The Annals of Probability
1988, Vol. 16, No. 1, 344–354

A DE FINETTI THEOREM FOR A CLASS OF PAIRWISE INDEPENDENT STATIONARY PROCESSES

BY JAMES B. ROBERTSON AND STEPHEN SIMONS

University of California at Santa Barbara

Consider a $\{0,1\}$-valued strictly stationary stochastic process $\{X_1, X_2, \dots\}$. Let k and l be natural numbers and define $y_i = 0$ or 1 according as $x_1 + \cdots + x_{i+k-1}$ is even or odd. Then, for $1 \le j \le l$ set $S_j(x_1 \cdots x_n) = \Sigma_{0 \le i \le m-1} y_{j+il}$. We consider all processes that have $(S_1, \dots, S_l)$ as sufficient statistics. We obtain explicit formulas for the distributions of the processes that are extreme points. We also represent these processes as finitary processes and use this representation to investigate their pairwise independence, ergodicity and mixing properties.

1. Introduction. Robertson and Womack (1985) considered a class of pairwise independent, ergodic stationary stochastic processes $\{X_n\colon n = 1, 2, \dots\}$ such that $P[X_n = 0] = P[X_n = 1] = \frac{1}{2}$. A one parameter family of joint distributions was constructed for such processes and sufficient statistics for this parameter were calculated by Robertson (1985). Whenever sufficient statistics are known, it is a natural problem to try to describe the set of all distributions which have these as sufficient statistics. [See, for example, Diaconis and Freedman (1980).] For example, if $X_1, X_2, \dots$ are independent identically distributed random variables with $P[X_n = 0] + P[X_n = 1] = 1$, then the partial sums $X_1 + \cdots + X_n$ are sufficient statistics. The processes for which $X_1 + \cdots + X_n$ is a sufficient statistic are the *exchangeable* processes. A de Finetti theorem then asserts that any exchangeable process is an average of (perhaps an infinite number) of independent identically distributed processes.

In this paper we will consider a generalization of the statistics considered by Robertson (1985). Next we shall describe all processes with these sufficient statistics. Our description will identify the extreme points of this convex set of distributions and then a theorem of Ressel (1985) will show that all of the processes are averages of these extreme point processes. In this case we obtain as extreme points many new processes. Finally we shall investigate the properties of these new processes. Finally we shall investigate the properties of these new processes. It will be shown, in particular, that the only pairwise independent processes with the sufficient statistics given by Robertson (1985) are the ones he described. We shall also identify which of these processes are ergodic, i.e., which are extreme points in the larger convex set of all stationary stochastic processes.

2. The statistics. We shall first describe the distributions of all two-valued stationary stochastic processes $\{X_n\colon n = 1, 2, \dots\}$. Let $\Gamma = \{0, 1\}$ and let Γ^*

Received January 1986; revised October 1986.
AMS 1980 *subject classifications*. Primary 60G10; secondary 28D05.
Key words and phrases. de Finetti theorem, stationary stochastic process, finitary process, pairwise independence, ergodic, weakly mixing, extreme point of compact convex set.

denote the free monoid with identity ε generated by Γ. That is, Γ^* consists of all words $w = x_1 \cdots x_n$ on Γ of finite length $|w| = n$, ε is the empty word and multiplication of two words is concatenation. For $n \geq 0$, we will denote the words of length n by Γ_n^*. We define a function P on Γ^* by

$$P(\varepsilon) = 1 \quad \text{and} \quad P(x_1 \cdots x_n) = P[X_1 = x_1, \dots, X_n = x_n]. \tag{2.1}$$

Kolmogorov's extension theorem ensures that the probabilities in (2.1) completely determine the distribution of the process. Functions P on Γ^* that yield distributions of stochastic processes are characterized by the following conditions:

$$P(\varepsilon) = 1, \tag{2.2}$$

$$P(w) \geq 0 \quad \text{for all } w \in \Gamma^*, \tag{2.3}$$

$$P(w) = \sum_{g \in \Gamma} P(wg) \quad \text{for all } w \in \Gamma^*. \tag{2.4}$$

The stationary processes are characterized by the additional condition

$$P(w) = \sum_{g \in \Gamma} P(gw) \quad \text{for all } w \in \Gamma^*. \tag{2.5}$$

The set $\Xi \subset [0,1]^{\Gamma^*}$ of all functions satisfying (2.2)–(2.5) is a compact convex set. Its extreme points correspond to the ergodic processes.

The distribution of the processes studied by Robertson (1985) is given by the following, where $0 \leq a \leq \frac{1}{2}$. If $n = 2u$ is even,

$$P(x_1 \cdots x_n) = \left(a^R\left(\tfrac{1}{2} - a\right)^{u-1-R} + a^S\left(\tfrac{1}{2} - a\right)^{u-1-S}\right)/8, \tag{2.6}$$

where R and S are statistics defined as follows: Set $y_i = 0$ or 1 according as there are an even or odd number of 1's in x_i, x_{i+1}, x_{i+2}. Now set

$$R(x_1 \cdots x_n) = \sum_{1 \leq i \leq u-1} y_{2i} \quad \text{and} \quad S(x_1 \cdots x_n) = \sum_{1 \leq i \leq u-1} y_{2i-1}. \tag{2.7}$$

The probabilities of words of odd length may be found by using (2.4) and (2.7). The process given by (2.7) has the following interesting properties [see Robertson (1985)]:

(a) $P(X_n = 0) = P(X_n = 1) = \frac{1}{2}$.

(b) $\{X_1, X_2, \dots\}$ is ergodic and pairwise independent.

(c) $P[X_1 = X_2 = X_3 = 1] = (\frac{1}{4} + a)/4$ assumes all values in the interval $[\frac{1}{16}, \frac{3}{16}]$, which are the only possible values for stationary processes satisfying (a) and (b).

Before proceeding, we shall generalize the statistics R and S. Let k and l be fixed positive integers. Let $w = x_1 \cdots x_n$ be a word where $n - k + 1 = ml$ is a multiple of l. Set y_i equal to 0 or 1 according as there are an even or odd number of l's in the sequence $x_i \cdots x_{i+k-1}$. Next, for $1 \leq j \leq l$, set $S_j(x_1 \cdots x_n) = \sum_{0 \leq i \leq m-1} y_{j+il}$. We note that when $k = 3$ and $l = 2$, S_1 and S_2 are, respectively, the statistics S and R introduced in (2.7). When $m = 0$, $S_j(x_1 \cdots x_{k-1}) = 0$ by convention. We are interested in the statistics $S = (S_1, \dots, S_l)$. In this paper, we want to describe all processes that have these sufficient statistics. That is, we want to describe all $P \in \Xi$ such that P is constant on the sets $\{w: S(w) = s\}$. The processes just discussed with distribution given by (2.6) are clearly of this type. In particular, for $k > 1$, $P(x_1 \cdots x_{k-1}) = 2^{1-k}$. [For $k = 1$ this formula can be interpreted to be (2.2).] We have suppressed the dependence of the S on $|w|$. We will show that dependence by defining a new function

(2.8) $\quad p(m, s) = 2^{k-1}P(w), \quad$ where $|w| = k - 1 + ml$ and $S(w) = s$.

For a fixed m, the s_i may assume all choices of integer values between 0 and m. Thus p is defined on the set

$$G = \{(m, r): m \geq 0,\ r \in \{0, \dots, m\}^l\}.$$

This is an abelian semigroup generated by the subset $F = \{(1, r): r \in \{0, 1\}^l\}$ of 2^l elements. The identity is $0 = (0, \dots, 0)$. Given a function p defined on G, we can define P on Γ_n^* by (2.8) for all n such that $n - k + 1$ is a multiple of l. Provided that

$$P(w) = \sum_{w' \in \Gamma_l^*} P(ww') \tag{2.9}$$

whenever $|w| - k + 1$ is a multiple of l, then P can be defined on Γ_n^* for $(m-1)l < n - k + 1 < ml$ using (2.4) and the definition on Γ_{n+1}^*. Then (2.4) will be satisfied for *all n*.

LEMMA 2.10. *Let p be a function on G and let P be defined on Γ^* as previously described. Then P satisfies* (2.2)–(2.5) *if and only if p satisfies* (2.2′)–(2.5′):

$$p(0) = 1, \tag{2.2′}$$

$$p(g) \geq 0 \quad \textit{for all } g \in G, \tag{2.3′}$$

$$p(g) = \sum_{f \in F} p(g + f) \quad \textit{for all } g \in G, \tag{2.4′}$$

$$p(m, r_1, \dots, r_l) = p(m, r_l, r_1, \dots, r_{l-1}) \quad \textit{for all } m \geq 0 \textit{ and } r \in \{0, \dots, m\}^l. \tag{2.5′}$$

PROOF. (2.2) and (2.4) $\Rightarrow$ (2.2′) $p(0) = 2^{k-1}P(w)$ for all $w \in \Gamma_{k-1}^*$, where $P(w)$ is constant by assumption. Applying (2.4) $k - 1$ times yields $2^{k-1}P(\varepsilon) = 2^{k-1}p(0)$. Now (2.2) implies that $p(0) = 1$.

(2.3) $\Leftrightarrow$ (2.3′) Every $p(g)$ is 2^{k-1} times a $P(w)$ and conversely. Thus $P(w) \geq 0$ for all w if and only if $p(g) \geq 0$ for all g.

(2.4) $\Rightarrow$ (2.4′) Let $g = (m, r) \in G$. Find $w \in \Gamma_{ml-k+1}^*$ such that $S(w) = r$. Applying (2.4) l times, we obtain (2.9). As w' runs through Γ_l^*, $S(ww')$ runs through $r + \{0, 1\}^l$. Hence, multiplying (2.9) by 2^{k-1}, we obtain

$$p(m, r) = \sum_{e \in \{0,1\}^l} p(m + 1, r + e).$$

This is exactly (2.4′).

(2.4′) $\Rightarrow$ (2.4) If $|w| - k + 1$ is a multiple of l, the preceding calculations can be reversed to give (2.9). It thus follows from the remarks preceding this lemma that (2.4) is true.

(2.4′) and (2.2′) $\Rightarrow$ (2.2) Using (2.4) as before, $2^{k-1}P(\varepsilon) = 2^{k-1}p(0)$ and thus, using (2.2′), we obtain (2.2).

(2.4) and (2.5) $\Rightarrow$ (2.5′) Let $e \in \{0,1\}^l$ with $e_l = 0$ and $s = r + e \in \{0,\dots, m+1\}^l$. Find $w = x_1 \cdots x_n \in \Gamma^*_{(m+1)l+k-1}$ such that $S(w) = s$. Since $s_l \le m$, we can choose w so that $\Sigma_{0\le i\le k-1} x_{(m+1)l+i}$ is even. Let w' be identical to w except that the last symbol is reversed and let w'' be w (or w') with the last symbol removed. Then $s(w') = s + (0,\dots,0,1)$. (2.4) and (2.5) imply that

$$\begin{aligned} p(m+1, s) &+ p(m+1, s+(0,\dots,0,1)) \\ &= 2^{k-1}(P(w) + P(w')) \\ &= 2^{k-1}P(w'') = 2^{k-1}(P(0w'') + P(1w'')) \\ &= p(m+1, s_l, s_1,\dots, s_{l-1}) + p(m+1, s_l+1, s_1,\dots, s_{l-1}). \end{aligned}$$

Summing this equation over all e and using (2.4′) yields (2.5′).

(2.5′) and (2.4′) $\Rightarrow$ (2.5) If $w \in \Gamma^*$ and $|w| = k + ml - 2$, we write $r = S(x_1 \cdots x_{k+ml-1})$, where $x_1 \cdots x_{k+ml-1}$ is either $w0$ or $w1$ chosen such that $\Sigma_{0\le i\le k-1} x_{ml+i}$ is even. Then we have

$$\begin{aligned} P(0w) + P(1w) &= 2^{1-k}(p(m, r_l, r_1,\dots, r_{l-1}) + p(m, r_l+1, r_1,\dots, r_{l-1})) \\ &= 2^{1-k}(p(m, r_1,\dots, r_l) + p(m, r_1,\dots, r_{l-1}, r_l+1)) \\ &= P(w0) + P(w1) \\ &= P(w), \end{aligned}$$

where we have used (2.5′) and (2.4). Now suppose that $w \in \Gamma^*$ and $|w| = k + ml - 2 - j$. Then

$$\begin{aligned} P(0w) + P(1w) &= \sum_{w'\in\Gamma_j^*} P(0ww') + \sum_{w'\in\Gamma_j^*} P(1ww') \\ &= \sum_{w'\in\Gamma_j^*} [P(0ww') + P(1ww')] \\ &= \sum_{w'\in\Gamma_j^*} P(ww') \\ &= P(w), \end{aligned}$$

where we have used the first case and (2.4) repeatedly. □

In Section 3, we shall consider only properties (2.2′)–(2.4′). Then in Section 4, we shall add condition (2.5′). The following is immediate from the preceding proof and shows that Section 3 applies to all stochastic processes with the given sufficient statistics, while the remainder of the paper applies to only those processes that are also stationary.

COROLLARY 2.11. *Conditions* (2.2)–(2.4) *are equivalent to conditions* (2.2′)–(2.4′).

3. A representation theorem. The situation described in Lemma 2.10, (2.2′)–(2.4′) satisfies the conditions of Theorem 5 in Ressel (1985). (For the notation of the paper, take $S = G$ and define β: $S \to \mathbb{R}$ by $\beta(s) = 1$ if $s \in F$ and $\beta(s) = 0$ if $s \notin F$.) We will summarize the relevant conclusions of that theorem. Let Π be the set of functions p: $G \to \mathbb{R}$ satisfying (2.2′)–(2.4′). Π is a compact convex subset of $\mathbb{R}^G$ whose extreme points ext Π are given by the set of characters

$$\text{(3.1)} \qquad \operatorname{ext}\Pi = \{\rho \in \Pi: \rho(s+t) = \rho(s)\rho(t) \text{ for all } s, t \in G\}.$$

Moreover, Π is a simplex and for every $p \in \Pi$, there exists a unique probability measure μ_p on ext Π such that

$$\text{(3.2)} \qquad p(s) = \int_{\operatorname{ext}\Pi} \rho(s)\, d\mu_p(\rho) \quad \text{for all } s \in G.$$

We now identify the elements of ext Π.

Theorem 3.3. *We have* ext $\Pi = \{p_x\colon\ x \in [0,1]^l\}$ *with* $p_x(m, r) = \Pi_{1 \le j \le l} x_j^{m - r_j}(1 - x_j)^{r_j}$ *for all* $(m, r) \in G$. *The mapping* $x \to p_x$ *is a homeomorphism.*

Proof. Of course each P_x is multiplicative, nonnegative and

$$\sum_{f \in F} p_x(f) = \sum_{r \in \{0,1\}^l} p_x(1, r) = \sum_{r \in \{0,1\}^l} \prod_{1 \le j \le l} x_j^{1 - r_j}(1 - x_j)^{r_j}$$
$$= \prod_{1 \le j \le l} \big(x_j + (1 - x_j)\big) = 1.$$

Since $\Sigma_{(1, r) \in F, r_j = 0} p_x(1, r) = x_j$ for $j = 1, \ldots, l$, the character p_x is determined by x. The mapping $x \to p_x$, being obviously continuous, is therefore a homeomorphism onto its image which equals ext Π, as we shall now demonstrate: Let $p \in \operatorname{ext}\Pi$ and set $x_j = \Sigma_{(1, r) \in F, R_j = 0} p_x(1, r)$. For $r \in \{0,1\}^l$, we have

$$\begin{aligned}
p(1, r) &= p(1, r) \prod_{1 \le i < l} 1 \\
&= p(1, r) \prod_{1 \le i < l} \sum_{e_i \in \{0,1\}^l} p(1, e_i) \quad \text{[from (3.1)]} \\
&= \sum_{e_1 \in \{0,1\}^l} \cdots \sum_{e_{l-1} \in \{0,1\}^l} p\Big(1, r + \sum_{1 \le i < l} e_i\Big) \quad \text{[renaming the indices]} \\
&= \sum_{\substack{e_1 \in \{0,1\}^l \\ e_{11} = r_1}} \cdots \sum_{\substack{e_l \in \{0,1\}^l \\ e_{ll} = r_l}} p\Big(l, \sum_{1 \le i \le l} e_i\Big) \quad \text{[from (3.1) again]} \\
&= \sum_{\substack{e_1 \in \{0,1\}^l \\ e_{11} = r_1}} \cdots \sum_{\substack{e_l \in \{0,1\}^l \\ e_{ll} = r_l}} \prod_{1 \le i \le l} p(1, e_i) \\
&= \prod_{1 \le i \le l} \sum_{\substack{e_i \in \{0,1\}^l \\ e_{ii} = r_i}} p(1, e_i) \\
&= p_x(1, r).
\end{aligned}$$

Since now $p(s) = p_x(s)$ for all $s \in F$, and both p and p_x are characters [(3.1) is satisfied], and since F generates G, we have $p = p_x$. □

4. The stationary points. Theorem 3.3 and eq. (3.2) characterize the convex set Π of all functions p satisfying (2.2′)–(2.4′) and their extreme points. We are, however, interested in the smaller convex set of all p that also satisfy (2.5′). We shall denote this smaller set by Π^*. For this purpose we introduce the transformation σ: $[0,1]^l \to [0,1]^l$ defined by $\sigma(x_1, \dots, x_l) = (x_2, \dots, x_l, x_1)$.

THEOREM 4.1. *Let $p \in \Pi$ correspond to the probability measure μ. Then $p \in \Pi^*$ if and only if μ is invariant under σ. Π^* is a simplex, and p is an extreme point of Π^* if and only if there exists $x \in [0,1]^l$ such that p has the form*

$$p(m,r) = (1/l) \sum_{1 \le j \le l} \prod_{1 \le i \le l} x_{i+j}^{m-r_i}(1 - x_{i+j})^{r_i} \tag{4.2}$$

for all $m \ge 0$ and $r \in \{0, \dots, m\}^l$, where we set $x_{i+l} = x_i$ for all i.

PROOF. Let $p \in \Pi^*$. Then by (2.5′) and Theorem 3.3 we have

$$\begin{aligned}
&\int_{[0,1]^l} \prod_{1 \le i \le l} y_i^{m-r_i}(1 - y_i)^{r_i}\, d\mu(y) \\
&\quad = p(m, r_1, \dots, r_l) \\
&\quad = p(m, r_l, r_1, \dots, r_{l-1}) \\
&\quad = \int_{[0,1]^l} y_1^{m-r_l}(1 - y_1)^{r_l} \prod_{2 \le i \le l} y_i^{m-r_{i-1}}(1 - y_i)^{r_{i-1}}\, d\mu(y) \\
&\quad = \int_{[0,1]^l} \prod_{1 \le i \le l} y_i^{m-r_i}(1 - y_i)^{r_i}\, d\mu(\sigma^{-1}(y)).
\end{aligned}$$

Thus, using the Stone–Weierstrass theorem, we see that $\mu(\sigma(A)) = \mu(A)$ for all Borel sets A and hence μ is invariant under σ. Conversely if μ is invariant under σ, then the first and last expressions in the preceding equations are equal and hence (2.5′) holds as desired. Next let $p \in \Pi^*$ and let μ be the measure, invariant under σ, given by Theorem 3.3. Since the orbits of σ form a partition of $[0,1]^l$, it follows that p is an extreme point of Π^* if and only if μ is concentrated on a single orbit of σ and μ assigns equal measure to each of the points of this orbit. Hence $p(m, r)$ is given by the formula (4.2). □

5. Finitary processes. We shall now turn our attention to the study of the properties of the points in ext Π^*. For example, the set of all P corresponding to $p \in \Pi^*$ is a subset of the convex set Ξ of all P satisfying (2.2)–(2.5). A major question we will answer is: When is a point in ext Π^* also in ext Ξ? That is, when is the process corresponding to a point in ext Π^* ergodic? To answer this type of question we shall use the concept of *finitary processes*. In this section we shall review these notions and some of the theorems in Robertson (1973).

We are given a real Hilbert space $(H, \langle \cdot, \cdot \rangle)$, two vectors ξ and η and, for all $g \in \Gamma$, bounded linear operators A_g on H. If $w = g_1 \cdots g_n \in \Gamma^*$, then we set $A_w = A_{g_1} \cdots A_{g_n}$. A_ε will, by convention, denote the identity operator. If M is a closed subspace of H, we let π_M denote the orthogonal projection of H onto M and ρ_M denote the injection map from M into H. If W is a subset of H, $\sigma(W)$ will denote the smallest closed subspace containing W. We suppose that the following conditions are satisfied:

$$\langle \eta, \xi \rangle = 1, \tag{5.1}$$

$$\text{for all } w \in \Gamma^*, \quad \langle \eta, A_W \xi \rangle \geq 0, \tag{5.2}$$

$$\sum_{g \in \Gamma} A_g \xi = \xi \quad \text{and} \quad \sum_{g \in \Gamma} A_g^* \eta = \eta. \tag{5.3}$$

We define a function P on Γ^* by

$$P(w) = \langle \eta, A_w \xi \rangle. \tag{5.4}$$

We simply call $H, \eta, \xi, \{A_g \colon g \in \Gamma\}$ a *system* if the preceding conditions are satisified. We say that a system is *reduced* if

$$\sigma\{A_w \xi \colon w \in \Gamma^*\} = \sigma\{A_w^* \eta \colon w \in \Gamma^*\} = H.$$

The function P defined by (5.4) satisfies conditions (2.2)–(2.5) and thus defines the distribution of a stationary stochastic process; cf. (2.1). Such processes are called *finitary* if we can find a system for which $\dim H < \infty$. For the definitions and properties of ergodicity and weak mixing, see, for example, Halmos (1956). The following theorem can essentially be found in Robertson (1973) and its proof will not be given here.

THEOREM 5.5. *Let* $H, \eta, \xi, \{A_g \colon g \in \Gamma\}$ *be a system such that* $\dim H < \infty$ *and* $\{X_n,\ n = 0, 1, \dots\}$ *a process it determines.*

(a) *If the process is ergodic and the system is reduced, then* 1 *is a simple eigenvalue of* $\Sigma_{g \in \Gamma} A_g$.

(b) *If* 1 *is a simple eigenvalue of* $\Sigma_{g \in \Gamma} A_g$, *then the process is ergodic.*

(c) *If the process is weakly mixing and the system is reduced, then* 1 *is a simple eigenvalue of* $\Sigma_{g \in \Gamma} A_g$ *and the only one of absolute value* 1.

(d) *If* 1 *is a simple eigenvalue and the only eigenvalue of absolute value* 1 *of* $\Sigma_{g \in \Gamma} A_g$, *then the process is weakly mixing.*

6. The representation. In this section, we shall describe a family of finitary systems and then show that these systems represent the extremal processes given in Theorem 4.1. Let k and l be natural numbers and let $p = (p_1, \dots, p_l) \in [0,1]^l$. Define $a(0)$ to be the $l \times l$ matrix given by $a(0)_{i,j} = p_i$ if $j = i + 1$ and 0 otherwise. (Here and in what follows, when dealing with indices that vary from 1 to l, addition will be interpreted modulo l.) $a(1)$ is defined in the same way as $a(0)$ except that p_i is replaced by $\bar{p}_i = 1 - p_i$. Thus $\sigma = a(0) + a(1)$ is the permutation matrix that sends $(x_1, \dots, x_l)$ to $(x_2, \dots, x_l, x_1)$. For $k = 1$, the system consists of $\mathbb{R}^l, \eta, \xi, \{a(0), a(1)\}$, where ξ and η are the vectors all of whose components are equal to $l^{-1/2}$. For $k > 1$, the system consists of $\mathbb{R}^{l2^{k-1}}, \eta, \xi, \{A(0), A(1)\}$, where $A(0)$, $A(1)$, ξ and η are as defined in the following text. $A(0)$ will be a $l \cdot 2^{k-1}$ dimensional square matrix. It will be partitioned into a 2^{k-1} dimensional square matrix $\tilde{A}$, whose entries will be l dimensional square matrices. The rows and columns of $\tilde{A}$ will be indexed by vectors $v = (v_1, \dots, v_{k-1}) \in \{0,1\}^{k-1}$. If $e \in \{0,1\}$, let

$$A(e)_{v,w} = \begin{cases} a(0), & \text{if } v = (e, w_1, \dots, w_{k-2}) \text{ and } v_1 + \cdots + v_{k-1} + w_{k-1} \text{ is even,} \\ a(1), & \text{if } v = (e, w_1, \dots, w_{k-2}) \text{ and } v_1 + \cdots + v_{k-1} + w_{k-1} \text{ is odd,} \\ 0, & \text{if } v \neq (e, w_1, \dots, w_{k-2}). \end{cases}$$

It is then easy to see that $\Sigma = A(0) + A(1)$ is a doubly stochastic matrix. Therefore, we may take η and ξ to be the $l \cdot 2^{k-1}$ dimensional vectors, all of whose components are equal to $(l \cdot 2^{k-1})^{-1/2}$. We think of η as a row vector and ξ as a column vector.

EXAMPLE. $k = 3$ and $l = 2$. Then, setting $a = p_1$ and $b = p_2$, we have

$$\tilde{A}(0) = \begin{bmatrix} a(0) & a(1) & 0 & 0 \\ 0 & 0 & a(1) & a(0) \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

$$\bar{A}(1) = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ a(1) & a(0) & 0 & 0 \\ 0 & 0 & a(0) & a(1) \end{bmatrix},$$

where

$$a(0) = \begin{bmatrix} 0 & a \\ b & 0 \end{bmatrix} \quad \text{and} \quad a(1) = \begin{bmatrix} 0 & 1-a \\ 1-b & 0 \end{bmatrix}.$$

THEOREM 6.1. *The probabilities given by* (5.4), *with the* η, ξ, $A(0)$ *and* $A(1)$ *previously given, are the same as the probabilities given by* (2.8) *and* (4.2). *More generally,* (5.4) *is true for words* $w \in \Gamma^*$ *of all lengths.*

PROOF OF THE $k = 1$ CASE. First suppose that $n = ml$ and $w = g_1 g_2 \cdots g_n \in \Gamma_n^*$. Then set $A_w = a(g_1) \cdots a(g_n)$. A_w will then have nonzero entries only where the permutation matrix σ^n has nonzero entries. The nonzero entry in the ith row will be equal to $y_i y_{i+1} \cdots y_{i+n-1}$, where $y_j = p_j$ or $1 - p_j$ according as $g_{j+1-i} = 0$ or 1. We note that, since $k = 1$, s_i in (2.8) is just the number of j's such that $j \equiv i \pmod{l}$ and $w_j = 1$. We also note that for η and ξ as given, $\eta B \xi$ is just equal to the sum of all the entries of B divided by l. These observations show that $\langle \eta, A_w \xi \rangle = \eta A_w \xi$ is given by the formula in Theorem 4.1. Let $1 < t < l$, $n = ml - t$ and $w = g_1 \cdots g_n \in \Gamma_n^*$. Then from (2.4),

$$P(w) = \sum_{w' \in \Gamma_t^*} P(ww').$$

Since $|ww'| = ml$, from what we have already proved,

$$\begin{aligned} P(w) &= \eta A(g_1) \cdots A(g_n)(A(0) + A(1))^t \xi \\ &= \eta A(g_1) \cdots A(g_n) \sigma^t \xi = \eta A(g_1) \cdots A(g_n) \xi. \end{aligned}$$

This completes the proof for the case $k = 1$. □

PROOF OF THE $k > 1$ CASE. Let $w = g_1 \cdots g_n \in \Gamma_n^*$, where $n = ml + k - 1$. Then the (v, u) entry of $\tilde{A}_w = \tilde{A}(g_1) \cdots \tilde{A}(g_n)$ is nonzero only if there exist $v^0 = v, v^1, v^2, \ldots, v^n = u$ such that

$$\text{for all } i = 1, \ldots, n, \qquad v^{i-1} = \left(g_i, v_1^i, \ldots, v_{k-2}^i\right).$$

This implies that $(g_1, \ldots, g_{k-1}) = (v_1, \ldots, v_{k-1})$. In other words, the only row of $\tilde{A}_w$ with any nonzero elements is the one indexed by $v = (g_1, \ldots, v_{k-1})$. For each u, there exists a unique chain $v^1, \ldots, v^n$ as before, namely,

$$v^i = \left(g_{i+1}, g_{i+2}, \ldots, g_{i+k-1}\right),$$

where we define $g_{n+i} = u_i$ for $1 \le i \le k - 1$. Hence the (v, u) entry of $\tilde{A}_w$ is $a(y_i) \cdots a(y_n)$, where y_i is 0 or 1 according as there are an even or an odd number of 1's in the sequence $g_i, \ldots, g_{i+k-1}$. Thus the sum of the elements of $\tilde{A}_w$ is the sum of the elements of $a(y_1) \cdots a(y_{n-k+1})(a(0) + a(1))^{k-1}$. Since $(a(0) + a(1))^{k-1}$ is a permutation matrix, this is identical with the sum of the elements of $a(y_1) \cdots a(y_{n-k+1})$. Now $y_1, \ldots, y_{n-k+1}$ are exactly as defined in Section 2. The rest of the details are similar to the $k = 1$ case. □

7. The pairwise independence of the process. Our original motivation was to give examples of stationary processes that were pairwise independent without being mutually independent. In this section we will describe which of the processes that we are now studying are pairwise independent.

THEOREM 7.1. *We consider the process discussed in Section* 6.

(a) *If* $k = 1$, *the process is pairwise independent if and only if all the* p_i *are equal. Thus, in this case, if the process is pairwise independent, it is also mutually independent.*

(b) *If* $k = 2$, *the process is pairwise independent if and only if for all* r $(1 \le r \le l)$,

$$\sum_{i=1}^{l} \left[\prod_{j=1}^{r} \left(p_{i+j} - \tfrac{1}{2}\right) \right] = 0.$$

(c) *If* $k > 2$, *the process is pairwise independent if and only if for all* $r \ge 1$,

$$\sum_{i=1}^{l} \left[\prod_{j=1}^{r-1} \left(p_{i+jk} - \tfrac{1}{2}\right)\left(p_{i+jk+1} - \tfrac{1}{2}\right) \right] = 0.$$

PROOF. The process is pairwise independent if and only if

$$\text{for all } n \ge 0, \qquad \eta A(0) \Sigma^n A(0) \xi = \left(\eta A(0) \xi\right)^2. \tag{7.2}$$

(a) Here $A(0) = a(0)$ and $\Sigma = \sigma$. Taking $n = l - 1$ in (7.2), we obtain

$$\frac{1}{l} \sum_{i=1}^{l} p_i^2 = \left(\frac{1}{l} \sum_{i=1}^{l} p_i \right)^2$$

and the result follows.

For (b) and (c) we define $\tilde{\eta}$ and $\tilde{\xi}$ by $\tilde{\eta}_v = \tilde{\xi}_v = (l \cdot 2^{k-1})^{-1/2}$ if $v_1 = 0$ and $\tilde{\eta}_v = \tilde{\xi}_v = 0$, otherwise. Then $\eta A(0) = \tilde{\eta}\Sigma$, $A(0)\xi = \tilde{\xi}$ and $\eta A(0)\xi = \frac{1}{2}$. Thus (7.2) holds if and only if

$$\text{for all } n \geq 1, \qquad \tilde{\eta}\Sigma^n\tilde{\xi} = \tfrac{1}{4}. \tag{7.3}$$

Let b be the $l \times l$ matrix defined by $b_{ij} = 2p_i - 1$ if $j = i + 1$ and $b_{ij} = 0$ otherwise. If x is *any* column vector of height l, then the following computational rules (7.4)–(7.7) are valid. In these rules α and β are column vectors of height $l \cdot 2^{k-1}$ that (as before) are partitioned into column vectors α_v and β_v of height l, where v runs through the set $\{0,1\}^{k-1}$.

(7.4) If $\alpha_v = x$ for all v, then $(\Sigma\alpha)_v = \sigma x$ for all v.

(7.5) Let $1 \leq r \leq k - 2$ and suppose that $\beta_v = x$ whenever $v \in \{0,1\}^{k-1}$ and $v_r = 0$, and $\beta_v = -x$, otherwise. This completely defines the vector β. Then $(\Sigma\beta)_v = \sigma x$ if $v_{r+1} = 0$ and $(\Sigma\beta)_v = -\sigma x$, otherwise.

(7.6) If $\beta_v = x$ whenever $v \in \{0,1\}^{k-1}$ and $v_{k-1} = 0$, and $\beta_v = -x$, otherwise, then $(\Sigma\beta)_v = bx$ if $v_1 + \cdots + v_{k-1}$ is even and $(\Sigma\beta)_v = -bx$, otherwise.

(7.7) If $\beta_v = x$ whenever $v \in \{0,1\}^{k-1}$ and $v_1 + \cdots + v_{k-1}$ is even, and $\beta_v = -x$, otherwise, then $(\Sigma\beta)_v = bx$ whenever $v_1 = 0$, and $(\Sigma\beta)_v = -bx$, otherwise.

Now let x be the column vector of height l with all components having the value $(l \cdot 2^{k-1})^{-1/2}/2$. Then $\tilde{\xi} = \alpha + \beta$, where α is as in (7.4) and β is as in (7.5) with $r = 1$. Since $\sigma x = x$, from (7.4),

$$\Sigma^n\tilde{\xi} = \Sigma^n\alpha + \Sigma^n\beta = \alpha + \Sigma^n\beta.$$

Thus

$$\tilde{\eta}\Sigma^n\tilde{\xi} = \tilde{\eta}\alpha + \tilde{\eta}\Sigma^n\beta = \tfrac{1}{4} + \tilde{\eta}\Sigma^n\beta.$$

Hence (7.3) holds if and only if

$$\text{for all } n \geq 1, \qquad \tilde{\eta}\Sigma^n\beta = 0. \tag{7.8}$$

(b) From (7.6) or (7.7) (which coincide in this case), (7.8) holds if and only if

$$\text{for all } n \geq 1, \qquad x^T b^n x = 0.$$

This gives the desired result.

(c) From (7.5)–(7.7) and the symmetries of the problem, we only have to verify (7.8) when n is a multiple of k. Thus, (7.8) holds if and only if

$$\text{for all } r \geq 1, \qquad x^T(b^2\sigma^{k-2})^r x = 0.$$

This gives the desired result. □

EXAMPLE. In our previous example, where $k = 3$ and $l = 2$, the result $r = 1$ condition becomes

$$\left(a - \tfrac{1}{2}\right)\left(b - \tfrac{1}{2}\right) + \left(b - \tfrac{1}{2}\right)\left(a - \tfrac{1}{2}\right) = 0.$$

It thus follows that for $k = 3$, the only pairwise independent processes correspond to $a = \frac{1}{2}$ (or by symmetry, $b = \frac{1}{2}$). These are just the processes treated in Robertson (1985).

8. Other properties of the process. In this section we shall investigate the ergodic and mixing properties of these processes. We shall not give a complete analysis, but do enough to indicate the general nature of these processes.

We shall first treat the case $k = 1$. $a(0) + a(1)$ is a permutation matrix independent of $(p_1, \ldots, p_l)$, and has eigenvalues equal to $e^{2\pi i j/l}$ for $j = 1, \ldots, l$. By Theorem 5.5(c), all of these processes are therefore ergodic. By Theorem 5.5(d), these processes will not be weakly mixing if the system is reduced. Whether or not the system is reduced depends on $(p_1, \ldots, p_l)$. If, for example, all of the p_i are equal, then the system can be reduced to a one dimensional space and the process is a Bernoulli process. On the other hand, if the p_i are all distinct, then the system is reduced and the system is not weakly mixing.

Next suppose that $k > 1$. In this case σ is a doubly stochastic matrix with period at least l. (The states that are l states apart form subclasses.) It may not be an irreducible stochastic matrix. In the preceding example, if we take $p_1 = p_2 = 0$, then σ is the permutation $(1,4,5,2,3,6)(7,8)$. In this case the probability space can also be taken to be finite. However, if none of the p_i is 0 or 1, then σ will be irreducible and hence the process will be ergodic. It also follows that the system will be reduced if all of the p_i are distinct. Thus the processes are "usually," but not always ergodic, but they are not "usually" weakly mixing.

REFERENCES

DIACONIS, P. and FREEDMAN, D. (1980). de Finetti's theorem for Markov chains. *Ann. Probab.* **8** 115–130.

HALMOS, P. (1956). *Lectures on Ergodic Theory*. Chelsea, New York.

RESSEL, P. (1985). de Finetti-type theorems: An analytical approach. *Ann. Probab.* **13** 898–922.

ROBERTSON, J. (1973). A spectral representation of the states of a measure preserving transformation. *Z. Wahrsch. verw. Gebiete* **27** 185–194.

ROBERTSON, J. (1985). Independence of fair coin tossing. *Math. Sci.* **10** 109–117.

ROBERTSON, J. and WOMACK, J. (1985). A pairwise independent stationary stochastic process. *Statist. Probab. Lett.* **3** 195–199.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF CALIFORNIA
SANTA BARBARA, CALIFORNIA 93106