HAUSDORFF DIMENSION IN PROBABILITY THEORY[1]

BY
PATRICK BILLINGSLEY

1. Summary

Let $(\Omega, \mathcal{B}, \mu)$ be a probability measure space on which is defined a stochastic process $\{x_n\}$ with finite state space. In §§1–3 we define a notion of fractional dimension, in terms of μ and $\{x_n\}$, for any set $M \subset \Omega$. If Ω is the unit interval, if μ is Lebesgue measure, and if $\sum_{n=1}^{\infty} x_n(\omega)s^{-n}$ is, for each ω, the base s expansion of ω, the definition reduces to the classical one due to Hausdorff. In §§4 and 5 we obtain, under the assumption that $\{x_n\}$ is a Markov chain, the dimensions of certain sets defined in terms of the asymptotic relative frequencies of the various transitions $i \to j$. In §7 these theorems are specialized to the case in which $\{x_n\}$ is independent. In the classical case these results become extensions of theorems due to Eggleston [4, 5] and Volkmann [16, 17]. In §6 we use the preceding theorems to obtain a result on "generalized Lipschitz conditions" on certain measures, a result which reduces in the classical case to one of Kinney [11]. In §8 the dimensions obtained in the first part of the paper are shown to be related to entropy and certain allied concepts of information theory.

2. Introduction and definitions

Let $\{x_1, x_2, \cdots\}$ be a stochastic process defined on a probability measure space $(\Omega, \mathcal{B}, \mu)$. Suppose that the state space of the process is a finite set σ, the states of which for notational convenience we take to be the first s integers: $\sigma = \{1, 2, \cdots, s\}$. Thus $x_n \in \sigma$ with probability one for all n. A set of the form

$$\{\omega : x_k(\omega) = a_n, k = 1, \cdots, n\},$$

where $(a_1, a_2, \cdots, a_n)$ is a sequence of states, we call a *cylinder* or, more specifically, an *n-cylinder*. (While the sets $\{\omega : x_1(\omega) = a \text{ or } b\}$ and $\{\omega : x_n(\omega) = a\}$ are cylinders according to the usual definition, they are not according to the one given above, which will be adhered to throughout the paper.)

If M is a subset of Ω, if $\rho > 0$, and if $\mathcal{V}$ is an enumerable (possibly finite) collection of cylinders, then we say that $\mathcal{V}$ is a ρ-covering of M provided $\mu(v) < \rho$ for all $v \in \mathcal{V}$ and $M \subset V = \bigcup \{v : v \in \mathcal{V}\}$. (We will consistently denote a collection of cylinders by a script letter, with or without subscripts, and the union of the collection by the corresponding Latin letter.) If $\alpha \geqq 0$,

Received December 12, 1958.

[1] Presented to the Institute of Mathematical Statistics August 26, 1958. Most of the work on this paper was done while the author was a National Science Foundation Postdoctoral Fellow at Princeton University.

let $L_\rho(M, \alpha) = \inf \sum \mu(v)^\alpha : \mathcal{V}$, where $\sum \mu(v)^\alpha : \mathcal{V}$ denotes $\mu(v)^\alpha$ summed over all $v \in \mathcal{V}$, and the infimum extends over all ρ-coverings $\mathcal{V}$ of M. If there exist no ρ-coverings of M, then $L_\rho(M, \alpha) = \infty$. As ρ decreases to 0, $L_\rho(M, \alpha)$ obviously increases to a limit $L(M, \alpha)$. It is easy to show that $L(\cdot, \alpha)$ is monotone and subadditive, that is, that it is an outer measure.

Suppose that $L(M, \alpha) < \infty$ and that $\varepsilon > 0$. Let $\mathcal{V}_n$ be an n^{-1}-covering of M with $\sum \mu(v)^\alpha : \mathcal{V}_r < L(M, \alpha) + 1$. If $n^{-1} < \rho$, then $L_\rho(M, \alpha + \varepsilon) < (L(M, \alpha) + 1)n^{-\varepsilon}$. Letting $n \to \infty$ we see that $L_\rho(M, \alpha)$, and hence $L(M, \alpha)$, is 0. Thus $L(M, \alpha) < \infty$ implies that $L(M, \alpha + \varepsilon) = 0$ for all $\varepsilon > 0$, and it follows that there exists an α_0, $0 \leqq \alpha_0 \leqq 1$, such that $L(M, \alpha) = \infty$ if $0 \leqq \alpha < \alpha_0$ and $L(M, \alpha) = 0$ if $\alpha_0 < \alpha \leqq 1$. This α_0 we take to be the Hausdorff dimension of M: $\alpha_0 = \dim M$.

We always have $0 \leqq \dim M \leqq 1$, although for example in the extreme case in which σ consists of a single element, it can be seen that $L(M, \alpha) = \infty$ for all α, $0 \leqq \alpha < \infty$, for every set M, even the empty set. However if $\mu\{\omega : x_n(\omega) = a_n, n = 1, 2, \cdots\} = 0$ for every sequence $\{a_n\}$ of states, that is, if every ∞-cylinder has measure 0, then we have $L(M, \alpha) = 0$ for $\alpha > 1$. Thus in this case, the only one of real interest, $\dim M$ is actually the point at which $L(M, \alpha)$ changes from ∞ to 0.

Note that $\dim M$ depends not only on M but on the measure μ and the process $\{x_n\}$ as well. If M is in the Borel field generated by $\{x_n\}$, then

$$L(M, 1) \geqq \mu(M)$$

so that $\mu(M) > 0$ implies $\dim M = 1$. At the other end of the spectrum, it can be shown that every countable set has dimension 0, provided every ∞-cylinder has measure 0. The above notion of dimension measures the magnitudes of sets in such a way that two sets of probability 0 can be compared to see which is "larger". It is the main purpose of the present paper to compute the dimensions of various sets where the strong law of large numbers is violated, under the assumption that $\{x_n\}$ is a Markov chain.

3. The Lebesgue case

In this section we relate the above definition of dimension to the classical one due to Hausdorff. Suppose Ω is the unit interval, $\mathcal{B}$ the collection of Borel subsets of Ω, and μ is Lebesgue measure. For any $\omega \in \Omega$ let

$$\omega = \sum_{n=1}^{\infty} x_n(\omega) s^{-n}$$

be the nonterminating base s expansion of ω. Then $\{x_1, x_2, \cdots\}$ is a stochastic process on $(\Omega, \mathcal{B}, \mu)$; it is, in fact, the classical model for independent trials with a fair s-sided die [1]. When $(\Omega, \mathcal{B}, \mu)$ and $\{x_n\}$ are defined in this way, we will say, using Doob's phrase, that we are in the *Lebesgue case.*

The definitions of $L_\rho(M, \alpha)$, $L(M, \alpha)$, and $\dim M$ given in §2 apply of course in the Lebesgue case. (It is of no importance that here the state space is $\{0, 1, \cdots, s-1\}$ rather than $\{1, 2, \cdots, s\}$.) Now in the Lebesgue case the cylinders are exactly the half-open intervals of the form

$$(3.1) \qquad (l/s^n, (l+1)/s^n],$$

where l and n are integers. Thus $L_\rho(M, \alpha) = \inf \sum |v|^\alpha : \mathfrak{V}$, where $|v|$ denotes the length of v, and the infimum extends over all coverings of M by intervals of the form (3.1) of length less than ρ. Let $L'_\rho(M, \alpha)$ be the same infimum but without the restriction that the intervals of the coverings have the specific form (3.1). Let $L'(M, \alpha)$ and $\dim' M$ be defined in terms of $L'_\rho(M, \alpha)$ just as, in §2, $L(M, \alpha)$ and $\dim M$ were defined in terms of $L_\rho(M, \alpha)$. Then $\dim' M$ is Hausdorff's original definition [10] of the fractional dimension of a subset M of the line. (Of course Hausdorff's definition applies in any metric space.) We will prove that

$$(3.2) \qquad \dim M = \dim' M,$$

thus showing that the definition of §2 is a generalization of that of Hausdorff.[2]

Since clearly $L_\rho(M, \alpha) \geqq L'_\rho(M, \alpha)$, we have $\dim M \geqq \dim' M$. In order to establish (3.2) it suffices to prove that $\dim M \leqq \dim' M$, and to prove this it is obviously enough to show that

$$(3.3) \qquad L_\rho(M, \alpha) \leqq (s^2+1)L'_\rho(M, \alpha).$$

Finally, (3.3) will follow easily if we can show that if u is any interval, then there exist s^2+1 intervals $v_1, \cdots, v_{s^2+1}$ of the form (3.1) such that

$$u \subset \bigcup_{i=1}^{s^2+1} v_i$$

and $|v_i| < |u|$, $i = 1, 2, \cdots, s^2+1$. To prove this last statement, let u be an interval with endpoints a and b, $a < b$. Define the integer n by $1/s^{n+1} < |u| \leqq 1/s^n$, and the integer k by $k/s^{n+2} < b \leqq (k+1)/s^{n+2}$. If $(k-s+1)/s^{n+2} \leqq a$, then

$$1/s^{n+1} < b - a \leqq (k+1)/s^{n+2} - (k-s+1)/s^{n+2},$$

which is impossible. If $(k-s^2)/s^{n+2} \geqq a$, then

$$1/s^n \geqq b - a > k/s^{n+2} - (k-s^2)/s^{n+2},$$

which is also impossible. Hence

$$\frac{k-s^2}{s^{n+2}} < a < \frac{k-s+1}{s^{n+2}} < \frac{k}{s^{n+2}} < b \leqq \frac{k+1}{s^{n+2}}.$$

---

[2] A referee points out that (3.2) has been proved for the case $s = 2$ by A. S. Besicovitch, *On existence of subsets of finite measure of sets of infinite measure*, Indag. Math., vol. 14 (1952), pp. 339–344.

From this it follows in the first place that the $s^2 + 1$ intervals

$$\left(\frac{k - s^2}{s^{n+2}}, \frac{k - s^2 + 1}{s^{n+2}}\right], \cdots, \left(\frac{k}{s^{n+2}}, \frac{k+1}{s^{n+2}}\right]$$

cover u, and in the second place that the common length $1/s^{n+2}$ of these intervals is less than $|u| = b - a$. Since these intervals are of the form (3.1), we have proved (3.3) and hence (3.2).

4. Preliminary lemmas

We now prove some preliminary results which we will need. Each theorem is a generalization of one which is well known in the Lebesgue case.

THEOREM 4.1. (i) *If* $M \subset M'$, *then* $\dim M \leqq \dim M'$. (ii) *If* $M = \bigcup_\Gamma M_\gamma$, *where* Γ *is any index set, then* $\dim M \geqq \sup_\Gamma \dim M_\gamma$. (iii) *If the index set* Γ *in* (ii) *is enumerable, then* $\dim M = \sup_\Gamma \dim M$.

Proof. Since (ii) follows immediately from (i), which is obvious, we need only prove $\dim M \leqq \sup_\Gamma \dim M_\gamma$ under the assumption that Γ is countable. But if $\alpha > \dim M_\gamma$ for each $\gamma \epsilon \Gamma$, then $L(M_\gamma, \alpha) = 0$, and since $L(\cdot, \alpha)$ is subadditive, $L(M, \alpha) = 0$, and the result follows. Part (iii) is in general not true if Γ is uncountable, and a basic problem is to prove it in special cases.

The next result generalizes Theorem 1 of [5].

THEOREM 4.2. *Suppose that for each* n, $\mathcal{V}_n$ *is an enumerable collection of cylinders and that* $M \subset \limsup_n V_n$. *If* $\sum_{n=1}^{\infty} \sum \mu(v)^\alpha : \mathcal{V}_n < \infty$, *then* $\dim M \leqq \alpha$.

Proof. Given $\rho > 0$, choose n_0 so large that

$$\textstyle\sum_{n \geqq n_0} \sum \mu(v) : \mathcal{V}_n \leqq \sum_{n \geqq n_0} \sum \mu(v)^\alpha : \mathcal{V}_n < \rho.$$

Then $\bigcup_{n \geqq n_0} \mathcal{V}_n$ is a ρ-covering of M, and hence $L_\rho(M, \alpha) \leqq \rho$. Thus $L(M, \alpha) = 0$, and the theorem follows.

If the stochastic process $\{x_n\}$ is altered on a set of probability zero, then for most purposes of probability theory the process is not essentially changed. The purpose of the present theory however is to analyze sets of probability zero, and hence we need some sort of assumption of a nonprobabilistic nature which links the process $\{x_n\}$ with the space Ω. The following condition seems to be the weakest one which leads to simple results.

CONDITION (C). *All but a countable number of sequences* $(a_1, a_2, \cdots)$ *of states have the property that either the set* $\{\omega : x_k(\omega) = a_k, k = 1, 2, \cdots\}$ *is nonempty, or else* $\mu\{\omega : x_k(\omega) = a_k, k = 1, 2, \cdots, n\} = 0$ *for some* n.

Condition (C), which is essentially a compactness condition, will be a hypothesis in most of the theorems which follow. Note that this condition is satisfied in the Lebesgue case; in fact since $\sum_{n=1}^{\infty} x_n(\omega) s^{-n}$ is in this case the nonterminating expansion of ω, the sequences $\{a_n\}$ which violate the requirements of the condition are exactly those with $a_n = 0$ for all sufficiently large n. A probability measure space to carry a stochastic process with given properties is ordinarily constructed by taking the countable combinatorial product of the state space with itself. In this case Condition (C) is always satisfied.

Theorem 4.2 enables one to find upper bounds for the dimension of a set. The following result, which is a generalization of Theorem 5 of Eggleston [5], enables one to find lower bounds. The two theorems used together enable one to find the exact dimensions of certain interesting sets (see §5).

THEOREM 4.3. *Suppose that for each integer n we have a finite set $\mathcal{V}_n$ of cylinders which are pairwise disjoint. Suppose that each element of $\mathcal{V}_{n+1}$ is contained in some element of $\mathcal{V}_n$, and that each element of $\mathcal{V}_n$ contains exactly ν_{n+1} elements of $\mathcal{V}_{n+1}$. Suppose further that if $v \in \mathcal{V}_n$, $v' \in \mathcal{V}_{n+1}$, and $v \supset v'$, then $\mu(v')/\mu(v) = \rho_{n+1}$, where $\rho_{n+1} > 0$ is independent of v and v'. Suppose that for all $\alpha < \alpha_0$ there exists an n_α such that*

$$\sum_{n \geqq n_\alpha} \frac{\rho_{n_\alpha} \cdots \rho_{n-1}}{(\rho_{n_\alpha} \cdots \rho_n)^{1+\alpha}(\nu_{n_\alpha} \cdots \nu_n)} < \infty. \tag{4.1}$$

If $V = \bigcap_{n=1}^{\infty} V_n$, then $\dim V \geqq \alpha_0$, *provided Condition* (C) *is satisfied.*

Proof. We may assume that $\mathcal{V}_1$ consists of exactly one cylinder. For otherwise we may remove from $\mathcal{V}_1$ all cylinders but one and remove from $\mathcal{V}_n$ any cylinder not contained in the remaining element of $\mathcal{V}_1$. Since this has the effect of decreasing V, the original V has dimension not less than α_0 if the modified one does.

Assume then that $\mathcal{V}_1$ consists of a single cylinder v_1, and define $\delta_n = \mu(v_1)\rho_2 \cdots \rho_n$ and $N_n = \nu_2 \cdots \nu_n$ for $n \geqq 2$. Then $\mathfrak{N}(\mathcal{V}_n) = N_n$, where $\mathfrak{N}$ denotes the number of elements in a set, and $\mu(v) = \delta_n$ for all $v \in \mathcal{V}_n$. It follows immediately from (4.1) that

$$\textstyle\sum_{n=3}^{\infty} \delta_{n-1}/\delta_n^{1+\alpha} N_n < \infty \tag{4.2}$$

for any $\alpha < \alpha_0$.

Given $\alpha < \alpha_0$, choose n_0 so that

$$\textstyle\sum_{n \geqq n_0}^{\infty} \delta_{n-1}/\delta_n^{1+\alpha} N_n < \frac{1}{2}, \tag{4.3}$$

which is possible by (4.2). Let $\mathfrak{U}$ be a finite or enumerable collection of cylinders such that

$$\textstyle\sum \mu(u)^{\alpha} : \mathfrak{U} < 1 \tag{4.4}$$

and

$$\mu(u) < \delta_{n_0} \quad \text{for} \quad u \in \mathfrak{U}. \tag{4.5}$$

We will show that $\mathcal{U}$ does not cover V. From this it will follow that

$$L(V, \alpha) \geqq L_{\delta_{n_0}}(V, \alpha) \geqq 1,$$

and hence that $\dim V \geqq \alpha$. Since α can be taken arbitrarily close to α_0, the theorem will follow.

Suppose then that $\mathcal{U}$ satisfies (4.4) and (4.5), where n_0 satisfies (4.3). Those sequences of states which violate the requirements of Condition (C) we will call the *exceptional* sequences. Let $a(j) = (a_1(j), a_2(j), \cdots)$, $j = 1, 2, \cdots$, be an enumeration of the exceptional sequences. For each $j = 1, 2, \cdots$, let $e_j = \{\omega : x_k(\omega) = a_k(j), k = 1, \cdots, l_k\}$ be a cylinder such that

$$(4.6) \qquad \mu(e_j) < [(1 - \textstyle\sum \mu(u)^\alpha : \mathcal{U})/2^j]^{1/\alpha}, \qquad \mu(e_j) < \delta_{n_0}.$$

Such a cylinder exists because $\{\omega : x_k(\omega) = a_k(j), k = 1, 2, \cdots\}$ is empty. Now let $\mathcal{W}$ be the collection $\mathcal{W} = \mathcal{U} \cup \{e_j : j = 1, 2, \cdots\}$. If $\mathcal{W}$ does not cover V, then certainly $\mathcal{U}$ does not. By (4.6) we have

$$(4.7) \qquad \textstyle\sum \mu(w)^\alpha : \mathcal{W} < 1,$$

and

$$(4.8) \qquad \mu(w) < \delta_{n_0} \quad \text{for} \quad \omega \in \mathcal{W},$$

that is, $\mathcal{W}$ satisfies the same conditions $\mathcal{U}$ does.

For $i = 1, 2, \cdots$, let $\mathcal{W}_i$ be the set of elements of $\mathcal{W}$ for which

$$(4.9) \qquad \delta_{n_0+i-1} \geqq \mu(w) > \delta_{n_0+i}.$$

By (4.7) each $\mathcal{W}_i$ contains finitely many elements, and by (4.8)

$$\mathcal{W} = \textstyle\bigcup_{i=1}^{\infty} \mathcal{W}_i.$$

An element of $\mathcal{V}_{n_0+i}$ which meets an element of $\mathcal{W}_i$ is contained in it, having a smaller probability. Now by (4.7) and (4.9)

$$(4.10) \qquad 1 > \textstyle\sum \mu(w)^\alpha : \mathcal{W}_i \geqq \mathfrak{N}(\mathcal{W}_i)\delta_{n_0+i}^\alpha.$$

Let $\mathcal{B}_i$ be the set of elements of $\mathcal{V}_{n_0+i}$ which meet elements of $\mathcal{W}_i$, i.e., which are contained in W_i. Then $B_i \subset W_i$, and by (4.9) and (4.10)

$$(4.11) \qquad \mu(B_i) \leqq \mu(W_i) \leqq \mathfrak{N}(\mathcal{W}_i)\delta_{n_0+i-1} < \delta_{n_0+i-1}/\delta_{n_0+i}^\alpha.$$

If $1 \leqq i \leqq t$, then each element of $\mathcal{V}_{n_0+t}$ is contained in exactly one element of $\mathcal{V}_{n_0+i}$, and each element of $\mathcal{V}_{n_0+i}$ contains exactly N_{n_0+t}/N_{n_0+i} elements of $\mathcal{V}_{n_0+t}$. Moreover if $v \in \mathcal{V}_{n_0+i}$ and $v' \in \mathcal{V}_{n_0+t}$, then $\mu(v')/\mu(v) = \delta_{n_0+t}/\delta_{n_0+i}$. It follows that if $\mathcal{C}_{t,i}$ is the set of elements of $\mathcal{V}_{n_0+t}$ which are covered by elements of $\mathcal{B}_i$, then

$$\mu(C_{t,i})/\mu(B_i) = N_{n_0+t}\,\delta_{n_0+t}/N_{n_0+i}\,\delta_{n_0+i}.$$

By (4.11) we have

$$\mu(C_{t,i}) < N_{n_0+t}\, \delta_{n_0+t}\, \delta_{n_0+i-1}/\delta_{n_0+1}^{1+\alpha}\, N_{n_0+i}\,. \tag{4.12}$$

If $\mathcal{C}_t = \bigcup_{i=1}^{t} \mathcal{C}_{t,i}$, then by (4.12) and (4.3)

$$\mu(C_t) < N_{n_0+t}\, \delta_{n_0+t}/2 = \mu(V_{n_0+t})/2.$$

Therefore $\mathcal{C}_t$ is a *proper* subset of $\mathcal{V}_{n_0+t}$.

It is easily seen that if an element v of $\mathcal{V}_{n_0+t}$ meets an element of $\bigcup_{i=1}^{t} \mathcal{W}_i$, then $v \in \mathcal{C}_t$. Thus the set $\mathcal{V}_t'$ of elements of $\mathcal{V}_{n_0+t}$ not met by any element of $\bigcup_{i=1}^{t} \mathcal{W}_i$ is nonempty. Clearly each element of $\mathcal{V}_{t+1}'$ is contained in some element of $\mathcal{V}_t'$. Let $\mathcal{V}_t''$ be the set of elements of $\mathcal{V}_t'$ containing infinitely many elements of $\bigcup_{j>t} \mathcal{V}_j'$. Since $\mathcal{V}_t'$ is finite, $\mathcal{V}_t''$ is nonempty. Since each element of $\mathcal{V}_t''$ clearly contains some element of $\mathcal{V}_{t+1}''$, it is possible to construct inductively a decreasing sequence $\{v_t''\}$ of cylinders with $v_t'' \in \mathcal{V}_t''$. Let

$$v'' = \bigcap_{t=1}^{\infty} v_t'' = \{\omega : x_k(\omega) = a_k,\ k = 1, 2, \cdots\}.$$

Suppose that $(a_1, a_2, \cdots)$ is one of the exceptional sequences, say $a(j)$. Then there exists some t_0 such that $v_t'' \subset e_j$ for all $t \geqq t_0$. Since $e_j \in \mathcal{W}$, $e_j \in \mathcal{W}_i$ for some i. If t exceeds both t_0 and i, then $v_t'' \subset e_j \in \mathcal{W}_1 \cup \cdots \cup \mathcal{W}_t$, which is impossible since $v_t'' \in \mathcal{V}_t'' \subset \mathcal{V}_t'$. Thus $(a_1, a_2, \cdots)$ is not one of the exceptional sequences. On the other hand, $\mu(v_t'') = \delta_{n_0+t} > 0$ for all t, which clearly implies that $\mu\{\omega : x_k(\omega) = a_k,\ k = 1, \cdots, l\} > 0$ for all l. Since $(a_1, a_2, \cdots)$ is not exceptional, it follows that v'' is nonempty.

By construction, $v'' \subset V_{n_0+t}$ for each t so that $v'' \subset V$. On the other hand, v'' is not contained in any element of $\bigcup_{i=1}^{t} \mathcal{W}_i$ for any t. Thus $\mathcal{W}$ does not cover V, and the proof of the theorem is complete.

5. Main theorems

In this section we compute the Hausdorff dimensions of various sets under the assumption that $\{x_n\}$ is a Markov chain with stationary transition probabilities.

For $i, j = 1, \cdots, s$, $n = 1, 2, \cdots$, and $\omega \in \Omega$, let $\delta_{ij}(\omega, n)$ be $1/n$ times the number of k, $1 \leqq k \leqq n$, such that $x_k(\omega) = i$ and $x_{k+1}(\omega) = j$. Thus $\delta_{ij}(\omega, n)$ is the relative frequency of the transition $i \to j$ in the first n steps of the process. Let

$$\begin{aligned} \delta_i(\omega, n) &= \delta_{i\cdot}(\omega, n) = \textstyle\sum_{j=1}^{s} \delta_{ij}(\omega, n), \\ \delta_{\cdot i}(\omega, n) &= \textstyle\sum_{j=1}^{s} \delta_{ji}(\omega, n). \end{aligned} \tag{5.1}$$

Then $\delta_{i\cdot}(\omega, n)$ and $\delta_{\cdot i}(\omega, n)$ are the relative frequencies of i among $\{x_1(\omega), \cdots, x_n(\omega)\}$ and $\{x_2(\omega), \cdots, x_{n+1}(\omega)\}$, respectively.

Assume now that $\{x_n\}$ is a Markov chain with transition matrix $P = (p_{ij})$. See [6] or [3] for the properties of Markov chains needed here. Let σ' be the set of ordered pairs (i, j) with $i, j \in \sigma$ and $p_{ij} > 0$. Consider points $\xi = (\xi_{ij})$ of $\mathfrak{N}(\sigma')$-dimensional Euclidean space with coordinates indexed by the elements (i, j) of σ' (taken in some arbitrary but fixed order). Let A be the set of ξ such that $\xi_{ij} \geqq 0$ for all $(i, j) \in \sigma'$ and $\sum_{\sigma'} \xi_{ij} = 1$. We denote by ρ the ordinary Euclidean metric on A. For $\xi \in A$ let $r_i(\xi) = \sum_j \xi_{ij}$ and $c_i(\xi) = \sum_j \xi_{ji}$, where the summation extends in the first case over those j such that $(i, j) \in \sigma'$, and in the second over those j such that $(j, i) \in \sigma'$. For $\xi \in A$ let

$$(5.2) \qquad H(\xi:P) = \frac{\sum_{\sigma'} \xi_{ij} \lg \xi_{ij}/r_i(\xi)}{\sum_{\sigma'} \xi_{ij} \lg p_{ij'}} = \frac{\sum_{\sigma'} \xi_{ij} \lg \xi_{ij}/c_j(\xi)}{\sum_{\sigma'} \xi_{ij} \lg p_{ij}},$$

where $0 \lg 0$ is taken to be 0. If $c_i(\xi) = 0$, then $\xi_{ij} = 0$ for all j, and we take $\xi_{ij} \lg \xi_{ij}/c_i(\xi) = 0$. The function $H(\xi:P)$, in terms of which our Hausdorff dimensions will be expressed, is essentially an entropy and is discussed in §8 below.

Let B be the set of points ξ of A such that $r_i(\xi) = c_i(\xi)$ for all i. Then B is a closed subset of A. If S is any subset of A and $\{\xi(n)\}$ is a sequence of points of A, we say that $\xi(n)$ approaches S as n approaches ∞, and write $\xi(n) \to S$, if $\lim_n \rho(\xi(n), S) = 0$.

For $n = 1, 2, \cdots$, and $\omega \in \Omega$, let $\delta(\omega, n)$ be the point of A with coordinates $\delta_{ij}(\omega, n)$ for all $(i, j) \in \sigma'$. Then $r_i(\delta(\omega, n)) = \delta_{i\cdot}(\omega, n)$ and $c_i(\delta(\omega, n)) = \delta_{\cdot i}(\omega, n)$. Since obviously $| \delta_{i\cdot}(\omega, n) - \delta_{\cdot i}(\omega, n) | \leqq 1/n$, $\delta(\omega, n) \to B$.

If $S \subset B$, let

$$(5.3) \qquad \begin{aligned} M(S) &= \{\omega : \delta(\omega, n) \to S\}, \\ N(S) &= \{\omega : \rho(\delta(\omega, n), S) = O(n^{-1})\}. \end{aligned}$$

If S consists of a single point ξ, we write $M(\xi)$ and $N(\xi)$. Our main result gives the dimensions of $M(S)$ and $N(S)$ under the assumption that $\{x_n\}$ is a regular Markov chain, that is, under the assumption that there are no transient states and just one ergodic class, with or without cyclically moving subsets.

THEOREM 5.1. *Suppose that $\{x_n\}$ is a regular Markov chain with transition matrix $P = (p_{ij})$. If $S \subset B$, then*

$$(5.4) \qquad \begin{aligned} \dim M(S) = \dim N(S) &= \dim \textstyle\bigcup_S M(\xi) \\ &= \dim \textstyle\bigcup_S N(\xi) = \sup_S H(\xi:P), \end{aligned}$$

provided Condition (C) *is satisfied.*

Note that we have made no assumption about the initial distribution. We prove Theorem 5.1 by a series of lemmas.

Let $(a_1, a_2, \cdots, a_{n+1})$ be a sequence of states. For $i, j = 1, \cdots, s$ let f_{ij} be the number of k, $1 \leqq k \leqq n$, such that $a_k = i$ and $a_{k+1} = j$. Let $f_{i\cdot} = \sum_{j=1}^{s} f_{ij}$ and $f_{\cdot i} = \sum_{j=1}^{s} f_{ji}$. The $s \times s$ matrix $F = (f_{ij})$ we will call the *transition count* of the sequence. Clearly $f_{i\cdot} - f_{\cdot i} = \delta_{i,a_1} - \delta_{i,a_{n+1}}$,

$i = 1, \cdots, s$. If F is any $s \times s$ matrix with nonnegative integral entries, we will call it *admissible* provided $\sum_{ij} f_{ij} = n$, $f_{i\cdot}, f_{\cdot i} > 0$, $i = 1, \cdots, s$, and $f_{i\cdot} - f_{\cdot i} = \delta_{iu} - \delta_{iv}$, $i = 1, \cdots, s$ for some u and v. If F is admissible, let $N_u^{(n)}(F)$ be the number of sequences $(u, a_1, a_2, \cdots, a_n)$ which have F as transition count. Lemma 5.2 below, which is due to Whittle [18], gives a simple expression for $N_u^{(n)}(F)$. This result has been proved by Whittle by integration methods and by Dawson and Good [2] and Goodman [9] by graph-theoretic methods. The following simple inductive proof may be of independent interest.

If $G = (g_{ij})$ is an $s \times s$ matrix, we denote by G_{ij} the $(i, j)^{\text{th}}$ cofactor of G. If this cofactor is independent of i, we denote it by $G_{\cdot j}$.

LEMMA 5.1. *If* $\sum_i g_{ij} = 0, j = 1, \cdots, s$, *then* $G_{ij} = G_{\cdot j}$ *does not depend on* i, *and*

$$G_{\cdot i} = \sum_j (\delta_{ij} - g_{ij}) G_{\cdot j}. \tag{5.5}$$

Proof. If the rank r of G is less than $s - 1$, then G_{ij} is identically 0, and the result is trivial. Since clearly $r < s$, we may assume $r = s - 1$. But then $\sum_l G_{li} g_{lj} = \delta_{ij} \det G = 0$, so that $(G_{1i}, \cdots, G_{si})$ must be a scalar multiple of $(1, \cdots, 1)$, since $\sum_l g_{lj} = 0$. Thus $G_{ij} = G_{\cdot j}$.[3] Clearly $\sum_j (\delta_{ij} - g_{ij}) G_{ij} = G_{\cdot i} - \det G = G_{\cdot i}$.

LEMMA 5.2. *If F is admissible and $n > s$, then*

$$N_u^{(n)}(F) = \left(\prod_j f_{\cdot j}! / \prod_{ij} f_{ij}!\right) \hat{F}_{\cdot u}, \tag{5.6}$$

where $\hat{F}$ is the matrix with entries $\delta_{ij} - f_{ij}/f_{\cdot j}$.

Proof. Since F is admissible, $f_{\cdot j} > 0$ and $f_{ij}/f_{\cdot j}$ is well defined. Since $\sum_i (\delta_{ij} - f_{ij}/f_{\cdot j}) = 0$, $\hat{F}_{ij} = \hat{F}_{\cdot j}$ is independent of i by Lemma 5.1.

The result being easy to establish for $n = s + 1$, assume (5.6) holds for $n - 1$. If $F(i, j)$ is F with the $(i, j)^{\text{th}}$ element diminished by 1, then clearly

$$N_u^{(n)}(F) = \sum_v N^{(n-1)}(F(u, v)), \tag{5.7}$$

where the summation extends over those v for which $f_{uv} > 0$. It suffices to verify that the right-hand side of (5.6) satisfies (5.7) or, since $\hat{F}_{uv}(u, v) = \hat{F}_{uv}$, that

$$F_{\cdot u} = \sum_{v=1}^{s} f_{uv} f_{\cdot u}^{-1} \hat{F}_{\cdot v},$$

which follows immediately from (5.5), taking $G = I - \hat{F}$.

LEMMA 5.3. *Suppose that $\{x_n\}$ is a Markov chain with transition matrix $P = (p_{ij})$. If $S \subset B$, then*

$$\dim M(S) \leqq \sup_{\xi \in S} H(\xi : P). \tag{5.8}$$

Proof. Note that we have not assumed Condition (C), and we have made

[3] This argument is due to O. S. Rothaus.

no assumptions on the chain $\{x_n\}$. Since $\delta(\omega, n)$ approaches S if and only if it approaches its closure, and $H(\cdot:P)$ is continuous on B, we may assume that S is closed. Then S is compact, and for any $\varepsilon > 0$ there exists a finite number L_ε of points $\xi(\varepsilon, l)$, $l = 1, \cdots, L_\varepsilon$ such that $\xi(\varepsilon, l) \in S$ and

$$S \subset \bigcup_{l=1}^{L_\varepsilon} \{\xi \in A: |\xi_{ij} - \xi_{ij}(\varepsilon, l)| < \varepsilon, (i, j) \in \sigma'\}. \tag{5.9}$$

Let

$$V_n(\varepsilon, l) = \{\omega: |\delta_{ij}(\omega, n) - \xi_{ij}(\varepsilon, l)| < \varepsilon, (i, j) \in \sigma'\},$$
$$V_n(\varepsilon) = \bigcup_{l=1}^{L_\varepsilon} V_n(\varepsilon, l).$$

Denote the right-hand side of (5.9) by E for the moment. Then E is open in A, $E \supset S$, and $V_n(\varepsilon) = \{\omega: \delta(\omega, n) \in E\}$. If $\omega \in M$, then $\omega \in V_n(\varepsilon)$ for all but a finite number of n. Therefore

$$M(S) \subset \liminf_{n\to\infty} V_n(\varepsilon), \tag{5.10}$$

for any $\varepsilon > 0$.

Clearly $V_n(\varepsilon, l)$ and $V_n(\varepsilon)$ are unions of collections $\mathcal{V}_n(\varepsilon, l)$ and $\mathcal{V}_n(\varepsilon) = \bigcup_{l=1}^{L_\varepsilon} \mathcal{V}_n(\varepsilon, l)$ of n-cylinders. We will show that for any given

$$\alpha > \sup_S H(\xi:P)$$

there is an $\varepsilon > 0$ such that

$$\sum_{n=1}^{\infty} \sum \mu(v)^\alpha : \mathcal{V}_n(\varepsilon) < \infty. \tag{5.11}$$

It will follow from (5.10) and Lemma 4.2 that $\dim M(S) \leqq \sup_S H(\xi:P)$.

Now clearly

$$\sum \mu(v)^\alpha : \mathcal{V}_n(\varepsilon, l) = \sum_F \sum_u N_u^{(n)}(F)\,(p_u \textstyle\prod_{ij} p_{ij}^{f_{ij}})^\alpha,$$

where the first summation extends over transition counts F satisfying

$$|f_{ij}/n - \xi_{ij}(\varepsilon, l)| < \varepsilon, \qquad (i, j) \in \sigma',$$

and $f_{ij} = 0$, $(i, j) \notin \sigma'$, and, for fixed F, the second summation extends over those u such that $f_{i\cdot} - f_{\cdot i} = \delta_{iu} - \delta_{iv}$ for some v. Let K_1 be an upper bound of the values of all determinants of order less than s having entries in $[0, 1]$. Then $K_1 < \infty$, and it follows from Lemma 5.2 that

$$\sum \mu(v)^\alpha : \mathcal{V}_n(\varepsilon, l) \leqq sK_1 \sum_F \left(\textstyle\prod f_{\cdot j}!/\prod f_{ij}!\right)\left(\prod p_{ij}^{f_{ij}}\right)^\alpha, \tag{5.12}$$

where the F-summation is the same as before. The inequality (5.12) will remain valid if we increase the range of the F-summation to include all $s \times s$ matrices of nonnegative integers satisfying

$$\sum_{i,j=1}^{s} f_{ij}/n = 1,$$
$$|f_{ij}/n - \xi_{ij}(\varepsilon, l)| < \varepsilon, \qquad (i, j) \in \sigma' \tag{5.13}$$
$$f_{ij} = 0, \qquad (i, j) \notin \sigma'.$$

In other words we drop the requirement that F satisfy $f_{i\cdot} - f_{\cdot i} = \delta_{iu} - \delta_{iv}$, $i = 1, \cdots, s$, for some u and v.

We consider only those ε such that

$$\varepsilon < 1/2s^4. \tag{5.14}$$

If i_0 and j_0 are appropriate functions of ε and l, with $(i_0, j_0) \in \sigma'$, then

$$\xi_{i_0 j_0}(\varepsilon, l) = \max_{(i,j)\in\sigma'} \xi_{ij}(\varepsilon, l) \geqq 1/s^2 \geqq \varepsilon s^2. \tag{5.15}$$

Now let $T(\varepsilon, l, n)$ be the $s \times s$ matrix with entries

$$\begin{aligned} t_{ij}(\varepsilon, l, n) &= [n(\xi_{ij}(\varepsilon, l) + \varepsilon)], && (i, j) \in \sigma', (i, j) \neq (i_0, j_0) \\ t_{ij}(\varepsilon, l, n) &= 0, && (i, j) \notin \sigma', \\ t_{i_0 j_0}(\varepsilon, l, n) &= n - \textstyle\sum_{(ij)\neq(i_0 j_0)} t_{ij}, \end{aligned} \tag{5.16}$$

where $[x]$ denotes the integral part of x.

For any matrix F with nonnegative integral entries satisfying $\sum_{ij} f_{ij} = n$ and $f_{ij} = 0$ for $(i, j) \notin \sigma'$, let $Q(F) = (\prod f_{\cdot j}!/(\prod f_{ij}!)(\prod p_{ij}^{f_{ij}})^{\alpha}$. Let $(i, j) \in \sigma'$ be distinct from (i_0, j_0), and let F' and F'' be two matrices of the above form satisfying

$$\begin{aligned} f''_{i_0 j_0} &= f'_{i_0 j_0} - 1 \\ f''_{ij} &= f'_{ij} + 1 \\ f''_{uv} &= f'_{uv} && \text{if } (u, v) \neq (i, j) \text{ and } (u, v) \neq (i_0, j_0). \end{aligned} \tag{5.17}$$

In other words, suppose F'' is obtained from F' by transferring one count from (i_0, j_0) to (i, j). Let $K_2 = \max_{(u,v)\in\sigma'} p_{uv}^{-1} < \infty$. If $j = j_0$, then

$$Q(F')/Q(F'') \leqq K_2(f'_{ij} + 1)/f'_{i_0 j_0}. \tag{5.18}$$

If $j \neq j_0$, then

$$Q(F')/Q(F'') \leqq K_2 f'_{\cdot j_0}/f'_{i_0 j_0}. \tag{5.19}$$

Suppose that

$$| f'_{ij}/n - \xi_{ij}(\varepsilon, l) | \leqq \varepsilon s^2, \qquad (i, j) \in \sigma'. \tag{5.20}$$

If (5.20) and (5.18) hold, then

$$Q(F')/Q(F'') \leqq 2K_2 f'_{ij}/f'_{i_0 j_0} \leqq 4K_2(1 + s^2)/s^2 = K_3$$

by (5.14) and (5.15). If (5.20) and (5.19) hold, then

$$Q(F')/Q(F'') \leqq 2K_2/s^2 \leqq K_3.$$

Therefore $Q(F')/Q(F'') \leqq K_3$ if (5.17) and (5.20) hold. If n exceeds some $n_1(\varepsilon)$, then (5.20) holds for $F' = T(\varepsilon, l, n)$, where $T(\varepsilon, l, n)$ is defined by (5.16). If F satisfies (5.13), then it can be carried into $T(\varepsilon, l, n)$ by at most

$2n\varepsilon$ transformations of the form $F' \to F''$. Moreover, every F' encountered in the process will satisfy (5.20). It follows that

$$Q(F) \leqq K_3^{2n\varepsilon} Q(T(\varepsilon, l, n)).$$

The number of matrices F satisfying (5.13) is at most $(2n\varepsilon)^{s^2}$. Thus by (5.12) and the definition of $\mathcal{V}_n(\varepsilon)$, we have, provided $n \geqq n_1(\varepsilon)$,

$$(5.21) \qquad \sum \mu(v)^\alpha : \mathcal{V}_n(\varepsilon) \leqq \sum_{l=1}^{L_\varepsilon} sK_1(2n\varepsilon)^{s^2} K_3^{2n\varepsilon} Q(T(\varepsilon, l, n)),$$

where K_1 and K_3 depend only on the p_{ij}.

Now let

$$\xi_{ij}^*(\varepsilon, l) = \xi_{ij}(\varepsilon, l) + \varepsilon, \qquad (i, j) \in \sigma', \quad (i, j) \neq (i_0, j_0),$$
$$\xi_{i_0 j_0}^*(\varepsilon, l) = 1 - \sum_{(ij) \neq (i_0 j_0)} \xi_{ij}(\varepsilon, l),$$

and let $\xi^*(\varepsilon, l)$ be the vector with components $\xi_{ij}^*(\varepsilon, l)$. It follows from (5.14) and (5.15) that $\xi^*(\varepsilon, l)$ lies in A and, in fact, has positive coordinates. Moreover,

$$(5.22) \qquad | \xi_{ij}^*(\varepsilon, l) - \xi_{ij}(\varepsilon, l) | < s^2 \varepsilon, \qquad (i, j) \in \sigma'.$$

By the definition (5.16)

$$(5.23) \qquad \lim_{n \to \infty} t_{ij}(\varepsilon, l, n)/n = \xi_{ij}^*(\varepsilon, l), \qquad (i, j) \in \sigma'.$$

Now fix α, ε and l. Since $\xi_{ij}^*(\varepsilon, l) > 0$, $t_{ij}(\varepsilon, l, n) \to \infty$. Applying Stirling's formula to $Q(T(\varepsilon, l, n))$ we have, as $n \to \infty$,

$$n^{-1} \lg Q(T(\varepsilon, l, n)) = \alpha \sum_{\sigma'} n^{-1} t_{ij}(\varepsilon, l, n) \lg p_{ij}$$
$$- \sum_{\sigma'} n^{-1} t_{ij}(\varepsilon, l, n) \lg t_{ij}(\varepsilon, l, n)/t_{\cdot j}(\varepsilon, l, n) + o(1).$$

From (5.23) it now follows that if $\eta > 0$ then there exists an $n_2(\alpha, \varepsilon, \eta)$ such that

$$(5.24) \qquad \begin{aligned} n^{-1} \lg Q(T(\varepsilon, l, n)) &\leqq \alpha \sum_{\sigma'} \xi_{ij}^*(\varepsilon, l) \lg p_{ij} \\ &- \sum_{\sigma'} \xi_{ij}^*(\varepsilon, l) \lg \xi^*(\varepsilon, l)/c_j(\xi^*(\varepsilon, l)) + \eta, \quad l = 1, 2, \cdots, L_\varepsilon, \end{aligned}$$

provided $n \geqq n_2(\alpha, \varepsilon, \eta)$. The functions

$$\sum_{\sigma'} \xi_{ij} \lg p_{ij} \quad \text{and} \quad \sum_{\sigma'} \xi_{ij} \lg \xi_{ij}/c_j(\xi)$$

being uniformly continuous on A, it follows from (5.22) that there is for each $\eta > 0$ an $\varepsilon_0(\eta)$ such that if $\varepsilon < \varepsilon_0(\eta)$, then

$$\sum_{\sigma'} \xi_{ij}^*(\varepsilon, l) \lg p_{ij} \leqq \sum_{\sigma'} \xi_{ij}(\varepsilon, l) \lg p_{ij} + \eta,$$

$$(5.25) \qquad \begin{aligned} &\sum_{\sigma'} \xi_{ij}^*(\varepsilon, l) \lg \xi_{ij}^*(\varepsilon, l)/c_j(\xi^*(\varepsilon, l)) \\ &\qquad \geqq \sum_{\sigma'} \xi_{ij}(\varepsilon, l) \lg \xi_{ij}(\varepsilon, l)/c_j(\xi(\varepsilon, l)) - \eta. \end{aligned}$$

If $\varepsilon < \varepsilon_0(\eta)$ and $n \geqq n_2(\alpha, \varepsilon, \eta)$, then by (5.24), (5.25), and (5.2),

$$(5.26) \qquad \begin{aligned} &n^{-1} \lg Q(T(\varepsilon, l, n)) \\ &\qquad \leqq \left(-\sum_{\sigma'} \xi_{ij}(\varepsilon, l) \lg p_{ij}\right)(H(\xi(\varepsilon, l):P) - \alpha) + (2 + \alpha)\eta. \end{aligned}$$

Now $-\sum_{\sigma'} \xi_{ij}(\varepsilon, l) \lg p_{ij} \geqq \min_{(ij)\epsilon\sigma'}(-\lg p_{ij}) = K_4^{-1} > 0$, where $K_4 < \infty$ depends only on the p_{ij}. If $d = \sup_S H(\xi:P)$, then from (5.26)

$$(5.27) \qquad n^{-1} \lg Q(T(\varepsilon, l, n)) \leqq K_4^{-1}(d - \alpha) + (2 + \alpha)\eta.$$

Given $\alpha > d$, select $\eta > 0$ so that

$$(5.28) \qquad 2\eta \lg K_3 + K_4^{-1}(d - \alpha) + (2 + \alpha)\eta < -\eta.$$

Then choose $\varepsilon < \min(1/2s^4, \varepsilon_0(\eta), \eta)$. If $n \geqq n_2(\alpha, \varepsilon, \eta)$, it follows by (5.21), (5.27), and (5.28) that

$$\sum \mu(v)^\alpha : \mathcal{V}_n(\varepsilon) \leqq L_\varepsilon sK_1(2n\varepsilon)^{s^2} e^{-\eta n}.$$

From this (5.11) follows, and Lemma 5.3 is proved.

LEMMA 5.4. *Suppose $\{x_n\}$ is a regular Markov chain with transition matrix $P = (p_{ij})$. If $\xi \in B$, then*

$$\dim N(\xi) \geqq H(\xi:P),$$

provided Condition (C) *is satisfied.*

Proof. Some of the $r_i(\xi)$ may be 0. We will suppose the last $s - s''$ of them are: $r_i(\xi) > 0$, $i \leqq s''$; $r_i(\xi) = 0$, $i > s''$. The $s'' \times s''$ matrix Z with entries given by $z_{ij} = \xi_{ij}/r_i(\xi)$ if $(i, j) \in \sigma'$, and by $z_{ij} = 0$ if $(i, j) \in \sigma'$ $(i, j \leqq s'')$ is a stochastic matrix. If $z = (r_1(\xi), \cdots, r_{s''}(\xi))$, then $zZ = z$, and hence z is a stationary distribution for Z. Since all z_i are positive, there are no transient states. We will suppose for notational convenience that there are two ergodic classes $\mathcal{A} = \{1, 2, \cdots, s'\}$ and $\mathcal{B} = \{s' + 1, \cdots, s''\}$. The cases of one or of three or more ergodic classes are handled in the same way. Let $a = \sum_{i,j=1}^{s'} \xi_{ij}$ and $b = \sum_{i,j=s'+1}^{s''} \xi_{ij}$. Then $a, b > 0$, $a + b = 1$. Let $\alpha_n = \max(2, [an])$ and $\beta_n = \max(2, [bn])$.

For each n let $(a_{n1}, \cdots, a_{n\alpha_n})$ be a sequence of states with

$$a_{n1} = a_{n\alpha_n} = 1, \qquad a_{nl} \in \mathcal{A}, \qquad (a_{nl}, a_{n,l+1}) \in \sigma',$$

and having transition count $F^{(n)} = (f_{ij}^{(n)})$ $(i, j = 1, \cdots, s')$, where

$$(5.29) \qquad | f_{ij}^{(n)}/\alpha_n - \xi_{ij}/a | = O(\alpha_n^{-1}).$$

Similarly, let $(b_{n1}, \cdots, b_{n\beta_n})$ be a sequence of states with

$$b_{n1} = b_{n\beta_n} = s' + 1, \qquad b_{nl} \in \mathcal{B}, \qquad (b_{nl}, b_{n,l+1}) \in \sigma',$$

and having transition count $G^{(n)} = (g_{ij}^{(n)})$ $(i, j = s' + 1, \cdots, s'')$, where

$$| g_{ij}^{(n)}/\alpha_n - \xi_{ij}/b | = O(\beta_n^{-1}).$$

It is easy to see that such sequences of states exist. Since P has only one ergodic class, there exist sequences $(c_1, \cdots, c_m)$ and $(d_1, \cdots, d_m)$ with $c_1 = d_m = 1$, $c_m = d_1 = s' + 1$ and

$$(c_l, c_{l+1}) \in \sigma', \qquad (d_l, d_{l+1}) \in \sigma', \qquad l = 1, \cdots, m - 1.$$

Let $\gamma_0 = 0$ and $\gamma_n = \alpha_1 + \cdots + \alpha_n + \beta_1 + \cdots + \beta_n + 2n(m-1)$. Let $\mathfrak{V}_n$ consist of all γ_n-cylinders $v = \{\omega : x_l(\omega) = e_l, l = 1, \cdots, \gamma_n\}$ determined by sequences $(e_1, \cdots, e_{\gamma_n})$ with the property that

$$(e_{\gamma_l+1}, \cdots, e_{\gamma_l+\alpha_l-1})$$

has transition count $F^{(l+1)}$, $(e_{\gamma_l+\alpha_l+m-1}, \cdots, e_{\gamma_l+\alpha_l+m+\beta_l})$ has transition count $G^{(l+1)}$, $(e_{\gamma_l+\alpha_l}, \cdots, e_{\gamma_l+\alpha_l+m-1}) = (c_1, \cdots, c_m)$, and

$$(e_{\gamma_{l+1}-m+1}, \cdots, e_{\gamma_{l+1}}) = (d_1, \cdots, d_m),$$

for $l = 0, 1, \cdots, n-1$. These sequences are exactly the ones which can be constructed in the following way. Take $\mathfrak{a}_1, \cdots, \mathfrak{a}_n, \mathfrak{b}_1, \cdots, \mathfrak{b}_n$ to be sequences of states with transition counts $F^{(1)}, \cdots, F^{(n)}, G^{(1)}, \cdots, G^{(n)}$, respectively. Form the compound sequence

$$\mathfrak{a}_1 \mathfrak{b}_1 \mathfrak{a}_2 \mathfrak{b}_2 \cdots \mathfrak{a}_n \mathfrak{b}_n .$$

Each $\mathfrak{a}_l$ starts and ends with 1, and each $\mathfrak{b}_l$ starts and ends with $s' + 1$. Between each $\mathfrak{a}_l$ and $\mathfrak{b}_l$ insert $(c_2, \cdots, c_{m-1})$, and between each $\mathfrak{b}_l$ and $\mathfrak{a}_{l+1}$ insert $(d_2, \cdots, d_{m-1})$.

Let $J^{(n)}$ be the $s \times s$ partitioned matrix

$$J^{(n)} = \begin{bmatrix} \sum_{l=1}^n F^{(l)} & 0 & 0 \\ 0 & \sum_{l=1}^n G^{(l)} & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Let J' and J'' be the transition counts of $(c_1, \cdots, c_m)$ and $(d_1, \cdots, d_m)$, respectively. Finally, let $K^{(n)} = (k_{ij}^{(n)}) = J^{(n)} + nJ' + (n-1)J''$. Then any sequence $(e_1, \cdots, e_{\gamma_n})$ satisfying the requirements of the preceding paragraph has transition count $K^{(n)}$.

It follows by (5.29) that $| \sum_{l=1}^n f_{ij}^{(l)} / \sum_{l=1}^n \alpha_l - \xi_{ij}/a | = O(\sum_{l=1}^n \alpha_l)^{-1}$. Since $\sum_{l=1}^n f_{ij}^{(l)} \leqq \gamma_n$ and $\gamma_n(a/\sum_{l=1}^n \alpha_l - 1/\gamma_n) = O(1)$, we have

$$\left| \sum_{l=1}^n f_{ij}^{(l)}/\gamma_n - \xi_{ij} \right| = O\left(\sum_{l=1}^n \alpha_l\right)^{-1}.$$

Similarly, $| \sum_{l=1}^n g_{ij}^{(l)}/\gamma_n - \xi_{ij} | = O(\sum_{l=1}^n \beta_l)^{-1}$. Since $n/\gamma_n \to 0$, it follows that

$$(5.30) \qquad \begin{aligned} | k_{ij}^{(n)}/\gamma_n - \xi_{ij} | &= O(\gamma_n^{-1}), & (i, j) &\in \sigma', \\ k_{ij}^{(n)} &= 0, & (i, j) &\notin \sigma'. \end{aligned}$$

Suppose $(e_1, e_2, \cdots)$ is an infinite sequence of states such that for each n, $(e_1, \cdots, e_{\gamma_n})$ has $K^{(n)}$ as transition count. If $H^{(n)}$ is the transition count for $(e_1, \cdots, e_n)$, it follows from (5.30) and the fact that $\gamma_n \sim n^2/2$ that $| k_{ij}^{(n)}/n - \xi_{ij} | = O(n^{-1})$, $(i, j) \in \sigma'$. Recalling the definition of $\mathfrak{V}_n$, we see that $\bigcap_{n=1}^\infty V_n \subset N(\xi)$. Therefore it suffices to prove

$$(5.31) \qquad \dim \bigcap_{n=1}^\infty V_n \geqq H(\xi : P).$$

Let ν_n be the number of v' of $\mathcal{V}_n$ contained in a single v of $\mathcal{V}_{n-1}$. Then ν_n does not depend on v, and in fact $\nu_n = N_1^{(\alpha_n)}(F^{(n)})N_{s'+1}^{(\beta_n)}(G^{(n)})$. From this and Lemma 5.2 we have

$$\nu_n = \frac{\prod_{j=1}^{s'} f_{\cdot j}^{(n)}!}{\prod_{i,j=1}^{s'} f_{ij}^{(n)}!}\hat{F}_{\cdot 1}^{(n)} \frac{\prod_{j=s'+1}^{s''} g_{\cdot j}^{(n)}!}{\prod_{i,j=s'+1}^{s''} g_{ij}^{(n)}!}\hat{G}_{\cdot s'+1}^{(n)}.$$

As $n \to \infty$, $F^{(n)}$ goes elementwise to the $s' \times s'$ matrix Y with entries y_{ij} given by

$$y_{ij} = \delta_{ij} - \xi_{ij}/r_j(\xi), \qquad i, j \leqq s', \quad (i, j) \in \sigma',$$

$$y_{ij} = \delta_{ij}, \qquad i, j \leqq s', \quad (i, j) \notin \sigma'.$$

The upper left-hand $s' \times s'$ submatrix of Z (defined above) is an indecomposable stochastic matrix since $\mathcal{A}$ is an ergodic class. It follows easily that the transpose $(I - Y)^*$ of $I - Y$ is an indecomposable stochastic matrix with stationary distribution $(r_1(\xi), \cdots, r_{s'}(\xi))$. Since

$$\sum_{i=1}^{s'} y_{ij} = 0, \qquad j = 1, \cdots, s',$$

we have $Y_{ij} = Y_{\cdot j}$ and $\sum_{i=1}^{s'} y_{ij} Y_{\cdot j} = 0$, $j = 1, \cdots, s'$ by Lemma 4.1. Since $(I - Y)^*$ has a unique stationary distribution, we must have

$$Y_{\cdot j} = kr_j(\xi), \qquad j = 1, \cdots, s'$$

for some constant k. If k were 0, the rank of Y^* would be less than $s' - 1$, which is impossible since $I - Y^*$ is indecomposable. Since $\hat{F}^{(n)} \to Y$ elementwise, we have $\hat{F}_{\cdot 1}^{(n)} \to Y_{\cdot 1}$. Because of its combinatorial significance, $\hat{F}_{\cdot 1}^{(n)}$ is positive, and hence k must be positive also. Thus $\hat{F}_{\cdot 1}^{(n)}$ goes to a positive constant, and it follows that

$$\lim_{n\to\infty} \alpha_n^{-1} \lg \hat{F}_{\cdot 1}^{(n)} = 0. \tag{5.32}$$

Now as $n \to \infty$ we have, by Stirling's formula and the fact that

$$\sum_{i,j=1}^{s'} \xi_{ij} \lg \xi_{ij}/r_j(\xi) = \sum_{i,j=1}^{s'} \xi_{ij} \lg \xi_{ij}/r_i(\xi)$$

since $\xi \in B$,

$$\alpha_n^{-1} \lg \prod_{j=1}^{s'} f_{\cdot j}^{(n)}!/\prod_{i,j=1}^{s'} f_{ij}^{(n)}! = -\sum_{i,j=1}^{s'} a^{-1}\xi_{ij} \lg \xi_{ij}/r_j(\xi) + o(1). \tag{5.33}$$

By (5.32), (5.33), similar expressions for the $g_{ij}^{(n)}$, and the fact that $\alpha_n/(\alpha_n + \beta_n) \to a$, we have

$$(\alpha_n + \beta_n)^{-1} \lg \nu_n = -\sum_{\sigma'} \xi_{ij} \lg \xi_{ij}/r_i(\xi) + o(1). \tag{5.34}$$

If $v \in \mathcal{V}_{n-1}$ and $v' \in \mathcal{V}_r$ and $v_n \subset v_{n-1}$, then

$$\mu(v')/\mu(v) = p_0 \prod_{i,j=1}^{s'} p_{ij}^{f_{ij}^{(n)}} \prod_{i,j=s'+1}^{s''} p_{ij}^{g_{ij}^{(n)}} = \rho_n,$$

where $p_0 = p_{c_1c_2} \cdots p_{c_{m-1}c_m} p_{d_1d_2} \cdots p_{d_{m-1}d_m}$ and ρ_n does not depend on v and v'. As $n \to \infty$,

$$\alpha_n^{-1} \lg \prod_{i,j=1}^{s'} p_{ij}^{f_{ij}^{(n)}} = \sum_{i,j=1}^{s'} a^{-1}\xi_{ij} \lg p_{ij} + o(1).$$

From this and a similar expression for the $g_{ij}^{(n)}$ we get

$$(\alpha_n + \beta_n)^{-1} \lg \rho_n = \sum_{\sigma'} \xi_{ij} \lg p_{ij} + o(1). \tag{5.35}$$

Given $\alpha < H(\xi:P)$, choose ε so that

$$0 < \varepsilon < -\tfrac{1}{6} \sum_{\sigma'} \xi_{ij} \lg p_{ij}(H(\xi:P) - \alpha). \tag{5.36}$$

By (5.34) and (5.35) there exists an n_0 such that if $n \geqq n_0$, then

$$(\alpha_n + \beta_n)^{-1} \lg \nu_n \geqq -\sum_{\sigma'} \xi_{ij} \lg \xi_{ij}/r_i(\xi) - \varepsilon,$$

$$\left| \sum_{\sigma'} \xi_{ij} \lg p_{ij} - (\alpha_n + \beta_n)^{-1} \lg \rho_n \right| \leqq \varepsilon.$$

It follows that if $n \geqq n_0$ then

$$\frac{1}{\gamma_n - \gamma_{n_0-1}} \lg \frac{\rho_{n_0} \cdots \rho_{n-1}}{(\rho_{n_0} \cdots \rho_n)^{1+\alpha}(\nu_{n_0} \cdots \nu_n)}$$

$$\leqq 4\varepsilon - \frac{\alpha_n + \beta_n}{\gamma_n - \gamma_{n_0-1}} \sum_{\sigma'} \xi_{ij} \lg p_{ij} - \alpha \sum_{\sigma'} \xi_{ij} \lg p_{ij} + \sum_{\sigma'} \xi_{ij} \lg \xi_{ij}/r_i(\xi).$$

Since the second term on the right goes to 0, and ε satisfies (5.36), there is some $n_1 \geqq n_0$ such that

$$\frac{1}{\gamma_n - \gamma_{n_0-1}} \lg \frac{(\rho_{n_0} \cdots \rho_{n-1})}{(\rho_{n_0} \cdots \rho_n)^{1+\alpha}(\nu_{n_0} \cdots \nu_n)} < -\varepsilon$$

provided $n \geqq n_1$. Since $\gamma_n \sim n^2/2$, it follows that (4.1) holds. Therefore (5.31), and hence Lemma 5.4, follows from Theorem 4.3.

We now prove Theorem 5.1. With the definition (5.3) we clearly have $N(\xi) \subset M(\xi)$, $M(S) \subset N(S)$, $\bigcup_S N(\xi) \subset N(S)$, and $\bigcup_S M(\xi) \subset M(S)$. It follows from Lemmas 5.3 and 5.4 and Theorem 4.1 that $\bigcup_S N(\xi)$, $\bigcup_S M(\xi)$, $M(S)$, $N(S)$ all have the same dimension, namely $\sup_S H(\xi:P)$, and Theorem 5.1 is proved. It follows also for example that

$$\dim \bigcup_S M(\xi) = \sup_S \dim M(\xi),$$

and if S is uncountable, we have an instance where Theorem 4.1 (iii) holds with an uncountable index set.

The dimensions of a number of interesting sets defined in terms of the asymptotic properties of the relative transition frequencies can be obtained through Theorem 5.1. In each case (5.4) gives the answer, and there is no more Hausdorff dimension theory involved. However, the supremum in (5.4) may be difficult to compute.

Before proceeding to an example we recast Theorem 5.1 in a more suggestive form.

THEOREM 5.2. *Suppose that* $\{x_n\}$ *is a regular Markov chain with transition matrix* $P = (p_{ij})$. *Suppose that* ϕ *is a continuous mapping of* A *into some regular space* Γ. *Suppose that* $G \subset \Gamma$ *and* $\phi^{-1}G \neq 0$. *Then*

$$\dim \{\omega:\phi(\delta(\omega, n)) \to G\} = \dim \bigcup_G \{\omega:\phi(\delta(\omega, n)) \to \gamma\}$$
$$= \sup \{H(\xi:P):\phi(\xi) \in \bar{G}\},$$

provided Condition (C) *holds.*

Proof. If $\gamma_n \in \Gamma$, then by $\gamma_n \to G$ we mean that if G' is any open set containing G, then $\gamma_n \in G'$ for all but a finite number of n. Let

$$S = \{\xi:\phi(\xi) \in \bar{G}\}.$$

Since ϕ is continuous, $\phi(\xi_n) \to G$ if and only if $\xi_n \to S$. Thus the theorem follows from Theorem 5.1.

As an example of Theorem 5.2, take $\phi(\xi) = H(\xi:P)$. If $0 \leqq \gamma \leqq 1$ and $G = \{\gamma\}$, then the ω set where

$$\lim_{n\to\infty} \frac{\sum_\sigma \delta_{ij}(\omega, n) \lg \delta_{ij}(\omega, n)}{\sum_\sigma \delta_{ij}(\omega, n) \lg p_{ij}} = \lim_{n\to\infty} H\big(\delta(\omega, n): P\big) = \gamma$$

has dimension γ. If we take $G = [0, \gamma]$, it follows that the ω set where

$$\lim_n H(\delta(\omega, n):P) \leqq \gamma$$

and the ω set where $\limsup_n H(\delta(\omega, n):P) \leqq \gamma$ each have dimension γ also. Further examples are given in §7.

In this section we have for simplicity confined ourselves to the regular case. It is possible to reformulate Theorem 5.1 so as to include the case of more than one ergodic class.

Since a Markov chain of arbitrary order can be reduced to one of first order, it is clear that if $\{x_n\}$ is such a chain, and if M is defined in terms of the asymptotic relative frequencies of occurrence of the various k-tuples of states, then one can in principle write down the dimension of M.

6. A theorem on generalized derivatives

As before let $(\Omega, \mathcal{B})$ be a measurable space, and let $\{x_n\}$ be a sequence of measurable functions. Suppose we have two probability measures μ and ν on $\mathcal{B}$. We will denote dimensions computed with respect to μ and ν (and $\{x_n\}$) by $\dim_\mu$ and $\dim_\nu$ respectively. Now suppose that $\{x_n\}$ is a regular Markov chain with respect to μ and with respect to ν. That is, suppose that

$$\mu\{x_{n+1} = u_{n+1} \| x_l = u_l, \, l \leqq n\} = p_{u_n u_{n+1}},$$

$$\nu\{x_{n+1} = u_{n+1} \| x_l = u_l, \, l \leqq n\} = q_{u_n u_{n+1}},$$

where $P = (p_{ij})$ and $Q = (q_{ij})$ are regular stochastic matrices. We assume that P and Q are distinct. Let the stationary probabilities corresponding to P and Q be $\{p_i\}$ and $\{q_i\}$, respectively. We do not assume however that $\mu\{x_1 = i\} = p_i$ or $\nu\{x_1 = i\} = q_i$. We will assume that Condition (C) is satisfied by $\{x_n\}$ together with μ.

For each $\omega \in \Omega$ and $n = 1, 2, \cdots$, let $v_n(\omega)$ be the n-cylinder

$$v_n(\omega) = \{\omega' : x_l(\omega') = x_l(\omega),\ l = 1, \cdots, n\}.$$

Now let

$$f_n(\omega) = \mu(v_n(\omega))/\nu(v_n(\omega)) \quad \text{if} \quad \nu(v_n(\omega)) > 0,$$
$$= 0 \qquad\qquad\qquad\qquad \text{otherwise.}$$

Let $f(\omega)$ be the Radon-Nykodym derivative of (the absolutely continuous component of) μ with respect to ν. Of course f is determined only to within a set of ν-measure 0. It follows from the theory of derivatives (see Chapter VII, §8, of [3]) that $f_n(\omega) \to f(\omega)$ except on a set of ν-measure 0. Since by the strong law of large numbers for Markov chains (see Chapter V, §6, of [3]) we have $\delta_{ij}(\omega, n) \to p_i\, p_{ij}$ on a set of μ-measure 1, and $\delta_{ij}(\omega, n) \to q_i\, q_{ij}$ on a set of ν-measure 1, and since $P \neq Q$, it is clear that μ and ν are singular with respect to each other. Hence $f(\omega) = 0$, and the limit

$$\text{(6.1)} \qquad \lim_{n\to\infty} \nu(v_n(\omega))/\mu(v_n(\omega))^{\alpha}$$

is ∞ if $\alpha = 1$, except on a set of ν-measure 0. On the other hand, the limit (6.1) is 0 if $\alpha = 0$. Thus it is natural to investigate the set of α such that (6.1) is 0 except on a set of ν-measure 0. This amounts to finding a generalized Lipschitz condition.

THEOREM 6.1. *Under the above assumptions there exists a set $E \in \mathcal{B}$ such that*

(i) $\nu(E) = 1$,

(ii) $\dim_\mu E = H(Q:P) = \sum q_i\, q_{ij} \lg q_{ij} / \sum q_i\, q_{ij} \lg p_{ij}$,

(iii) *if $\omega \in E$, then the limit* (6.1) *equals* 0 *if* $0 \leqq \alpha < H(Q:P)$ *and equals* ∞ *if* $\alpha > H(Q:P)$.

Proof. Let E be the ω set where $\delta_{ij}(\omega, n) \to q_i\, q_{ij}$. Then (i) follows from the strong law of large numbers and (ii) from Theorem 5.1. Now if $\omega \in E$, then

$$\nu(v_n(\omega)) = \nu\{\omega' : x_1(\omega') = x_1(\omega)\} \prod_{ij} q_{ij}^{n q_i q_{ij} + o(n)},$$

and hence

$$\text{(6.2)} \qquad \lg \nu(v_n(\omega)) = n \sum q_i\, q_{ij} \lg q_{ij} + o(n).$$

Similarly,

$$\text{(6.3)} \qquad \lg \mu(v_n(\omega)) = n \sum q_i\, q_{ij} \lg p_{ij} + o(n).$$

But (iii) follows immediately from (6.2) and (6.3).

If $(\Omega, \mathcal{B}, \{x_n\}, \mu)$ constitutes the Lebesgue case, and ν is a second probability measure on $(\Omega, \mathcal{B})$ satisfying the above conditions, then Theorem 6.1 reduces to a result of Kinney [11]. In this case (iii) becomes an ordinary Lipschitz condition.

7. The independent case

We now specialize the results of §5 to the case in which the process $\{x_n\}$ is independent. Let A' be the set of vectors $\zeta = (\zeta_1, \cdots, \zeta_s)$ of s-space such that $\zeta_i \geqq 0$ and $\sum_\sigma \zeta_i = 1$. Let ρ' denote Euclidean distance on A'. If $\delta_i(\omega, n)$ is defined by (5.1), and $\delta'(\omega, n) = (\delta_1(\omega, n), \cdots, \delta_s(\omega, n))$, then $\delta'(\omega, n) \in A'$. If ζ, $p \in A'$, let $H(\zeta:p) = \sum_\sigma \zeta_i \lg \zeta_i / \sum_\sigma \zeta_i \lg p_i$. For an information-theoretic interpretation of $H(\zeta:p)$ see §8. If $S' \subset A'$, let

$$M'(S') = \{\omega:\delta'(\omega, n) \to S'\}$$

$$N'(S') = \{\omega:\rho'(\delta'(\omega, n), S') = O(n^{-1})\}.$$

THEOREM 7.1. *Suppose $\{x_n\}$ is an independent process with*

$$\mu\{\omega:x_n(\omega) = i\} = p_i > 0.$$

If $S' \subset A'$, then

$$\begin{aligned} \dim M'(S') &= \dim N'(S') = \dim \textstyle\bigcup_{S'} M'(\zeta) \\ &= \dim \textstyle\bigcup_{S'} N'(\zeta) = \sup_{S'} H(\zeta:p), \end{aligned} \tag{7.1}$$

provided Condition (C) *holds.*[4]

Proof. Let S be the set of $\xi \in B$ such that $(r_1(\xi), \cdots, r_s(\xi)) \in S'$. It follows from Theorem 5.1 that the sets listed in (7.1) have $\sup_S H(\xi:P)$ as their common dimension, where $P = (p_{ij})$ with $p_{ij} = p_j$. Since a conditional entropy never exceeds the corresponding entropy [14],

$$\sum_\sigma r_i(\xi) \left[- \sum_{\sigma'} \frac{\xi_{ij}}{r_i(\xi)} \lg \frac{\xi_{ij}}{r_i(\xi)} \right] \leqq - \sum_\sigma c_j(\xi) \lg c_j(\xi),$$

or, assuming $\xi \in B$,

$$- \sum_{\sigma'} \xi_{ij} \lg \xi_{ij}/r_i(\xi) \leqq - \sum_\sigma r_i(\xi) \lg r_i(\xi).$$

Hence

$$\sup_S H(\xi:P) \leqq \sup_S \frac{\sum_\sigma r_i(\xi) \lg r_i(\xi)}{\sum_\sigma r_i(\xi) \lg p_i} = \sup_{S'} H(\zeta:p).$$

On the other hand, if $\zeta \in S'$, then $(\zeta_i \zeta_j) \in S$. Hence

$$\sup_S H(\xi:P) \geqq \sup_{S'} H((\zeta_i \zeta_j):P) = \sup_{S'} H(\zeta:p),$$

proving the theorem.

Consider Theorem 7.1 with $\zeta_i \equiv p_i$, and suppose S' consists of the single point $(p_1, \cdots, p_s)$. We get

$$\dim \{\omega: | \delta_i(\omega, n) - p_i | = O(n^{-1}), i = 1, \cdots, s\} = 1.$$

[4] A referee points out that in the Lebesgue case with $s = 2$ and $S' = \{\zeta:0 \leqq \zeta_1 \leqq \alpha\}$, the dimension of $M'(S')$ has been obtained by A. S. BESICOVITCH, *On the sum of digits of real numbers represented in the dyadic system*, Math. Ann., vol. 110 (1934), pp. 321–330.

It is an immediate consequence of the law of the iterated logarithm [6] that

$$\mu\{\omega: |\delta_i(\omega, n) - p_i| = o(n^{-1} \lg\lg n)^{1/2}, i = 1, \cdots, s\} = 0.$$

Thus if one considers the set where $\delta_i(\omega, n) \to p_i$ at a certain rate and then increases the rate, the measure of the set drops to 0 long before the dimension drops below 1.

Theorem 7.1 can be specialized to the following one just as in §5.

THEOREM 7.2. *Suppose that* $\{x_n\}$ *is an independent process with*

$$\mu\{\omega: x_n(\omega) = i\} = p_i > 0.$$

Suppose that ϕ *is a continuous mapping of* A' *into some regular space* Γ. *Suppose that* $G \subset \Gamma$ *and* $\phi^{-1}G \neq 0$. *Then*

$$\dim\{\omega:\phi(\delta'(\omega, n)) \to G\} = \dim \mathsf{U}_G\{\omega:\phi((\delta'(\omega, n)) \to \gamma\}$$
$$= \sup\{H(\zeta:p):\phi(\zeta) \in \bar{G}\},$$

provided Condition (C) *holds.*

We indicate briefly some applications of Theorems 7.1 and 7.2. In the Lebesgue case with S' consisting of a single point, Theorem 7.1 gives

$$\sum \zeta_i \lg \zeta_i/\lg s^{-1}$$

as the classical Hausdorff dimension of the ω set where $\delta_i(\omega, n) \to \zeta_i$ for each i, a result originally due to Eggleston [4]. In [7] Good gives the dimension of $M'(\zeta)$ in the general independent case. Good's formulation of the problem is somewhat different from the present one since he takes Ω to be the unit interval, μ to be Lebesgue measure, and then defines $\{x_n\}$ as what he calls a "generalized decimal".

If $0 \leqq \gamma \leqq s - 1$, let r be the root of $\sum_{i=0}^{s-1} (i - \gamma)z^i = 0$ lying between 0 and 1, and let $f(\gamma) = \lg(1 + r + \cdots + r^{s-1})/\lg s$. Eggleston [5] has shown in the Lebesgue case that the dimension of the ω set where

$$\lim_n \sum_{k=1}^n x_k(\omega)/n = \gamma$$

is $f(\gamma)$. He has shown further that if $\gamma \leqq (s - 1)/2$, then $f(\gamma)$ is also the dimension of the ω set where $\lim_n \sum_{k=1}^n x_k(\omega)/n \leqq \gamma$. (A simplified proof of this latter fact has been given by Kinney [11].) These results follow from Theorem 7.2 if one takes $\phi(\zeta) = \sum_{i=0}^{s-1} i\zeta_i$ and computes the supremum involved by Lagrange multipliers. Theorem 7.2 then also gives $f(\gamma)$ as the dimension of the ω set where

$$\limsup_n \sum_{k=1}^n x_k(\omega)/n \leqq \gamma,$$

provided $\gamma \leqq (s - 1)/2$, as well as results on the rate of convergence.

If Γ is Euclidean and ϕ is a linear transformation, $\phi(\zeta) = R\zeta$, we have as a special case of Theorem 7.1

$$\dim\{\omega:R\delta'(\omega, n) \to \gamma\} = \sup\{H(\zeta:p):R\zeta = \gamma\}, \tag{7.2}$$

assuming that $\gamma \in RA'$. A result essentially equivalent to (7.2) has been obtained in the Lebesgue case by Volkmann [17, Part IV] under the additional assumption that the matrix R of the transformation has the special property that in each column there is exactly one positive entry, the others being 0. Volkmann's result appears quite different from (7.2), even in the Lebesgue case, because his method gives the supremum in a different form. The equivalence of the two forms can be established by maximizing $H(\xi:p)$ subject to $R\zeta = \gamma$ by the method of Lagrange multipliers.

As a final illustration, consider the set

$$T(\gamma) = \{\omega: -\textstyle\sum_{i=1}^{s} \delta_i(\omega, n) \lg p_i \to \gamma\},$$

where $\min_i \lg p_i^{-1} \leqq \gamma \leqq \max_i \lg p_i^{-1}$. If $\gamma = -\sum_{i=1}^{s} p_i \lg p_i$, the entropy [14] of $(p_1, \cdots, p_s)$, it follows as a special case of the Shannon-McMillan theorem [13, 14] that $\mu(T(\gamma)) = 1$, so that $\dim T(\gamma) = 1$. For general γ let r be the positive root of $\sum_{i=1}^{s} (\gamma + \lg p_i)p_i^z = 0$. Then in the general independent case $\dim T(\gamma) = r + \gamma^{-1} \sum_{i=1}^{s} p_i^r$. Again this result is obtained by Lagrange multipliers.

8. Connection with information theory

If $p = (p_1, \cdots, p_s)$ and $q = (q_1, \cdots, q_s)$ are two probability distributions on $\sigma = \{1, 2, \cdots, s\}$, then, as in §7, define

$$H(q:p) = \textstyle\sum q_i \lg q_i / \sum q_i \lg p_i,$$

provided $q_i = 0$ for any i such that $p_i = 0$. If $p_i \equiv 1/s$ then $H(q:p)$ becomes $H(q) = -\sum q_i \lg q_i / \lg s$, the *relative entropy* of q [14].[5] For a general p it is thus natural to call $H(q:p)$ the *relative entropy of q with respect to p*. Now Leibler and Kullback [12] define the function

$$I(q:p) = \textstyle\sum q_i \lg (q_i/p_i)$$

and interpret it as the mean information per observation for discrimination between p and q when q is the true distribution. They then define (in case $p_i = 0$ if and only if $q_i = 0$) $J(q, p) = I(q:p) + I(p:q)$ as the divergence between the distributions p and q.[6] It is a simple matter to show that

$$H(q:p) = H(q)/(H(q) + I(q:p)).$$

We now turn to the Markov case. Let P and Q be regular stochastic matrices with stationary distributions p and q, where we assume that $q_{ij} = 0$ if $p_{ij} = 0$. As in §5 we set

$$H(Q:P) = \textstyle\sum_{ij} q_i q_{ij} \lg q_{ij} / \sum_{ij} q_i q_{ij} \lg p_{ij}.$$

If $p_{ij} \equiv 1/s$, this becomes $H(Q) = -\sum_{ij} q_i q_{ij} \lg q_{ij}/\lg s$, the relative

[5] That the dimension, in the Lebesgue case, of the ω set where $\delta_i(\omega, n) \to q_i$, is this relative entropy has been noted by Good [8] and Shannon [15].

[6] Their definitions apply of course to distributions on spaces much more general than σ.

entropy [14] of a Markov source with matrix Q. Such a source we will denote by Q itself. Just as in the independent case, we take $H(Q:P)$ to be the relative entropy of Q with respect to P.

Suppose we want to decide whether a source is governed by Q or by P. If we have received a number of symbols from the source, and the last one is i, then according to the Leibler-Kullback definition, the mean amount of information the next symbol from the source contains for making the decision is $\sum_j q_{ij} \lg (q_{ij}/p_{ij})$, if the source is in fact governed by Q. Averaging over i we have

$$I(Q:P) = \sum_{ij} q_i q_{ij} \lg (q_{ij}/p_{ij})$$

as the information per symbol for discrimination between P and Q when the source is actually Q, an extension of the Leibler-Kullback notion. If $q_{ij} = 0$ implies $p_{ij} = 0$ as well as conversely, we may take

$$J(Q, P) = I(Q:P) + I(P:Q)$$

as a definition of divergence between P and Q. As in the independent case it is easy to show that

$$H(Q:P) = H(Q)/(H(Q) + I(Q:P)).$$

Let σ_n be the set of sequences $a = (a_1, a_2, \cdots, a_n)$ with $a_l \in \sigma$. Let $p^{(n)}$ be the distribution on σ_n defined by $p^{(n)}(a) = p_{a_1} p_{a_1 a_2} \cdots p_{a_{n-1} a_n}$, and define $q^{(n)}$ similarly in terms of Q. Then $p^{(n)}$ $[q^{(n)}]$ simply gives the probabilities of obtaining the various sequences of length n from the source P $[Q]$. Let $I(q^{(n)}:p^{(n)})$ be the ordinary Leibler-Kullback function, and take

$$I_n(Q:P) = I(q^{(n)}:p^{(n)}).$$

One can show that $\lim_n n^{-1} I_n(Q:P) = I(Q:P)$, a result well known in the case $p_{ij} \equiv 1/s$ [14].

We have shown that the dimensions obtained in the preceding sections all have information-theoretic interpretations. The appearance of the supremum in the theorems shows that a "maximum-entropy principle" is a central feature of the theory.

References

1. E. Borel, *Les probabilités dénombrables et leurs applications arithmétiques*, Rend. Circ. Mat. Palermo, vol. 27 (1909), pp. 247–271.
2. W. Dawson and I. J. Good, *Exact Markov probabilities from oriented linear graphs*, Ann. Math. Statistics, vol. 28 (1957), pp. 946–956.
3. J. L. Doob, *Stochastic processes*, New York, 1953.
4. H. G. Eggleston, *The fractional dimension of a set defined by decimal properties*, Quart. J. Math. Oxford Ser., vol. 20 (1949), pp. 31–36.
5. ———, *Sets of fractional dimensions which occur in some problems of number theory*, Proc. London Math. Soc. (2), vol. 54 (1951–1952), pp. 42–93.
6. W. Feller, *An introduction to probability theory and its applications*, 2nd ed., New York, 1957.

7. I. J. GOOD, discussion of a paper, J. Roy. Statist. Soc. Ser. B, vol. 13 (1951), pp. 61–62.
8. ———, discussion of a paper, Transactions of the Institute of Radio Engineers, Professional Group on Information Theory, February, 1953, p. 170.
9. L. A. GOODMAN, *Exact probabilities and asymptotic relationships for some statistics from m-th order Markov chains*, Ann. Math. Statistics, vol. 29 (1958), pp. 476–490.
10. F. HAUSDORFF, *Dimension und äusseres Mass*, Math. Ann., vol. 79 (1919), pp. 157–179.
11. J. R. KINNEY, *Singular functions associated with Markov chains*, Proc. Amer. Math. Soc., vol. 9 (1958), pp. 603–608.
12. S. KULLBACK AND R. A. LEIBLER, *On information and sufficiency*, Ann. Math. Statistics, vol. 22 (1951), pp. 79–86.
13. B. MCMILLAN, *The basic theorems of information theory*, Ann. Math. Statistics, vol. 24 (1953), pp. 196–219.
14. C. E. SHANNON, *A mathematical theory of communication*, Bell System Tech. J., vol. 27 (1948), pp. 379–423, 623–656.
15. ———, discussion of a paper, Transactions of the Institute of Radio Engineers, Professional Group on Information Theory, February, 1953, pp. 173–174.
16. B. VOLKMANN, *Über Klassen von Mengen natürlicher Zahlen*, J. Reine Angew. Math., vol. 190 (1952), pp. 199–230.
17. ———, *Über Hausdorffsche Dimensionen von Mengen, die durch Zifferneigenschaften charakterisiert sind*, Math. Zeitschr., *I*, vol. 58 (1953), pp. 284–287; *II*, vol. 59 (1954), pp. 247–254; *III*, vol. 59 (1954), pp. 259–270; *IV*, vol. 59 (1954), pp. 425–433.
18. P. WHITTLE, *Some distribution and moment formulae for the Markov chain*. J. Roy. Statist. Soc. Ser. B, vol. 17 (1955), pp. 235–242.

THE UNIVERSITY OF CHICAGO
CHICAGO, ILLINOIS