PACIFIC JOURNAL OF MATHEMATICS
Vol. 76, No. 1, 1978

CONSTRUCTION OF GENERALIZED NORMAL NUMBERS

F. J. MARTINELLI

Let x be a real number, $0 \leqq x < 1$, and let $0.x_1x_2 \cdots$ be its expansion in the base B. Let $N(b, n)$ be the number of occurrences of the digit b in x up to x_n. Then x is called *digit normal* (in the base B) if

$$\lim_{n\to\infty} \frac{N(b, n)}{n} = \frac{1}{B}$$

for each of the B possible values of b. Let γ be any fixed B-ary sequence of length L and $N(\gamma, n)$ be the number of indices k for which $x_kx_{k+1} \cdots x_{k+L-1}$ is γ, that is, $N(\gamma, n)$ is the number of times γ appears in the first n digits of x. Then x is *normal* (in the base B) if

$$\lim_{n\to\infty} \frac{N(\gamma, n)}{n} = B^{-L}$$

for each of the B^L possible sequences γ, and B^{-L} is called the limiting frequency of γ in x.

The purpose of this paper is to construct a generalized normal number (in the base 2) in which these frequencies are weighted. For example, we will obtain infinite binary decimals in which the limiting frequency of occurrence of ones is 1/3 (in general, $p < 1$) rather than 1/2; consequently, any binary string γ of length L will have limiting frequency

$$(1/3)^K(2/3)^{L-K}$$

where K is the number of ones in γ.

For simplicity, the construction will be in the base 2, since generalization to the integer base B is straightforward and need be only briefly described.

Borel [1] proved that almost every number x, relative to Lebesque measure, is normal in the base 10. The simplest construction of a normal number is due to Champernowne [2] who showed that if the natural numbers are arranged in increasing order, the resulting decimal

$$0.12345678910111213 \cdots$$

is normal in the base 10. Copeland and Erdös [3] showed that certain subsequences of the natural numbers, arranged as above, are also normal; in particular for the sequence of primes,

$$0.13571113 \cdots$$

is normal in the base 10. Ito and Shiokawa [4] generalized Champernowne's construction to arbitrary real bases.

A generalized normal number may be defined as follows. Let $0 < p < 1$ and $q = 1 - p$. The infinite binary decimal x is called *digit p-normal* if

$$\lim_{n \to \infty} \frac{N(1, n)}{n} = p .$$

Any binary sequence of finite length will be called a string. If γ is a string of length L containing K ones and J zeros, then x is *L-digit p-normal* if

$$\lim_{n \to \infty} \frac{N(\gamma, n)}{n} = p^K q^J$$

for every such γ. Finally, x is *p-normal* if it is L-digit p-normal for every positive integer L; normality in the previous sense is the case $p = 1/2$.

In [6] Postnikov and Pjateckii-Sapiro construct a generalized normal number by adding certain digits to the Champernowne sequence and use a criterion obtained by the latter author [8] to prove p-normality. In this paper, a much simpler construction is described and the proof that the resulting decimal is p-normal proceeds from first principles.

Generalized normal numbers may be used to generate sequences of random variates having specified distribution functions (see [7] and [5] and the references therein). In [5] Knuth constructs a 1/2-normal number in the base 2 (which would yield uniformly distributed random variates) but points out that the construction is not of "practical value as a computer method for random number generation." The algorithm described here for p-normal numbers is easily programmed on a digital computer.

Construction. For any integers k, n with $0 \leqq k \leqq n$, let $A(k, n)$ be the concatenation in lexicographic order of all possible strings of length n having exactly k ones. Denote the length of a sequence S by $|S|$; for example $|A(k, n)| = n\binom{n}{k}$. If $\{S_i\}$ is a sequence of strings, let $S_1 S_2 S_3 \cdots$ be the concatenation of these strings in order. If $S_i = S$ for $1 \leqq i \leqq m$, denote $S_1 S_2 \cdots S_m$ by (mS).

THEOREM. *Let $0 < p < 1$ and let $k_n = [pn]$ for each $n = 1, 2 \cdots$. Let $A_n = A(k_n, n)$. Then*

$$x = 0.A_1(2A_2)(3A_3) \cdots$$

is p-normal.[1] (*Here* [.] *is the greatest integer function.*)

Proof. Let γ be any fixed string of length L containing K ones and J zeros. Let $B = b_1 b_2 \cdots b_{|B|}$ be any string and let B' be B concatenated on the right with $L - 1$ zeros. Define $N(\gamma, B)$ to be the number of indices k for which the sequence $b'_k b'_{k+1} \cdots b'_{k+L-1}$ in B' equals γ. Let $N_0(\gamma, A_n)$ denote the number of substrings $b_k b_{k+1} \cdots b_{k+L-1}$ in A_n which are equal to γ and for which $k \equiv i (\mathrm{mod}\ n)$, where $1 \leqq i \leqq n - L + 1$. An elementary counting argument shows that

$$N_0(\gamma, A_n) = (n - L + 1)\binom{n - L}{k_n - K}$$

and that

$$N_0(\gamma, A_n) \leqq N(\gamma, A_n) \leqq N_0(\gamma, A_n) + \binom{n}{k_n}(L - 1) .$$

To complete the proof of the theorem, we will need the following lemmas.

LEMMA 1.

$$\lim_{n \to \infty} \frac{N(\gamma, A_n)}{|A_n|} = p^K q^J .$$

Proof. Let $j_n = n - k_n$ and assume that $k_n \geqq K$ and $j_n \geqq J$. Then, since

$$|A_n| = n\binom{n}{k_n} ,$$

$$\frac{N_0(\gamma, A_n)}{|A_n|} = \frac{n - L + 1}{n} \frac{(n - L)!}{n!} \frac{k_n!}{(k_n - K)!} \frac{j_n!}{(j_n - J)!}$$

$$= \frac{1}{n} \frac{\prod_{\alpha=0}^{K-1}(k_n - \alpha) \prod_{\alpha=0}^{J-1}(j_n - \alpha)}{\prod_{\alpha=0}^{L-2}(n - \alpha)} .$$

Now let $p_n = k_n/n$ and $q_n = j_n/n$; then

$$|p_n - p| = \left|\frac{[pn]}{n} - \frac{pn}{n}\right| = \frac{1}{n}|[pn] - pn| \leqq 1/n$$

and $p_n \to p$. Hence, $q_n \to q$. Then

[1] The lexicographic ordering in the $A(k_n, n)$ strings is not needed in the proof; in fact the theorem is true for any ordering so that for each p, the method yields an uncountable family of p-normal numbers.

$$\frac{N_0(\gamma, n)}{|A_n|} = \frac{1}{n} \frac{\prod_{\alpha=0}^{K-1}(np_n - \alpha)\prod_{\alpha=0}^{J-1}(nq_n - \alpha)}{\prod_{\alpha=0}^{L-2}(n - \alpha)}$$

$$= \frac{(np_n)^K \prod_{\alpha=0}^{K-1}\left(1 - \frac{\alpha}{np_n}\right)\cdot(nq_n)^J \prod_{\alpha=0}^{J-1}\left(1 - \frac{\alpha}{nq_n}\right)}{n^L \prod_{\alpha=0}^{L-2}\left(1 - \frac{\alpha}{n}\right)}$$

$$= p_n^K q_n^J \cdot \frac{\prod_{\alpha=0}^{K-1}\left(1 - \frac{\alpha}{np_n}\right)\prod_{\alpha=0}^{J-1}\left(\frac{\alpha}{nq_n}\right)}{\prod_{\alpha=0}^{L-2}\left(1 - \frac{\alpha}{n}\right)}$$

and

$$\lim_{n\to\infty} \frac{N_0(\gamma, A_n)}{|A_n|} = p^K q^J .$$

Furthermore,

$$\frac{N_0(\gamma, A_n)}{|A_n|} \leqq \frac{N(\gamma, A_n)}{|A_n|} \leqq \frac{N_0(\gamma, A_n)}{|A_n|} + \frac{L-1}{n}$$

and

$$\lim_{n\to\infty} \frac{L-1}{n} = 0 ;$$

hence the lemma is proved.

We will need the following lemma, whose proof is omitted.

LEMMA 2. *Let $\{a_n\}$ be a sequence of real numbers which converges to a real number a, and let $\{b_n\}$ be a sequence of positive numbers such that $\sum b_n = \infty$. Then*

$$\lim_{n\to\infty} \frac{a_1 b_1 + \cdots + a_n b_n}{b_1 + \cdots + b_n} = a .$$

LEMMA 3.

$$\overline{\lim_{n\to\infty}} \frac{|A_{n+1}|}{|A_n|} = \max(p^{-1}, q^{-1}) .$$

Proof. From the definitions,

$$|A_n| = n\begin{bmatrix} n \\ k_n \end{bmatrix} = \frac{nn!}{k_n!\,j_n!} = \frac{nn!}{[pn]!(n - [pn])!}$$

and

$$\frac{|A_{n+1}|}{|A_n|} = \left(1 + \frac{1}{n}\right)(n+1)\frac{[pn]!(n-[pn])!}{[p(n+1)]!(n+1-[p(n+1)])!} .$$

For each value of n, $[p(n+1)]$ assumes one of the values $[pn]$ or $[pn]+1$. In the first case,

$$\frac{|A_{n+1}|}{|A_n|} = \left(1 + \frac{1}{n}\right)\left(1 - \frac{[pn]}{n+1}\right)^{-1} \longrightarrow (1-p)^{-1} = q^{-1}$$

and in the second case,

$$\frac{|A_{n+1}|}{|A_n|} = \left(1 + \frac{1}{n}\right)\left(\frac{[pn]}{n+1} + \frac{1}{n+1}\right)^{-1} \longrightarrow p^{-1}$$

as $n \to \infty$. This proves the lemma.

Returning to the proof of the theorem, let $\{a_{m,k}\}$ be the increasing sequence of positive integers defined by

$$a_{m,k} = \sum_{i=1}^{m-1} i|A_i| + k|A_m|; \qquad m \geqq 1,\ 1 \leqq k \leqq m .$$

Let b_l denote the lth integer in the sequence $\{a_{m,k}\}$. Let m_0 be the smallest integer for which A_{m_0} has a least $L-1$ leading zeros and let $b_{l_0} = a_{m_0,1}$. Then for any $l \geqq l_0$,

$$N(\gamma, b_l) = N(\gamma, b_{l_0-1}) + \sum_{i=m_0}^{m-1} if_i|A_i| + f_m k|A_m|$$

where $f_i = N(\gamma, A_i)/|A_i|$ and $b_1 = a_{m,k}$. Hence, by Lemmas 1 and 2,

$$\lim_{l\to\infty} \frac{N(\gamma, b_l)}{b_l} = p^K q^J .$$

Now let t be an integer, $b_l \leqq t \leqq b_{l+1}$ for some $b_l = a_{m,k}$ Then

$$\frac{N(\gamma, t)}{t} \leqq \frac{N(\gamma, b_{1+1})}{b_l} \leqq \frac{N(\gamma, b_1)}{b_l} + \frac{|A_m|}{b_l}$$

and

$$\frac{N(\gamma, t)}{t} \geqq \frac{N(\gamma, b_l)}{b_{l+1}} \geqq \frac{N(\gamma, b_{l+1})}{b_{l+1}} - \frac{|A_m|}{b_{l+1}} .$$

Since

$$\frac{|A_m|}{b_{l+1}} \leqq \frac{|A_m|}{b_l} = \frac{|A_m|}{\sum_{i=1}^{m-1} i|A_i| + k|A_m|} \leqq \frac{|A_m|}{(m-1)|A_{m-1}|} ,$$

the theorem follows from the fact that $|A_m|/|A_{m-1}|$ is bounded.

For the base $B > 2$, there must be specified B positive real numbers $p_0, \cdots, p_{B-1}$ such that

$$\sum_{i=0}^{B-1} p_i = 1 .$$

A number x may be constructed as in the theorem using sequences

$$A_n = A(k_n^0, k_n^1, \cdots, k_n^{B-1}, n)$$

of length n containing k_n^b b's for each digit $b = 0, 1, \cdots, B-1$. The k_n^b are determined recursively by

$$\begin{aligned} k_n^0 &= [p_0 n] \\ k_n^0 + k_n^1 &= [(p_0 + p_1)n] \\ &\vdots \\ k_n^0 + \cdots + k_n^{B-1} &= [(p_0 + \cdots + p_{B-1})n] = n . \end{aligned}$$

The resulting number x will be $(p_0, p_1, \cdots, p_{B-1})$-normal in the base B.

ACKNOWLEDGMENTS. The author is indebted to Professors H. M. Hilden and T. S. Pitcher of the University of Hawaii and Professor P. C. Shields of the University of Toledo for their helpful suggestions and comments.

REFERENCES

1. E. Borel, *Lecons sur la Theories des Fonctions*, (2nd ed., 1914), 182-216.
2. D. G. Champernowne, J. London Math. Soc., **8** (1933), 254-260.
3. A. H. Copeland and P. Erdös, Bull. Amer. Math. Soc., **52** (1946), 857-860.
4. S. Ito and I. Shiokawa, J. Math. Soc. Japan, **27** (1975), 20-23.
5. D. E. Knuth, *BIT* **5** (1965), 246-250.
6. A. G. Postnikov and I. I. Pjateckii-Sapiro, Izv. Akad. Nauk SSSR Ser. Mat., **21** (1957), 501-514. (Russian).
7. A. G. Postnikov, *Selected Transl. in Math. Statist. and Probability*, **13** (1973), 41-122.
8. I. I. Pjateckii-Sapiro, Izv. Akad. Nauk SSSR Ser. Mat., **15** (1951), 47-52. (Russian).

Received May 10, 1977.

UNIVERSITY OF HAWAII AT MANOA
HONOLULU, HI 96822