

REGULAR SEQUENCES AND FREQUENCY DISTRIBUTIONS

E. B. LEACH

1. **Introduction.**¹ We consider first sequences of digits in the range $(0, \dots, m-1)$. Such a sequence determines a number between 0 and 1 in the m -ary system, from which we obtain both a topology and a measure on the class of all such sequences, using Lebesgue measure on the real line.

If f is such a sequence, and n a positive integer, we may associate with f a sequence g whose terms are themselves finite sequences of digits, or *words*. $g(i) = (f(i), f(i+1), \dots, f(i+n-1))$. Sequences g formed in this way are not arbitrary; for any i , $g(i)$ ends in the subword of $(n-1)$ digits: $(f(i+1), \dots, f(i+n-1))$, which is the initial subword of $g(i+1)$. Conversely, if g is any sequence of n -digit words, such that every pair of consecutive words from g has this property, there is a unique sequence f of digits that is associated with g as described above. We will call such a sequence g *regular*; this term will be generalized in the next section.

In this paper we consider the frequencies with which words appear in a regular sequence g . They may all appear with equal limiting frequency m^{-n} , and in fact this happens for all g arising from sequences f of digits with the exception of a set of f 's falling within a set of measure 0.² On the other hand, the frequencies of a single regular sequence g may wander near to every conceivable value. This event is rare in the sense of measure theory, but common in a topological sense: namely, the set of f 's giving rise to g 's that do not have this property form a set of the first category.³ The phrase "conceivable value" here is subject to a condition imposed by regularity of the sequence g .

The purpose of the present paper is to describe the limit distributions (which are described as sets rather than points) of an arbitrary regular sequence. The set is a compact connected subset of the dis-

Presented to the Society, April 17, 1959; received by the editors September 11, 1959.

¹ The author is indebted to Professor Bayard Rankin, whose problem of constructing sequences with certain random properties motivated the work of this paper.

² This is proved in [4, Theorem 26].

³ We will not prove this assertion here. The proof depends on continuity at most points of certain frequency distributions depending on f , and follows the usual lines of a proof involving category of sets.

tribution space involved. The description we give of the set is complete, for if we are given a set with that description, there is a regular sequence of words having that set as its limit distribution set.

The existence theory can be used to give constructions which are simple or not, according to the conditions imposed. It is, for example, rather easy to construct a periodic sequence f of digits in which every n -digit combination occurs exactly once with its first digit in a single period of f . By stringing together several periods of these sequences for different values of n we obtain a single sequence f (certainly not periodic), in which the n -digit words are uniformly distributed for every n .

We are putting the theory in a somewhat more general setting in the next few sections, not because we have other examples in mind, but because the exposition seems more natural. The special case of sequences of words arising from sequences of digits will be carried along as an example.

2. Limit distributions of sequences. Let S and T be nonempty finite sets and let θ' and θ'' be maps from S into T . A *regular sequence* f with values in S is a sequence satisfying for every integer $i \geq 1$:

$$(1) \quad \theta''(f(i)) = \theta'(f(i+1)).$$

We will assume that there is a periodic regular sequence f_0 whose image is the entire set S .

In the case of sequences of words, S is the set of n -letter words and T the set of $(n-1)$ -letter words. If $s = (a_1, a_2, \dots, a_n)$, then $\theta'(s)$ and $\theta''(s)$ are the sequences (a_1, \dots, a_{n-1}) and (a_2, \dots, a_n) respectively. It is trivially possible to write down a periodic sequence of digits in which every n -letter word occurs at least once.

For sequences with values in S or T , frequency distributions are points in the real vector spaces R^S or R^T with bases S and T . A sequence f with values in S (regular or not) determines a sequence ν_f of integer-valued points in R^S . For any positive integer i and $s \in S$, $\nu_f(i, s)$ will denote the s th coordinate of the term $\nu_f(i)$ of the sequence. We define this to be the number of positive integers j such that $j \leq i$ and $s = f(j)$. Division by i gives the sequence μ_f of *distributions* associated with f .

The *limit distribution* of a sequence f is a subset E_f of R^S defined as follows: $x \in E_f$ if and only if there is a subsequence of μ_f converging to x .

The terms in μ_f and points of E_f have non-negative coordinates, and the sum of coordinates of each point is 1. Points with this description will be called *distribution points* of R^S . These points form

a compact subset of R^S . It is convenient to use the sum of absolute values of coordinates as a norm in R^S , so that every distribution point has norm 1.

The limit distribution set of a sequence f is itself a closed and therefore a compact set. If the terms of μ_f are regarded as one-point sets, E_f is the upper closed limit of the sequence μ_f , in the terminology of Hausdorff's *Mengenlehre*.

The maps θ' and θ'' determine natural linear transformations that we will also call θ' and θ'' of R^S into R^T . If $x \in R^S$, $\theta'(x)$ is a point x' whose coordinate for any $t \in T$ is:

$$(2') \quad x'(t) = \sum_{\theta'(s)=t} x(s).$$

Similarly, $\theta''(x) = x''$, with coordinates:

$$(2'') \quad x''(t) = \sum_{\theta''(s)=t} x(s).$$

A *regular* point in R^S is a point x satisfying $\theta'(x) = \theta''(x)$.

If f is any sequence with values in S , composition with θ' and θ'' gives sequences f' and f'' with values in T . The frequencies $\nu_{f'}$ and $\nu_{f''}$ are readily verified to be the result of applying the transformations (2') and (2'') to the corresponding terms of ν_f . The distributions $\mu_{f'}$ and $\mu_{f''}$ are similarly computed from μ_f .

The linear transformations θ' and θ'' map the limit distribution E_f into the distributions $E_{f'}$ and $E_{f''}$, subsets of R^T . Each map is into simply by continuity of θ' and θ'' . It is also onto. For suppose a subsequence of $\mu_{f'}$ converges to a point $x' \in E_{f'}$. The corresponding subsequence of μ_f may not converge, but has a convergent subsequence converging to a point x of E_f , and clearly $\theta'(x) = x'$.

If f is regular, the sequences f' and f'' satisfy the identity $f''(i) = f'(i+1)$. The limit frequencies of f' and f'' are the same, and it is easy to show that if $x \in E_f$, $\theta'(x) = \theta''(x)$, i.e., x is a regular point of R^S .

In comparing $\mu_f(j)$ and $\mu_f(j+1)$, it is seen that the relative frequency of the element $f(j+1)$ increases by an amount no greater than $1/(j+1)$, while the other components decrease by amounts leaving the sum fixed. So, $|\mu_f(j+1) - \mu_f(j)| \leq 2/(j+1)$. This implies a final property of the limit distribution: E_f is connected. Otherwise, E_f would be a disjoint union of two nonempty subsets E_1, E_2 that are separated by a positive distance δ . Infinitely many terms of μ_f are within $\delta/3$ of E_1 , and infinitely many others are within $\delta/3$ of E_2 . For indices j so large that $2/(j+1) < \delta/3$, the sequence μ_f can cross from points near to E_1 to points near to E_2 , only via intermediate

points, further than $\delta/3$ from both E_1 and E_2 ; and μ_f contains infinitely many terms of this kind. A convergent subsequence of these exists that converges to a point x that is no nearer than $\delta/3$ to either E_1 or E_2 , a contradiction, because $x \in E_f$.

In summary, the limit distribution of a regular sequence f is a non-empty compact connected set of regular distribution points.

3. Periodic regular sequences. Let x be a regular distribution point whose coordinates are all rational and strictly positive. Multiplication by a suitable integer gives a regular point y with positive integer coordinates. We will show that there is a periodic regular sequence f with period p , such that $\nu_f(p) = y$. The sequence f therefore has for limit distribution the single point x .

To find f we consider finite sequences g with values in S , defined on the domain $(1, 2, \dots, N)$, satisfying the regularity condition (1) for indices $i < N$, and the inequality, for each $s \in S$:

$$(3) \quad \nu_g(N, s) \leq y(s).$$

From this, N is no greater than the norm of y , and there exists a sequence g of this kind with N as large as possible. Certainly, $N \geq 1$, since regularity does not enter until the second term. This sequence g satisfies two equations:

$$(4) \quad \theta''(g(N)) = \theta'(g(1)),$$

$$(5) \quad \nu_g(N) = y.$$

The first of these implies that g can be extended to be periodic of period N , and the second that this extension has the frequency property required of the sequence f . To prove (4), suppose the two sides are different; let $t = \theta''(g(N))$. Then t occurs once more in the sequence $\theta'' \circ g$ than it does in the sequence $\theta' \circ g$, since by regularity, t occurs in $\theta'(g(2)), \dots, \theta'(g(N))$ exactly as often as it occurs in $\theta''(g(1)), \dots, \theta''(g(N-1))$. Letting g' and g'' stand for $\theta' \circ g$ and $\theta'' \circ g$, we have a sequence of relations involving this t :

$$(6) \quad \sum_{\theta'(s)=t} \nu_g(N, s) = \nu_{g'}(N, t) < \nu_{g''}(N, t) = \sum_{\theta''(s)=t} \nu_g(N, s) \\ \leq \sum_{\theta''(s)=t} y(s) = \sum_{\theta'(s)=t} y(s).$$

The first two equalities give the frequencies of g' and g'' in terms of that of g , using equations (2') and (2''), while the last is the condition of regularity on y . From the composite strict inequality, there is an element $s \in S$ such that $\theta'(s) = t$ and $\nu_g(N, s) < y(s)$. Then we may extend g by defining $g(N+1) = s$, to be regular and satisfy condition

(3), with N replaced by $N+1$. This violates the maximal property of N , and (4) is proved.

To prove (5), suppose for some $s \in S$, $\nu_o(s) < y(s)$. Then, considering the periodic regular sequence f_0 mapping onto S , there are integers $i < j$, such that $f_0(i) = g(1)$ and $f_0(j) = s$. There is then a smallest integer k , with $i < k$, such that $\nu_o(N, f_0(k)) < y(f_0(k))$. Then $\nu_o(N, f_0(k-1)) > 0$, for either $k-1 = i$ and $f_0(k-1) = g(1)$ or else $\nu_o(N, f_0(k-1)) = y(f_0(k-1)) > 0$. In either case, $f_0(k-1)$ occurs somewhere in g . In view of (4), cyclic rotations of g have the same properties as g , and one of these, \bar{g} has $f_0(k-1)$ as its last term. Now \bar{g} may be extended by defining $\bar{g}(N+1) = f_0(k)$, with the extended sequence regular and satisfying the inequality corresponding to (3), and again the maximal property of N is violated.

In simple cases, construction of f may be made rather readily along these lines. As an example, we construct a periodic regular sequence of three digit words with digits from the binary system, each to occur once in each period. Identifying these words by digits 0 through 7 in the natural way, we start with 0 and extend, keeping track of regularity to obtain the sequence 0 1 2 4, which cannot be extended, since 4 must be followed by 0 or 1 which have already occurred. But cyclic rotation now preserves regularity, and putting 1 last we may continue to extend to obtain the sequence 2 4 0 1 3 6 5, at which point another impasse is encountered. The remaining number 7 must be preceded by 3, and one whole period of the desired sequence is 6 5 2 4 0 1 3 7. By writing the first digits of the binary numbers involved, we obtain the sequence 1 1 0 1 0 0 0 1 \dots of digits in which every three digit combination occurs once in each period.

The condition that every coordinate of the distribution x be positive is essential. In the above example, the distribution assigning to 0 and 7 the frequency $1/2$ and to all others the value 0 is regular. But in a sequence of binary digits in which 0 0 0 and 1 1 1 occurs infinitely often, every other three digit word must also occur, with the exception of 0 1 0 and 1 0 1. A sequence of digits with this distribution of three digit words can be easily obtained by alternating increasingly long strings of zeros with equally long strings of ones, so that other words occur, but with frequency 0.

4. Regular sequences with given limit distributions. Let E be a nonempty compact connected set of regular distributions in R^S . We will show that there is a regular sequence f such that $E_f = E$.

The set P of regular distributions with strictly positive coordinates is dense in the set of all regular distributions. We will assume that the corresponding assertion with "strictly positive" deleted is known.

The stronger assertion follows from the assumption (see §2) that there is a regular sequence f_0 whose limit distribution x_0 belongs to P . Then if x is any regular distribution with rational coordinates, the interior of the line segment joining x to x_0 consists of regular distribution with strictly positive coordinates, and points from P may be found along this segment, arbitrarily near to x .

We will first prove that there is a sequence h of points from P satisfying:

- (i) Every point of E is the limit of a subsequence of h .
- (ii) No point outside E is the limit of a subsequence of h .
- (iii) $\lim_{j \rightarrow \infty} (h(j) - h(j+1)) = 0$.⁴

First, for each integer $k > 0$, there is a finite set of spheres covering E , each with center in E and radius $1/k$, because of compactness of E . Since E is connected, these may be written in a finite sequence $(S_{k1}, S_{k2}, \dots, S_{kN_k})$, possibly with repetitions in such a way that each sphere overlaps the adjacent spheres in the sequence. We may also choose the first term to overlap the last term in the corresponding sequence of spheres of radius $1/(k-1)$. Next we define a sequence $\{j_k\}$ of integers by:

$$(7) \quad j_1 = 0; \quad j_{k+1} = j_k + N_k, \quad \text{if } k \geq 1.$$

A sequence S of spheres is next defined by:

$$(8) \quad S(j) = S_{k \ j-j_k}, \quad \text{if } j_k < j \leq j_{k+1}.$$

The sequence of radii of the spheres has the limit 0. Each sphere in S has center in E , and contains a point of P . Consecutive spheres in S overlap, and every point of E is covered infinitely often by the spheres of S .

We now define a sequence h of points from P by letting $h(j)$ be any point of P that is in $S(j)$. The properties of the sequence of spheres imply the desired properties of h .

For each integer $j > 0$, there is a periodic regular sequence f_j having h_j as its limit distribution, which may be chosen to start with a fixed element of s_0 of S , independent of j . A sequence f having E as its limit distribution can be given by stringing together a number of periods of each of the sequences f_j .

Specifically, we give by induction a sequence $\{i_j\}$ of integers, with $i_1 = 0$ and $i_j < i_{j+1}$, and define f by the formula:

$$(9) \quad f(i) = f_j(i - i_j), \quad \text{if } i_j < i \leq i_{j+1}.$$

⁴ A theorem similar to this has been proved by Sekanina [6]; however, an English version may be useful. The proper setting for the theorem seems to be a compact metric space, here, the space of regular frequency distributions.

Regularity is assured by requiring that $i_j - i_{j-1}$ be a multiple of the period of f_{j-1} if $j \geq 2$, so that in switching from the sequence f_{j-1} to f_j , the first term s_0 satisfies regularity with respect to its predecessor.

The choice of i_j is also to be made so that if $i_j < i \leq i_{j+1}$, $\mu_f(i)$ will lie within a distance $2/j$ of the line segment joining $h(i-1)$ to $h(j)$. This condition implies that the sequence μ_f has the same upper closed limit as h , namely the set E itself, and f satisfies the conditions we have imposed. Suppose i_1, \dots, i_{j-1} have been chosen so that the terms in μ_f satisfy the stated limitations for $i \leq i_j$, for every choice of i_j . The sequence f consists, for $i > i_{j-1}$ of terms from i_{j-1} , and can be continued to have as limit frequency, $h(j-1)$. We make i_j so large that $|\mu_f(i_j) - h(j-1)| < 1/j$. In the sequence f_j , there is an integer N , so large that if $n > N$, $\mu_{f_j}(n)$ is nearer than $2/j$ to $h(j)$. Our final restriction on i_j is that it be so large that $2N/i_j < 1/j$.

Now, if $i_j < i \leq i_{j+1}$, f has the form (9). If $i - i_j \leq N$, $\mu_f(i)$ is nearer than $2N/i_j$ to $\mu_{f_j}(i_j)$, which in turn is nearer than $1/j$ to $h(j)$. If $i - i_j > N$, $\mu_f(i)$ is an average of $\mu_{f_j}(i_j)$ and $\mu_{f_j}(i - i_j)$, and since each of these frequencies is nearer than $2/j$ to ends of the segment $(h(j-1), h(j))$, $\mu_f(i)$ is nearer than $2/j$ to an intermediate point on that segment. This completes the proof of existence of a regular sequence whose limit distribution is E .

5. Sequences of digits. For sequences of digits, a more general problem may be posed by assigning to each positive integer n a set E_n of frequency distributions of n -digit words, each of which is to be the limit distribution of the sequence of n -digit words associated with a single sequence of digits. For the case when E_n consists of a single point, such sequences have been considered by Ville [7, pp. 18-37]. Other authors have considered sequences with related distribution properties, again considering one-point distributions [1; 2; 4; 5].

In addition to the previous conditions of regularity on each set E_n , this sequence must satisfy the condition that for each $n \geq 1$, E_n be the image under the map described above (the common value of θ' and θ'' of §2) of E_{n+1} . Then if for each n , f_n is a sequence of digits whose n -digit words have the limit distribution E_n , we may piece together sections of these to give a single sequence f solving the problem. We omit details of the construction, which are similar to those of the preceding section.

We conclude by giving two other ways of constructing periodic sequences whose n -digit words occur with equal frequency, for some fixed value of n . For the sake of brevity, we offer in proof only the observation that the methods work in the examples given.

The first method is inductive, making use of a sequence f' of period m^{n-1} , such that every $(n-1)$ -digit word occurs exactly once in each period. For each digit j we define a sequence f_j by:

$$(10) \quad f_j(1) = j; \quad f_j(i+1) = f_j(i) + f'(i) \pmod{m}.$$

The sequences are periodic with period m^{n-1} , except that if $n=2$ and m is even, the period is $2m^{n-1}$. Periods of these sequences may be pieced together to form a single sequence f of period m^n , whose n -digit words occur once in every period.

For example, if f' is the sequence 11010001, of binary digits given in §3, application of (10) gives the two sequences:

$$01001111 \dots, \quad 10110000 \dots$$

of period eight. If the second of these is permuted cyclically to start with the same three digits 010 as the first, we may alternate periods of each to give:

$$0100111101011000 \dots$$

in which every four-digit word occurs with the same frequency.

The following construction described to me by Professor Rankin⁵ is simpler, although the period of the resulting sequence, $n \cdot m^n$, is greater. n sequences are written to form an $n \times \infty$ matrix. The j th sequence consists of m^{j-1} repetitions of each digit, repeated cyclically. The columns of the matrix are then joined into a single sequence with the property that every word of n digits occurs n times in a period. For $m=2$, $n=3$, the matrix is:

$$\begin{array}{l} 01010101 \dots \\ 00110011 \dots \\ 00001111 \dots \end{array}$$

and the sequence is:

$$000100010110001101011111 \dots$$

If the sequences resulting from either of these constructions are strung together appropriately, the result is a sequence that is random in a certain sense.

REFERENCES

1. D. G. Champernowne, *The construction of decimals normal in the scale of ten*, J. London Math. Soc. vol. 8 (1933) pp. 254-260.

⁵ A construction similar to this is given in [1].

2. A. H. Copeland, *Admissible numbers in the theory of probability*, Amer. J. Math. vol. 50 (1928) pp. 535–552.
3. F. Hausdorff, *Mengenlehre*, 3rd ed., Berlin, 1935.
4. B. Rankin, *Computable probability spaces*, Acta Math. vol. 103 (1960) pp. 89–122.
5. H. Reichenbach, *The theory of probability*, Berkeley, 1949.
6. M. Sekanina, *On a certain characteristic of compact connected sets in euclidean space*, Časopis Pěst. Mat. vol. 82 (1957) pp. 129–136 (Czech) (reviewed in Math. Reviews vol. 19 (1958) p. 667).
7. J. Ville, *Étude critique de la notion de collectif*, Paris, 1939.

CASE INSTITUTE OF TECHNOLOGY

EXACT SEPARATION OF RECURSIVELY ENUMERABLE SETS WITHIN THEORIES

HILARY PUTNAM AND RAYMOND M. SMULLYAN

Introduction. Ehrenfeucht and Feferman [1] have recently shown that every r.e. (recursively enumerable) set is representable in every consistent axiomatizable extension (T) of Robinson's system R (see [2, p. 53]). In this paper we extend the above result and show that any such theory (T) has the stronger property that for any two disjoint r.e. sets A, B there is a formula $F(x)$ which represents A and whose negation $F'(x)$ (i.e. $\sim F(x)$) represents B .

The proof of Ehrenfeucht and Feferman uses Myhill's result [3] that every creative set is universal. Our proof analogously uses the recent result first proved by Muchnik [4], and independently by Smullyan [5]—that every effectively inseparable pair of r.e. sets is doubly universal (cf. Theorem A, §3).

1. Terminology. We let (T) be any theory in standard formalization (cf. Tarski [2]). For any number n we let \bar{n} (written " Δ_n " in Tarski) be the numeral associated with n . For any formula $F(x)$ we let F_T be the set of all numbers n such that $F(\bar{n})$ is provable in (T) and we let F_R be the set of all n such that $F(\bar{n})$ is refutable in (T)—i.e. such that $F'(\bar{n})$ is provable in (T). (For a consistent theory (T), the sets F_T, F_R are, of course, disjoint.) Let A, B be disjoint number sets. We say that F (i.e. $F(x)$) *represents* A iff $A = F_T$. We say that F *separates* A from B within (T), or that F *separates* the pair (A, B) in (T), iff $A \subseteq F_T$ and $B \subseteq F_R$. (This means that F, F' respectively represent supersets of A, B .) And we say that F *exactly* separates the pair (A, B)

Presented to the Society, April 25, 1959 under the title *Double isomorphism of Rosser theories*; received by the editors October 5, 1959.