MATHEMATICS

ON NON-NORMAL NUMBERS

BY

C. M. COLEBROOK AND J. H. B. KEMPERMAN

(Communicated at the meeting of September 30, 1967)

1. *Introduction*

Let $s \geqslant 2$ be an integer, to be kept fixed. A real number $0 \leqslant x < 1$ is said to be normal to the base s when its expansion

$$x = \cdot\, x_1 x_2 x_3 \ldots = \sum_{c=1}^{\infty} x_c s^{-c}, \qquad (x_c \in \{0, 1, \ldots, s-1\}),$$

to the base s is such that each possible block of digits occurs with its "proper" frequency. More precisely, for each $k = 1, 2, \ldots$ and each of the s^k blocks $A = (a_1, \ldots, a_k)$ consisting of k digits $0 \leqslant a_i \leqslant s-1$, the occurrence of $(x_{c+1}, \ldots, x_{c+k}) = A$ happens with an asymptotic frequency s^{-k}, $(c = 0, 1, \ldots)$.

Let K denote the additive group of real numbers modulo one. Further $C(K)$ will denote the collection of all complex-valued continuous functions on K. It will be convenient to think of $f \in C(K)$ as a continuous function on the reals of period 1.

A sequence of points $\{u_j\}$ in K is said to have the asymptotic distribution ν when

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} f(u_j) = \nu(f) = \int_K f\, d\nu \text{ for each } f \in C(K).$$

Here, ν denotes a probability measure on K, (that is, a nonnegative measure of total mass 1). As was shown by WALL (see [5]), a number $x \in K$ is normal to the base s if and only if the corresponding sequence $\{s^j x\} = \{x, sx, s^2x, \ldots\}$ in K is uniformly distributed; that is, when $\{s^j x\}$ has the Lebesgue measure λ on K as its asymptotic distribution. More generally, a number $x \in K$ will be said to be ν-normal when the sequence $\{s^j x\}$ has the asymptotic distribution ν. Here, ν denotes a probability measure on K, necessarily invariant under the (many to one) transformation $x \to sx$ of the additive group K onto itself. The set of all such measures ν on K will be denoted by $I(s)$.

---

[1]) The second author's contribution was supported in part by the National Science Foundation, Grant GP-5801.

Naturally, it is quite possible that the sequence $\{s^j x\}$ has no asymptotic distribution at all. In general, for each $x \in K$, let $V(x, s)$ denote the collection of all accumulation points (in the weak*-topology) of the sequence of probability measures $\{\nu_1, \nu_2, \ldots\}$ defined by

$$\nu_n(f) = \frac{1}{n} \sum_{j=0}^{n-1} f(s^j x), \qquad f \in C(K).$$

As is easily seen, $V(x, s)$ is a non-empty closed and connected subset of $I(s)$. Conversely [2], given any closed and connected non-empty subset V of $I(s)$, there always exists a number $x \in K$ such that $V(x, s) = V$. In particular, given $\nu \in I(s)$, there always exists a number $x \in K$ which is ν-normal to the base s, (that is, $V(x, s) = \{\nu\}$), a result due to Pjateckii–Shapiro [6].

The question arises what can be said about the behavior of x with respect to several bases. The ultimate goal would be to characterize those sequences $\{V_s;\ s = 2, 3, \ldots\}$ for which there exists at least one $x \in K$ such that $V(x, s) = V_s$ for all s.

The bases r and s are said to be equivalent ($r \sim s$) if there exist integers m, n and $s_1 \geq 2$ with $r = s_1^m$ and $s = s_1^n$ (otherwise, $r \not\sim s$). If so then $V(x, r)$ and $V(x, s)$ are strongly related, in fact, both uniquely determine the set $V(x, s_1)$. In particular, see [7], if $x \in K$ is normal to one base then also to every equivalent base.

Conjecture. Let $\{s_1, s_2, \ldots\}$ be a given sequence of mutually non-equivalent bases. For each $q = 1, 2, \ldots$, choose V_q in an arbitrary manner as a non-empty closed and connected subset of $I(s_q)$. Then one can find at least one number $x \in K$ such that $V(x, s_q) = V_q$ for all q.

At the present, we are a far way from proving or disproving our conjecture. The strongest known result in this direction is the following result due to Schmidt [7], [8]. Choose A and B as arbitrary sets of integers ≥ 2 such that $a \not\sim b$ whenever $a \in A$ and $b \in B$. Then one can find at least one number $x \in K$ which is normal to each base $a \in A$ and simultaneously non-normal to each base $b \in B$.

In particular, there exists a number x which is non-normal to a given base s and simultaneously normal to each base $r \not\sim s$, see [7]. For $s = 3$ this result is due to Cassels [1]. It is the purpose of the present paper to prove the following related result.

Theorem 1.1. *Given the integer $s \geq 2$ and the number $x \in K$ one can always find a number $z \in K$ such that*

$$V(z, r) = V(x, r) \text{ for each } r \sim s, \tag{1.1}$$

while

$$V(z, r) = \{\lambda\} \text{ for each } r \not\sim s. \tag{1.2}$$

As an immediate consequence we have:

Theorem 1.2. *Let $s \geqslant 2$ be a given integer and $\nu \in I(s)$ a given probability measure on K. Then there exists a number $z \in K$ which is ν-normal to the base s and simultaneously λ-normal to each base $r \nsim s$.*

Proof. Choose first $x \in K$ such that $V(x, s) = \{\nu\}$, and then apply Theorem 1.1.

Our proof of Theorem 1.1 is closely related to the proof of Schmidt [7]; see also [8] and [9].

2. *Preliminaries*

Let $x \in K$ be a given number, $s \geqslant 2$ a given base. As is easily seen, there exists a unique integer $s_1 \geqslant 2$ such that $r \sim s$ if and only if $r = s_1{}^m$ for some positive integer m. In proving Theorem 1.1, we may as well assume that $s = s_1$ in which case (1.1) is equivalent to

$$V(z, s^m) = V(x, s^m) \text{ for all } m = 1, 2, \ldots \tag{2.1}$$

A sufficient condition for (2.1) is that

$$\lim_{n \to \infty} n^{-1} \sum_{j=0}^{n-1} (f(s^{jm} z) - f(s^{jm} x)) = 0, \text{ for } f \in C(K);\ m = 1, 2, \ldots \tag{2.2}$$

Let

$$x = \sum_{c=1}^{\infty} x_c s^{-c}, \quad z = \sum_{c=1}^{\infty} z_c s^{-c}, \qquad (x_c, z_c \in \{0, 1, \ldots, s-1\}). \tag{2.3}$$

Let further $N(n)$ denote the number of $c = 1, \ldots, n$ with $z_c \neq x_c$. A sufficient condition for (2.2) is that

$$N(n) = o(n) \text{ as } n \to \infty, \tag{2.4}$$

as follows easily from the uniform continuity of the $f \in C(K)$.

Consider a fixed sequence $\{\varepsilon_c; c = 1, 2, \ldots\}$ such that

$$\begin{cases} \varepsilon_c = +1 \text{ if } 0 \leqslant x_c \leqslant s-2, \\ = -1 \text{ if } \quad x_c = s-1. \end{cases} \tag{2.5}$$

Next, let $\{d_c\}$ be a fixed sequence satisfying

$$0 < d_{c+1} \leqslant d_c \leqslant \tfrac{1}{2}; \quad \lim_{c \to \infty} d_c = 0, \tag{2.6}$$

$(c = 1, 2, \ldots)$. Finally, let $y_1, y_2, \ldots$ be *independent* random variables, y_c having the distribution defined by

$$y_c \in \{0, \varepsilon_c\}, \quad Pr(y_c = \varepsilon_c) = d_c. \tag{2.7}$$

Lemma 2.1. *The number $z \in K$ defined by*

$$z = x + y, \quad y = \sum_{c=1}^{\infty} y_c s^{-c} \tag{2.8}$$

satisfies condition (2.1) *with probability* 1.

Proof. Let $z_c = x_c + y_c$. By (2.5) and (2.7), we have that $z_c \in \{0, 1, \ldots, s-1\}$ for all c. Moreover, $\sum_{c=1}^{\infty} z_c s^{-c} = z$, thus, we have the situation (2.3) with $z_c - x_c = y_c \in \{0, \varepsilon_c\}$. It suffices to show that (2.4) holds with probability 1, equivalently, that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{c=1}^{n} |y_c| = 0 \text{ with probability 1.}$$

This follows immediately from $E(|y_c|) = d_c \to 0$ and the following classical criterion due to Kolmogorov, see [4] p. 238, 253, 259.

Lemma 2.2. *Let $\{U_j\}$ be a given sequence of complex-valued independent random variables such that $|U_j| \leqslant 1$. Then*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} E(U_j) = 0 \text{ implies that } \lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} U_j = 0,$$

with probability 1. *(The converse is obvious.)*

From now on, the random variable $z = x + y$ will be as in (2.8). For each base r, let D_r denote the set of numbers which are non-normal to the base r. In view of Lemma 2.1, it suffices to prove that for each fixed base $r \not\sim s$ we have $z \notin D_r$ with probability 1. At first sight, this might seem like an easy problem since the set D_r has Lebesgue measure zero. However, also the support S_y of the random variable y (and hence $S_z = x + S_y$) is a set of Lebesgue measure 0. For $s \geqslant 3$ this assertion is rather obvious (y_c having only two possible values); if $s = 2$ the assertion can easily be deduced from Lemma 2.1 and the fact that D_s has Lebesgue measure zero. For a related result, see [3].

Let us introduce the random variables

$$U\{w\} = e^{2\pi i w z} = e^{2\pi i w(x+y)}, \qquad (w = 0, \pm 1, \pm 2, \ldots),\ U\{-w\} = \overline{U\{w\}}. \tag{2.9}$$

Lemma 2.3. *Suppose that, for each choice of the base $r \not\sim s$ and each choice of the positive integer h, we have*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} U\{hr^j\} = 0, \text{ with probability 1.} \tag{2.10}$$

Then z satisfies (1.2) *with probability* 1.

Proof. Consider a fixed base $r \not\sim s$. By (2.9) and (2.10),

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} f(r^j z) = \int_K f\, d\lambda, \tag{2.11}$$

with probability 1, whenever f is a trigonometric polynomial

$$f(v) = \sum_{h=-H}^{H} b_h e^{2\pi i h v}.$$

By Weyl's [10] criterion (the trigonometric polynomials being dense in

$C(K)$ with the supremum norm), we have with probability 1 that (2.11) holds for each $f \in C(K)$, in other words, that $V(z, r) = \{\lambda\}$.

The random variables $U_j = U(hr^j)$ $(j = 1, 2, \ldots)$ occurring in (2.10) are clearly *not* independent. Thus, Lemma 2.2 is of no use in establishing results of the type (2.10). Instead, we shall use:

Lemma 2.4. *Let $\{U_j\}$ be a sequence of complex-valued random variables such that $|U_j| \leqslant 1$. Suppose further that there exist constants C and $\gamma > 1$ such that*

$$(2.12) \qquad E\left(\left|\frac{1}{n}(U_1 + \ldots + U_n)\right|^2\right) \leqslant C (\log n)^{-\gamma} \text{ for all } n = 1, 2, \ldots$$

Then $\lim\limits_{n\to\infty} \frac{1}{n} \sum\limits_{j=1}^{n} U_j = 0$ *with probability* 1.

Proof. Choose the positive constants η and δ such that $\eta + \delta < \gamma - 1$, and put $1 + \eta + \delta = \alpha\gamma$, thus, $0 < \alpha < 1$. Let further $n_k = 1 + [\exp k^\alpha]$, $(k = 1, 2, \ldots)$, $n_k \uparrow +\infty$. Let A_k denote the event defined by

$$\left|\frac{1}{n_k}(U_1 + \ldots + U_{n_k})\right|^2 > k^{-\eta}.$$

Then, using (2.12),

$$Pr(A_k) \leqslant k^\eta E\left(\left|\frac{1}{n_k}(U_1 + \ldots + U_{n_k})\right|^2\right) \leqslant k^\eta C (\log n_k)^{-\gamma} \leqslant C k^{-1-\delta}.$$

It follows that $\sum Pr(A_k) < \infty$ so that (with probability 1) A_k will happen for only finitely many k. In particular,

$$\lim_{k\to\infty} \frac{1}{n_k}(U_1 + \ldots + U_{n_k}) = 0,$$

with probability 1. This yields the stated assertion since $|U_j| \leqslant 1$ and $n_{k+1}/n_k \to 1$.

Combining the Lemmas 2.3 and 2.4, we have

Lemma 2.5. *Suppose that, for each choice of the base $r \nsim s$ and each choice of the positive integer h, one can find constants C and $\gamma > 1$ such that*

$$(2.13) \qquad \frac{1}{n^2} \sum_{j=1}^{n} \sum_{k=1}^{n} |E(U\{h(r^j - r^k)\})| \leqslant C (\log n)^{-\gamma} \qquad \text{for all } n = 1, 2, \ldots$$

Then z satisfies (1.2) *with probability* 1.

Thus, also in view of Lemma 2.1, it suffices for the proof of Theorem 1.1 to exhibit at least one sequence $\{d_c\}$ for which the conditions of Lemma 2.5 are fulfilled.

3. *Upper bound on $E(U)$.*

Here, and further on, w will denote an integer. We have

$$U\{w\} = e^{2\pi i w z} = e^{2\pi i w x} e^{2\pi i w y},$$

where x is a real constant. Further, $y = \sum_{c=1}^{\infty} y_c s^{-c}$ with the y_c as independent random variables. In fact, for θ real,

$$|E(e^{2\pi i\theta y_c})| = |(1-d_c)+d_c e^{2\pi i\theta\varepsilon_c}| = [1-4d_c(1-d_c)\sin^2\pi\theta]^{\frac{1}{2}} \leqslant$$
$$\leqslant \exp[-2d_c(1-d_c)\sin^2\pi\theta] \leqslant \exp[-d_c\sin^2\pi\theta],$$

since $\varepsilon_c = \pm 1$ and $0 < d_c \leqslant \frac{1}{2}$. We conclude that

$$|E(U\{w\})| = |\prod_{c=1}^{\infty} E(\exp(2\pi i w s^{-c} y_c))| \leqslant \exp[-\sum_{c=1}^{\infty} d_c \sin^2 \pi w s^{-c}].$$

This in turn yields

$$|E(U\{w\})| \leqslant \exp[-\sum_{c=1}^{\infty} t_c \phi(ws^{-c})], \tag{3.1}$$

where

$$t_c = d_c \sin^2 \pi s^{-2},$$

while ϕ denotes the function on the reals defined by

$$\left.\begin{aligned} \phi(\theta) &= 1 \text{ if } s^{-2} \leqslant \theta - [\theta] \leqslant 1 - s^{-2}, \\ &= 0, \text{ otherwise,} \end{aligned}\right\} \tag{3.2}$$

(with $[\theta]$ as the integral part of θ). Observe that $\phi(\theta)=0$ when θ is an integer and also when $|\theta| < s^{-2}$. Moreover, $\varphi(\theta) = \varphi(-\theta)$; $\varphi(\theta+1) = \varphi(\theta)$.

Let us further introduce

$$\Phi(w) = \sum_{c=1}^{\infty} \phi(ws^{-c}) = \sum_{c=-\infty}^{+\infty} \phi(ws^{-c}), \text{ thus, } \Phi(sw) = \Phi(w) = \Phi(-w) \geqslant 0. \tag{3.3}$$

Assuming $w > 0$, consider the expansion

$$w = \ldots w_2 w_1 w_0 = \sum_{c=0}^{\infty} w_c s^c, \qquad w_c \in \{0, 1, \ldots, s-1\}, \tag{3.4}$$

with only finitely many w_c non-zero. Observe that

$$ws^{-c} - [ws^{-c}] = \sum_{j=1}^{c} w_{c-j} s^{-j} = 0 \cdot w_{c-1} w_{c-2} \ldots w_0,$$

hence, $\phi(ws^{-c}) = 1$ when the pair of digits (w_{c-1}, w_{c-2}) is *good* in the sense that it is distinct from both pairs $(0, 0)$ and $(s-1, s-1)$, (a terminology due to Schmidt). Consequently, if $w > 0$ then $\Phi(w)$ *is not smaller than the number of good pairs* (w_{c-1}, w_{c-2}) in the expansion (3.4) of w to the base s, $(c = 1, 2, \ldots;\ w_c = 0$ if $c \leqslant -1)$.

Lemma 3.1. *We have for each integer* w *that*

$$|E(U\{w\})| \leqslant \exp[-\beta(w)\Phi(w)]. \tag{3.5}$$

Here, $\beta(w)$ *is the function defined by*

$$\beta(w) = d_m \sin^2 \pi s^{-2} \text{ when } s^{m-2} \leqslant |w| < s^{m-1}, \qquad (m = 2, 3, \ldots;\ \beta(0) = 0). \tag{3.6}$$

Proof. Given $w>0$, let $m \geqslant 2$ denote the unique integer such that $s^{m-2} \leqslant w < s^{m-1}$. Then $c \geqslant m+1$ would imply that $ws^{-c} < s^{m-1-c} \leqslant s^{-2}$, hence, $\phi(ws^{-c})=0$. Therefore, $\{t_c\}$ being non-increasing,

$$\sum_{c=1}^{\infty} t_c \phi(ws^{-c}) = \sum_{c=1}^{m} t_c \phi(ws^{-c}) \geqslant t_m \Phi(w) = \beta(w)\Phi(w),$$

by (3.3). Thus (3.1) implies (3.5).

4. *Proof of Theorem* 1.1.

Let s be a fixed base and let Φ be as in Section 3; clearly, Φ depends on s. It will be convenient to introduce the following property.

Property A. A function $\omega(x)$ on $[1, +\infty)$ will be said to have Property A when

(i) $\omega(x)$ tends to $+\infty$ in a non-decreasing manner as x tends to $+\infty$.

(ii) For each base $r \not\sim s$ and each positive integer h one can find constants $C>0$ and $\gamma>1$ such that, for all large n,

$$\# \{(j, k): 1 \leqslant j, k \leqslant n, \Phi(hr^j - hr^k) \leqslant \omega(n) \log\log n\} \leqslant Cn^2(\log n)^{-\gamma}. \tag{4.1}$$

Lemma 4.1. *Let $\omega(x)$ be any function satisfying Property A. Then Theorem* 1.1 *holds; more precisely, under the choice*

$$d_c = \min\{\tfrac{1}{2}, \eta/\omega(\sqrt{c})\}, \quad (c=1, 2, \ldots), \tag{4.2}$$

of $\{d_c\}$ we have with probability 1 *that the random number $z=x+y$ satisfies both* (1.2) *and* (2.1). *Here, η denotes any positive constant such that $\eta > 1/\sin^2 \pi s^{-2}$.*

Proof. Choose $\{d_c\}$ as in (4.2). Let $h \geqslant 1$ and $r \geqslant 2$ be given integers such that $r \not\sim s$. It suffices to show that (2.13) holds for some choice of the constants C and $\gamma>1$. In view of (3.5) and (4.1) it suffices to show that, for some $\gamma>1$, we have

$$\exp[-\beta(hr^n)\omega(n)\log\log n] = 0((\log n)^{-\gamma}) \text{ as } n \to \infty.$$

Equivalently, we must have that

$$\liminf_{n\to\infty} [\omega(n)\beta(hr^n)] > 1. \tag{4.3}$$

Put $K=\eta \sin^2 \pi s^{-2}$, thus, $K>1$. By (3.6) and (4.2) we have for n sufficiently large that $\beta(hr^n) = K/\omega(\sqrt{m})$. Here, m is the integer defined by $s^{m-2} \leqslant hr^n < s^{m-1}$. Hence, for n sufficiently large we have $\sqrt{m} \leqslant n$, thus, $\omega(\sqrt{m}) \leqslant \omega(n)$, yielding (4.3). This completes the proof of Lemma 4.1.

Theorem 1.1 is now obtained by invoking the following result. It implies that any function $\omega(x)$ satisfying $\omega(x) \uparrow +\infty$ and

$$\omega(x) = o(\log x/\log\log x), \quad \text{as } x \to +\infty,$$

does have Property A. Actually, Lemma 4.2 is much stronger than necessary for our purpose and it would be of interest to find a *simple* proof of the fact that there exists at least one function having Property A.

Lemma 4.2. *For each choice of the positive integers h and $r \geqslant 2$, $r \not\sim s$, one can find positive constants C, α and δ such that, for all $n = 1, 2, \ldots$,*

$$\# \{(j, k) : 1 \leqslant j, k \leqslant n, \Phi(hr^j - hr^k) \leqslant \alpha \log n\} \leqslant Cn^{2-\delta}. \tag{4.4}$$

The proof of Lemma 4.2 is analogous to a proof in [7] pp. 665–669. The following is a quick sketch in several steps of a proof of Lemma 4.2 which may be regarded as a simplified version of the implicit proof contained in [7]. Lemma 4.2 will be reduced to:

Lemma 4.3. *Let h and $r \geqslant 2$ be positive integers such that at least one prime divisor p of s is not divisible on r. Then there exist positive constants C, α and δ such that the inequality*

$$\# \{j = 1, \ldots, n : \Phi_N(hr^j + u) \leqslant \alpha \log n\} \leqslant Cn^{1-\delta} \tag{4.5}$$

holds for each choice of the integers $n \geqslant 1$ and u. Here, the integer N is defined by $s^{N-1} < n \leqslant s^N$.

Further, the function Φ_N is defined by

$$\Phi_N(w) = \sum_{c=1}^{N} \phi(ws^{-c}), \qquad (N = 1, 2, \ldots). \tag{4.6}$$

If w is a positive integer as in (3.4) then $\Phi_N(w)$ is easily seen to be no smaller than the *number of good pairs* (w_{c-1}, w_{c-2}) with $1 \leqslant c \leqslant N$, $(w_{-1} = 0)$. From the properties of the function ϕ,

$$\Phi_N(-w) = \Phi_N(w) \leqslant \Phi(w).$$

Moreover, $\Phi_N(w + bs^m) = \Phi_N(w)$ as soon as b and m are integers with $m \geqslant N$. It follows that

$$\Phi(ws^\lambda + ys^\mu) = \Phi(w + ys^{\mu-\lambda}) \geqslant \Phi_N(w), \tag{4.7}$$

provided y, λ and μ are integers satisfying $\mu - \lambda \geqslant N$.

Step (i). We assert that Lemma 4.2 is a consequence of Lemma 4.3. Namely, applying (4.5) with $u = -hr^k$ and summing over $k = 1, \ldots, n$, one obtains (4.4) whenever some prime divisor of s is not divisible on r.

It remains to consider the case that each prime divisor of s is also a prime divisor of r. Let r and s have factorizations

$$r = p_1^{\varrho_1} \ldots p_k^{\varrho_k};\ s = p_1^{\sigma_1} \ldots p_k^{\sigma_k} \text{ with } \frac{\sigma_k}{\varrho_k} \leqslant \ldots \leqslant \frac{\sigma_1}{\varrho_1},$$

and all ϱ_i positive. Then $R = r^{\sigma_1}/s^{\varrho_1}$ is an integer with $R \geqslant 2$; (if $R = 1$

then $r \sim s$). Further, p_1 is a prime dividing s but not R. It follows from Lemma 4.3 that

$$(4.8) \qquad \# \{\lambda = 1, \dots, m : \Phi_M(hr^q R^\lambda) \leqslant \alpha \log m\} \leqslant C m^{1-\delta},$$

for each choice of the integers $m \geqslant 1$ and $q = 0, 1, \dots, \sigma_1 - 1$. Here, C, α and δ denote positive constants while $s^{M-1} < m \leqslant s^M$. Thus, $M \sim \log m / \log s$ when m is large.

In proving (4.4), consider a pair of integers j and k with $1 \leqslant j \leqslant k \leqslant n$ and write

$$j = \lambda\sigma_1 + q, \quad k = \mu\sigma_1 + q' \text{ with } 0 \leqslant q, \ q' < \sigma_1,$$

($0 \leqslant \lambda$, $\mu \leqslant n/\sigma_1$ and $\lambda \leqslant \mu$). Then one has

$$hr^j - hr^k = [hr^q R^\lambda] s^{\varrho_1 \lambda} - [hr^{q'} R^\mu] s^{\varrho_1 \mu}.$$

Hence, using (4.7),

$$\Phi(hr^j - hr^k) \geqslant \Phi_M(hr^q R^\lambda) \text{ provided } (\mu - \lambda)\varrho_1 \geqslant M.$$

The latter is true for all but $0(nM)$ pairs $1 \leqslant j, \leqslant k \leqslant n$. Applying (4.8) with $m = [n/\sigma_1]$ (thus $M = 0(\log n)$) and summing over q, one obtains a result of the type (4.4).

Step (ii). It remains to prove Lemma 4.3. From now on $h \geqslant 1$, $r \geqslant 2$ and p are fixed integers such that p is a prime dividing s but not r. Let o_k denote the order of r modulo p^k, that is, the smallest positive integer with $r^m \equiv 1 \pmod{p^k}$. We assert that, for some positive constant ε,

$$(4.9) \qquad o_k \geqslant \varepsilon p^k \text{ for all } k > 0.$$

First observe that, for $c \geqslant 1$,

$$a \equiv 1 + qp^c (\text{mod } p^{c+1}) \text{ implies } a^p \equiv 1 + qp^{c+1} (\text{mod } p^{c+2})$$

unless both $c = 1$ and $p = 2$. Let $p \geqslant 3$ and consider

$$r^{(p-1)p^j} \equiv 1 + qp^{c+j} \not\equiv 1 (\text{mod } p^{c+j+1}).$$

It holds for $j = 0$ with a unique maximal $c \geqslant 1$ and q prime to p. By induction, it holds for all $j \geqslant 0$. Hence, $o_{c+j+1} > p^j$ for all $j \geqslant 0$, proving (4.9) when $p \geqslant 3$. If $p = 2$ one uses instead

$$2^{2^{1+j}} \equiv 1 + 2^{c+j} \not\equiv 1 (\text{mod } 2^{c+j+1}).$$

Step (iii). Define g as the largest integer such that p^g divides h. Consider a pair of distinct non-negative integers j_1 and j_2. By (4.9), we have $|j_1 - j_2| \geqslant \varepsilon p^{k-g}$ as soon as $hr^{j_1} \equiv hr^{j_2} \pmod{p^k}$, hence, as soon as $hr^{j_1} \equiv hr^{j_2}$ (mod s^k). Consequently, introducing

$$(4.10) \qquad N_k(t) = \# \{j = 1, \dots, s^k : hr^j \equiv t \ (\text{mod } s^k)\},$$

we have the upperbound

$$N_k(t) \leqslant 1 + s^k(\varepsilon p^{k-g})^{-1} \leqslant 1 + (p^g/\varepsilon)(s/2)^k, \tag{4.11}$$

holding for each choice of the positive integer k and the residue class $t = 0, 1, \ldots, s^k - 1$.

Step (iv). For $k = 1, 2, \ldots,$ consider the function Ψ_k with domain $G_k = \{0, 1, \ldots, s^k - 1\}$ defined as follows. Let $t \in G_k$ have the expansion

$$t = t_0 + t_1 s + \ldots + t_{k-1} s^{k-1}, \quad t_i \in \{0, 1, \ldots, s-1\}. \tag{4.12}$$

Then

$$\Psi_k(t) = \# \{i = 1, \ldots, k-1 : (t_i, t_{i-1}) \neq (0, 0), (s-1, s-1)\}. \tag{4.13}$$

Consider further the quantity

$$M_k(b) = \# \{t \in G_k : \Psi_k(t) \leqslant bk\}, \tag{4.14}$$

where b is a positive parameter. We assert that for each positive number $\varrho > \frac{1}{2}$ there exists a positive number $b_0(\varrho)$ such that

$$M_k(b) = O(2^{\varrho k}) \text{ as } k \to \infty, \text{ as soon as } 0 < b < b_0(\varrho). \tag{4.15}$$

One proof based on Stirling's formula may be found in **[7]** p. 667. A second proof would be as follows.

Let k be fixed, $m = [k/2]$ so that $k = 2m + q$ with $q = 0$ or 1. Let $f(t)$ denote the function on G_k defined as in (4.13) but with i restricted to the odd integers $i = 1, 3, \ldots, 2m-1$, (so that the pairs counted do not overlap). In particular, $f(t) \leqslant \Psi_k(t)$. As is easily seen,

$$\sum_{t \in G_k} u^{f(t)} = s^q \prod_{i=1}^{m} [1 + 1 + u + \ldots + u] = s^q [2 + (s^2 - 2) u]^m.$$

Here, u is an auxiliary variable. Assuming that $0 < u < 1$ we have that $\Psi_k(t) \leqslant bk$ implies $u^{f(t)} \geqslant u^{bk}$. Hence, by (4.14),

$$M_k(b) \leqslant (s/u^b)^q [2u^{-2b} + (s^2 - 2) u^{1-2b}]^m \text{ for each } 0 < u < 1.$$

By choosing b as a sufficiently small number and $u = b$, the quantity $[\cdot]$ can be brought arbitrarily close to 2, (since $x^{-x} \to 1$ as $x \downarrow 0$). This proves the assertion (4.15).

Step (v). End of proof of Lemma 4.3. It suffices to establish (4.5) for n of the form $n = s^k$, $(k = 1, 2, \ldots)$. In this case $N = k$ and $\alpha \log n = bk$ where $b = \alpha \log s$.

Observe that $\Phi_k(w)$ is periodic of period s^k. Hence, if $w \equiv t \pmod{s^k}$ with $t \in G_k$ then $\Phi_k(w) = \Phi_k(t) \geqslant \Psi_k(t)$ by (4.13) and the remark following (4.6). Therefore, by (4.10), the left hand side of (4.5) (with $n = s^k$ and $N = k$) has the upperbound $\Sigma' N_k(t)$ where we sum over those $t \in G_k$ for which $\Psi_k(t+u) \leqslant bk$; here $t+u$ is to be interpreted modulo s^k. Moreover, by (4.11) and (4.14), we have the upperbound (independent of u):

$$\sum_t{}' N_k(t) \leqslant M_k(b)[1+(p^g/\varepsilon)(s/2)^k] = O(2^{-(1-\varrho)k}\, s^k),$$

as soon as $0 < b < b_0(\varrho)$, by (4.15). Here, ϱ can be any number with $\frac{1}{2} < \varrho < 1$. Consequently, we have for each $\delta < \log 2/\log s^2$ that (4.5) holds with a suitable constant C (depending on h and r) as soon as $b = \alpha \log s$ is sufficiently small, $0 < \alpha < \alpha_0(\delta)$, where $\alpha_0(\delta)$ is independent of h and r.

The University of Rochester

References

1. Cassels, J. W. S., On a problem of Steinhaus about normal numbers, Colloquium Math., 7, 95–101 (1959).
2. Colebrook, C. M., On the Hausdorff dimension of certain sets of non-normal numbers, forthcoming.
3. Kakutani, S., On equivalence of infinite product measures, Ann. Math., **49**, 214–224 (1948).
4. Loève, M., Probability Theory, second edition, (Princeton, 1960).
5. Niven, I., Irrational Numbers, Carus Mathematical Monograph no. 11, (New York, 1956).
6. Pjateckii-Shapiro, I. I., On the laws of distribution of the fractional part of an exponential function, (Russian), Izv. Akad. Nauk SSSR, Ser. Mat., 15, 47–52 (1951).
7. Schmidt, W., On normal numbers, Pacific J. Math., **10**, 661–672 (1960).
8. ———, Über die Normalität von Zahlen zu verschiedenen Basen, Acta Arith., 7, 299–309 (1962).
9. ———, Normalität bezüglich Matrizen, J. reine angew. Math., **214/215**, 227–260 (1964).
10. Weyl, H., Über die Gleichverteilung von Zahlen mod 1, Math. Ann., **77**, 313–352 (1916).