$ (this is an additive operation). Independence of the $x$ variables implies

$$
\frac{\Delta x_j}{\Delta x_r} = 0, \quad j \neq r, \quad \frac{\Delta x_r}{\Delta x_r} = 1,
$$

10
proving all but the last equality. But this follows because it is already known that \( f \) is a tail function with respect to the \( x \)'s. It can easily be proved directly: let \( n_k \) be a subsequence with \( \lim\sup_k f_{n_k} = f \). Put \( g_m = \sup_{k \geq m} f_{n_k} \). Then the partial difference of \( g_m \) with respect to \( x_r, r \leq n_m \) is bounded by \( 1/n_m \). Using relations 6 to 9 we see that \( \lim g_m = f \) has partial difference 0. This completes the proof.

**Lemma 4.** At the pair \((\nu, \nu^{(1)})\), \( \Delta h_n \) can be represented as a total difference

\[
\Delta h_n = h_n(\nu^{(1)}) - h_n(\nu) = h_n(u_1 + \Delta u_1, u_2 + \Delta u_2, \cdots) - h_n(u_1, u_2, \cdots)
\]

\[
= \sum_{i \geq 1} \frac{\Delta h_n}{\Delta u_i} \Delta u_i = \sum_{i \geq 1} (\Delta h_n,i/\Delta u_i) \Delta u_i, \quad \Delta u_i \neq 0 \quad (13)
\]

where

\[
\Delta h_n,i = \Delta h_n,i(\nu, \nu^{(1)}) =
\]

\[
h_n(u_1, \cdots, u_{i-1}, u_i + \Delta u_i, u_{i+1} + \Delta u_{i+1}, \cdots) -
\]

\[
h_n(u_1, \cdots, u_{i-1}, u_i, u_{i+1} + \Delta u_{i+1}, \cdots).
\]

The formally infinite sum of relation 13 reduces to a finite sum. More precisely, given the pair \((\nu, \nu^{(1)})\), there exists an integer \( m \) such that the partial differences \( \Delta h_n,i/\Delta u_i = 0 \) for all \( i > m \). The number of non-vanishing terms in the sum depends on \( \nu \) and \( n \).

Proof: The pair \((\nu, \nu^{(1)})\) is a generic input. The recipe given in relation 13 for decomposing \( \Delta h_n \) is given by the definitions stated in relations 6 through 9. The pair \((\nu, \nu^{(1)})\) corresponds to the pair \((\omega, \omega^{(1)})\). To see that the sum in relation 13 is finite, note that the function \( h_n = f_n \) only depends on \( x_1, x_2, \cdots x_n \). So given the pair \((\nu, \nu^{(1)})\) lemma 2 proves the existence of an integer \( m \) such that for all \( i > m \)

\[
\frac{\Delta x_j}{\Delta u_i} = 0, \quad 1 \leq j \leq n.
\]

and therefore the terms \( \Delta h_n,i \) of relation 14 are 0 for \( i > m \). Thus the terms in the sum of relation 13 vanish for \( i > m \) and the formula of relation 13 represents a finite sum. This concludes the proof of the lemma.

### 3.3 Solution of problem: theorem 3

We know that if the \( x \) variables are independent, the contribution to changes in the averages \( f_n \) of any change in a single \( x_j \) tends to 0 as \( n \) tends to infinity
(corollary [1]). But how about the partial differences of \( h_n \) with respect to a single fixed \( u_i \)? Does a change in \( u_i \) induce changes in \( h_n \) that die out in the limit? The answer is not obvious. Heuristic considerations suggest why we might expect this to be so: \( h_n \) depends on larger and larger initial segments of \( u \) variables as \( n \) increases. Each \( u \) variable depends on just a finite number of \( x \) variables (lemma [2]) and each \( x \) variable has negligible effect on \( h_n = f_n \) for large \( n \). It seems reasonable to suspect that change in a single \( u \) variable is not going to have much of an effect on \( h_n \) for large \( n \).

We now set out to prove that this suspicion is true. The basic result leading to the solution of the problem follows.

**Theorem 3** Assume that the \( u \) variables are functions of independent \( x \) variables. Let \( (\nu, \nu^{(1)}) \) be the input pair in lemma [3] and \( \Delta h_n \) its total difference in accordance with relation [13]. Then
(a): For all \( i \), the partial differences of \( h_n \) with respect to \( u_i \) in relation [13] satisfy
\[
\lim_{n} \frac{\Delta h_n}{\Delta u_i} = 0. \tag{15}
\]
(b): Condition \((TU)\) is true.

Proof of part (a): The \( i \)th partial differences \( \Delta u_i \) referenced in (a) are all non-zero, so let \( r \) be a fixed positive integer with \( \Delta u_r = \pm 1 \). Consider relation [14]. The right hand side expresses \( \Delta h_{n,r} \) as the difference \( h_n(\nu_2) - h_n(\nu_1) \) evaluated at the two points
\[
\nu_2 = (u_1, \cdots, u_{r-1}, u_r + \Delta u_r, u_{r+1}, \Delta u_{r+1}, \ldots) \quad \text{and} \quad \nu_1 = (u_1, \cdots, u_{r-1}, u_r, u_{r+1}, \Delta u_{r+1}, \ldots) \tag{16}
\]

The irrationality of \( \nu^{(1)} \) implies that \( \nu_1 \) and \( \nu_2 \) are also irrational. For \( k = 1, 2 \), let \( \nu_k \) correspond to \( \omega_k \) (i.e., \( \omega_k = \sqrt{k} \)) and let \( \omega_k = (x(1,k), x(2,k), \cdots) \). Put \( x_{(j,2)} - x_{(j,1)} = \Delta x_j' \) and let the differences in the \( u \) coordinates at \((\nu_1, \nu_2)\) be denoted by \( \Delta u_r' \). Then \( \Delta u_i' = 0 \) for \( i \neq r \), \( \Delta u_r' = \Delta u_r \). The idea now is to change our focus; let us study the functions \( f_n \) and \( h_n \) at the pairs \((\omega_1, \omega_2)\) and \((\nu_1, \nu_2)\). Use the pairs \((\omega_1, \omega_2)\) and \((\nu_1, \nu_2)\) as the input pairs in lemmas [3] and [4] respectively to get the representations
\[
\Delta h_n' = \Delta f_n' = f_n(\omega_2) - f_n(\omega_1) = \frac{1}{n} \sum_{1 \leq j \leq n} \Delta x_j' \tag{17}
\]
and
\[
\Delta h_n' = h_n(\nu_2) - h_n(\nu_1) = \frac{\Delta h_n'}{\Delta u_r} \Delta u_r, \quad \text{where} \quad \frac{\Delta h_n'}{\Delta u_r} = \pm \frac{1}{n} \sum_{1 \leq j \leq n} \Delta x_j'. \tag{18}
\]
To summarize what we have done above: we have obtained a new pair \((\nu_1, \nu_2)\) and induced pair \((\omega_1, \omega_2)\). Both pairs are used as input pairs for the representations of lemmas 3 and 4 where primes are used to distinguish \(\Delta x\) and \(\Delta u\) variables from the generic inputs. Since the initial pair chosen above is in terms of \(u\) variables and the induced variables are in terms of \(x\) variables, this indicates that the \(u\) variables could naturally be considered independent and the \(x\)'s dependent. But due to the symmetry it is legitimate to think of the process in reverse: that is, let us assume that the \(x\) variables (the \(\Delta x'\)) have been chosen independently and induced the representation for \((\nu_1, \nu_2)\), so the \(u\) variables are dependent. This is what we will assume during the remainder of the argument.

From relation 18,

\[
\Delta h'_n = \frac{\Delta h'_n}{\Delta u_r} \Delta u_r = h_n(\nu_2) - h_n(\nu_1) = \frac{\Delta h_n}{\Delta u_r} \Delta u_r = \Delta h_{n,r},
\]

and so

\[
\frac{\Delta h'_n}{\Delta u_r} = \frac{\Delta h_n}{\Delta u_r} \quad \text{(19)}
\]

and

\[
\Delta f'_n = \Delta h'_n = \frac{\Delta h_n}{\Delta u_r} \Delta u_r. \quad \text{(20)}
\]

Since \(u_r\) is a function of the \(x\) variables, at the pair \((\omega_1, \omega_2)\) relations 6 through 9 give the representation

\[
\Delta u_r = \sum_{j \geq 1} \frac{\Delta u_r}{\Delta x'_j} \Delta x'_j. \quad \text{(21)}
\]

By lemma 2 there exists \(N = N((\omega_1, \omega_2), r)\) such that the changes \(\Delta x'_j, j > N\) cause no change in \(u_r\), that is,

\[
\frac{\Delta u_r}{\Delta x'_j} = 0, \quad j > N.
\]

It follows that relation 21 gives a finite decomposition.

The original input pair \((\nu, \nu^{(1)})\) in lemma 4 is assumed to be any generic pair, that is, our notation has referred to variables. We now choose an arbitrary fixed input pair \((\nu, \nu^{(1)})\). Boldface characters are used to denote fixed pairs corresponding to the variables denoted by the same notation without boldface. Thus \((\omega_1, \omega_2)\) and \((\nu_1, \nu_2)\) denote the associated pairs
induced by \((\nu, \nu^{(1)})\) using relation 16. For \((\omega_1, \omega_2)\) specify an integer \(p \leq N(\omega_1, \omega_2)\), an index with

\[
\frac{\Delta u_r}{\Delta x'_p} \Delta x'_p \neq 0, \quad \text{evaluated at } (\omega_1, \omega_2). \tag{22}
\]

Such \(p\) exists by relation 21 since \(\Delta u_r \neq 0\). We also specify an arbitrary subsequence \(n_k((\nu_1, \nu_2))\) such that

\[
\lim_{k} \frac{\Delta h_{n_k}}{\Delta u_r} \text{ converges to some constant } c, \quad \text{evaluated at } (\nu_1, \nu_2). \tag{23}
\]

Return now to the general formulas in terms of variables. Replace \(n\) by \(n_k\) in relation 20, take the limsup of both sides of the equation, and substitute the finite representation of relation 21 for \(\Delta u_r\) to get

\[
\Delta f' = \Delta h' = \limsup_k \Delta h'_{n_k} = \limsup_k \frac{\Delta h_{n_k}}{\Delta u_r} \Delta u_r = \sum_j \limsup_k \frac{\Delta h_{n_k}}{\Delta u_r} \frac{\Delta u_r}{\Delta x'_j} \Delta x'_j. \tag{24}
\]

We are going to take the partial difference of the far left and far right hand sides of relation 24 with respect to the variable \(x'_p\) to get

\[
0 = \frac{\Delta f'}{\Delta x'_p} = \sum_j \limsup_k \frac{\Delta h_{n_k}}{\Delta u_r} \frac{\Delta u_r}{\Delta x'_j} \frac{\Delta x'_j}{\Delta x'_p}. \tag{25}
\]

The first equality of relation 25 follows from corollary 1. On the right hand side first note that the partial difference operator is additive. Then recall that by relations 18 and 19 the limsup term is a tail function with respect to the \(x\) variables so is not a function of \(x'_p\). Changing \(x'_p\) does not effect this term and it acts like a constant when taking the partial difference. Thus the partial difference of the right hand side of relation 24 is given by the right hand side of relation 25. Now use the property that if the \(x\) variables are independent, changes in \(x'_p\) cause no change in \(x'_j\), for \(j \neq p\), so the partial differences are equal to 0 for \(j \neq p\) and, by relation 22, equal to 1 for \(j = p\). (see also relation 12). It follows that relation 25 simplifies to

\[
0 = \limsup_k \frac{\Delta h_{n_k}}{\Delta u_r} \frac{\Delta u_r}{\Delta x'_p}. \tag{26}
\]

Relation 26 is an identity, valid for all values of the variables. Evaluating it at the fixed value \((\nu_1, \nu_2)\) and using relation 23 relation 26 becomes

\[
0 = c \cdot \frac{\Delta u_r}{\Delta x'_p}(\omega_1, \omega_2). \tag{27}
\]
The partial difference term here is not equal to 0 by relation 22; therefore \( c = 0 \). The subsequence \( n_k \) is associated with an arbitrary limit point of \((\nu_1, \nu_2)\) so the above argument shows this limit point is unique, that is, for the fixed pair \((\nu_1, \nu_2)\)

\[
\lim_{n} \frac{\Delta h_n}{\Delta u_r} = 0.
\]

Since \( r \) was an arbitrary index from the set of indices \( i \) in relation 13, the limit relation is valid for all indices \( i \) for the fixed pair \((\nu_1, \nu_2)\). Finally, the pair \((\nu_1, \nu_2)\) is itself arbitrary in the domain of possible choices, so the proof of part (a) of the theorem is complete.

Proof of part (b): Let \( n_k \) be any subsequence with \( \lim \sup_{k} f_{n_k} = f \). In terms of the \( h \) functions, this is the same as \( \lim \sup_{k} h_{n_k} = h \). Put \( g_m = \sup_{k \geq m} h_{n_k} \). By part (a) of the theorem, for \( m \) large enough the partial difference of \( g_m \) with respect to any fixed \( u_i \) is less than \( \varepsilon \). The definitions in relations 6 to 9 then prove \( \lim g_m = h \) has partial difference with respect to \( u_i \) equal to 0 (see also proof of corollary 1). Thus changing \( u_i \) does not change \( h \). Given a finite set of indices, change \( u \) variables one index at a time, producing no change at each stage. Hence \( h \) is only a function of its tail, and this is Condition (TU). So Condition (TU) is valid and the proof of theorem 1 is complete.

4 Appendix

4.1 Probabilistic motivation for proof that Condition (TU) holds

The strategy of the proof is partially suggested by a heuristic probability argument. Assume the expansion \( \omega = (x_1, x_2, \ldots) \) represents a sequence of independent random variables with regard to some basic measure \( \pi \). Then \( \nu = (u_1, u_2, \ldots) \), the square of \( \omega \), is a sequence of random variables depending upon \( \omega \). Suppose the variables \( u_i \) can be expressed as a function of a finite number of the \( x \)'s, for example, suppose \( u_1 = \alpha(x_1, x_2, \ldots, x_N) \). The function \( f \) of relation 4 being a tail function, is a function of the \( x_j \) for \( j > N \) and so \( f \) is stochastically independent of the random variables \( x_1, x_2, \ldots, x_N \). By a well known result in probability this implies that \( f \) is also independent of \( u_1 \). The result can be interpreted to mean that no information about \( u_1 \) can tell us anything about \( f \), at least up to \( \pi \) sets of measure 0, and, more generally, that no finite sequence \( u_1, u_2, \ldots u_r \) is informative about \( f \). This is a stochastic version of Condition (TU).
Unfortunately it does not seem easy to get a rigorous probability argument along the lines described above. The proof given in this paper uses no formal probability, but it imitates the strategy of the probability model using functional counterparts to stochastic properties. For example, partial differences are defined and partial differences converging to 0 are used as a functional substitute for stochastic independence. From lemma 2, \( u_1 \), say, can be expressed as a function \( \alpha(x_1, x_2, \cdots, x_N) \) of at most a finite number of \( x \) variables. Since the contributions to the averages \( f_n \) for each of \( x_1, x_2, \cdots, x_N \) die out in the limit, we wish to conclude that property carries over to \( h_n \) relative to \( u_1 \). If the \( x \) variables are functionally independent, this is proved in theorem 3 and would be the analog of the probability theorem that if \( f \) is stochastically independent of finite sets of \( x \)'s and \( u_1 \) is a function of a finite number of \( x \)'s, then \( f \) is stochastically independent of \( u_1 \).

4.2 Sketch of main steps in proof of theorem 3

Given the generic input pair \( (\nu, \nu^{(1)}) \) of lemma 4, another pair \( (\nu_1, \nu_2) \) is defined with total difference \( \Delta h_n' \) and lemma 4 representation

\[ \Delta h_n' = \Delta u_r \Delta u_r. \]

The right hand side of this equation is the \( r \)th partial difference of relation 13. Let \( (\omega_1, \omega_2) \) be the corresponding pair induced by \( (\nu_1, \nu_2) \) and having differences \( \Delta x'_j \).

There is a finite sum representation

\[ \Delta u_r = \sum_{j \geq 1} \frac{\Delta u_r}{\Delta x'_j} \Delta x'_j. \]

Choose a fixed pair \( (\nu, \nu^{(1)}) \) with \( (\omega_1, \omega_2) \) and \( (\nu_1, \nu_2) \) the fixed pairs associated with it. There exists an index \( p \) such that

\[ \frac{\Delta u_r}{\Delta x'_p} \Delta x'_p \neq 0, \quad \text{evaluated at } (\omega_1, \omega_2). \]

Also, there exists a subsequence \( n_k \) such that

\[ \lim_{k} \frac{\Delta h_{n_k}}{\Delta u_r} \text{ converges to some constant } c, \quad \text{evaluated at } (\nu_1, \nu_2). \]
Go back now to the general formulas. From * and **, using the subsequence \( n_k \), and taking limsups, we have

\[
\Delta f' = \Delta h' = \sum_j \limsup_k \frac{\Delta h_{n_k}}{\Delta u_r} \frac{\Delta u_r}{\Delta x_j} \Delta x_j'.
\]

Take the partial difference of both sides with respect to \( x_p' \) to get

\[
0 = \frac{\Delta f'}{\Delta x_p'} = \sum_j \limsup_k \frac{\Delta h_{n_k}}{\Delta u_r} \frac{\Delta u_r}{\Delta x_j} \frac{\Delta x_j'}{\Delta x_p'} = \limsup_k \frac{\Delta h_{n_k}}{\Delta u_r} \frac{\Delta u_r}{\Delta x_p'}.
\]

The left hand equality follows from corollary \( \Box \) and the right hand equality by the independence of the \( x \) variables. This relation holds for variables. Evaluate it at the fixed pair \((\nu, \nu^{(1)})\) to get the equation

\[
0 = c \cdot \frac{\Delta u_r}{\Delta x_p'} (\nu_1, \nu_2),
\]

using ****. Then *** proves \( c = 0 \) and since the subsequence \( n_k \) is arbitrary, the limit exists for the fixed \((\nu, \nu^{(1)})\). The choice of the fixed point is also arbitrary, and part (a) of the theorem follows.
References

[1] Samuel Goldberg, *Introduction to Difference Equations*, John Wiley, 1958.

[2] G. H. Hardy, *A Course of Pure Mathematics*, Cambridge University Press, 1952.

[3] G. H. Hardy and E. M. Wright, *An Introduction to the Theory of Numbers*, Oxford University Press, 1960.

[4] Richard Isaac, [arXiv:math/0512404](http://arxiv.org/abs/math/0512404), 2006.

[5] Richard Isaac, arXiv:math/1104.1616, 2011.

[6] John Lamperti, *Probability: A survey of the Mathematical Theory, second edition*, John Wiley, 1996.

[7] Ivan Niven, *Irrational Numbers*, Mathematical Association of America (Carus Mathematical Monograph), distributed by John Wiley, 1956.

Emeritus Professor

Lehman College and Graduate Center, CUNY

email: richard.isaac@lehman.cuny.edu