The Annals of Probability
1980, Vol. 8, No. 1, 115–130

DE FINETTI'S THEOREM FOR MARKOV CHAINS

BY P. DIACONIS[1] AND D. FREEDMAN[2]

Bell Laboratories, Murray Hill, New Jersey and University of California at Berkeley

Let $Z = (Z_0, Z_1, \cdots)$ be a sequence of random variables taking values in a countable state space I. We use a generalization of exchangeability called partial exchangeability. Z is partially exchangeable if for two sequences $\sigma, \tau \in I^{n+1}$ which have the same starting state and the same transition counts, $P(Z_0 = \sigma_0, Z_1 = \sigma_1, \cdots, Z_n = \sigma_n) = P(Z_0 = \tau_0, Z_1 = \tau_1, \cdots, Z_n = \tau_n)$. The main result is that for recurrent processes, Z is a mixture of Markov chains if and only if Z is partially exchangeable.

1. Introduction. Let $Z = (Z_0, Z_1, Z_2, \cdots)$ be a sequence of random variables taking values zero or one. In the classical Bayesian approach to statistics, if Z_i are independent and identically distributed with unknown success probability p, it is customary to choose a prior distribution μ on the Borel sets of the unit interval and describe the joint distribution of the process as follows: for any sequence of zeros and ones $e_0, e_1, \cdots, e_n$, letting

$$S = \Sigma_{i=0}^{n} e_i,$$

$$P(Z_0 = e_0, \cdots, Z_n = e_n) = \int_0^1 p^S (1-p)^{n+1-S} \mu(dp). \tag{1}$$

In developing his subjective theory of probability, de Finetti introduced the notion of exchangeability: P is exchangeable if

$$P(Z_0 = e_0, \cdots, Z_n = e_n) = P(Z_{\pi(0)} = e_0, \cdots, Z_{\pi(n)} = e_n), \tag{2}$$

where π is any permutation of $\{0, 1, 2, \cdots, n\}$.

A basic result due to de Finetti is that a probability measure P on infinite sequences Z satisfies (2) for every $n > 0$ if and only if P satisfies (1) for some μ. This result is developed further in Hewitt and Savage (1955); for a recent survey, see Kingman (1978).

Suppose now that Z is a Markov chain with unknown stationary transition matrix

$$T = \begin{pmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{pmatrix}.$$

Received June 1978.

[1] Research partially supported by National Science Foundation Grant MPS 74-21416 and by the Energy Research and Development Administration under Contract EY-76-C-03-0515.

[2] Research partially supported by National Science Foundation Grant MCS 575-09459.

AMS 1970 *subject classifications.* Primary 60J05; secondary 62A15.

Key words and phrases. Mixture of Markov chains, de Finetti's theorem, extreme point representations, zero-one laws.

For simplicity, assume in this introduction that $Z_0 = 0$. The classical Bayesian description of this process (see for example, Martin (1967)) involves a prior distribution ν on the Borel sets of the 2×2 stochastic transition matrices

$$\begin{pmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{pmatrix}.$$

Writing a_{ij} for the number of transitions from state i to state j in the sequence $0, e_1, e_2, \cdots, e_n$, this yields

$$(3)\quad P\{Z_0 = 0, Z_1 = e_1, \cdots, Z_n = e_n\} = \int p_{00}^{a_{00}} p_{01}^{a_{01}} p_{10}^{a_{10}} p_{11}^{a_{11}}\, \nu(dp_{00}, dp_{01}, dp_{10}, dp_{11}).$$

The main result of this paper gives an appropriate notion of symmetry, a generalization of exchangeability called partial exchangeability, so that (3) holds for some ν if and only if P is partially exchangeable, provided P is recurrent. For transient P, this characterization fails as shown by example (19) below. Briefly, P is partially exchangeable if P assigns probability to all strings $0, e_1, e_2, \cdots, e_n$ with the same transition count matrix

$$\begin{pmatrix} a_{00} & a_{01} \\ a_{10} & a_{11} \end{pmatrix}.$$

This result is hinted at by de Finetti (1959), (1974, pages 217–220). A proof when Z is stationary is in Freedman (1962). For a discussion of the result see Diaconis and Freedman (1978a, c).

de Finetti (1959, 1974) sketches a general theorem for finite partially exchangeable processes $Z_0, Z_1, \cdots, Z_n$. He appears to assert that the finite results imply results for infinite sequences by passage to the limit. While this argument works for exchangeable sequences (Diaconis (1977), Diaconis and Freedman (1978b)), it breaks down for transient Markov chains, as shown in Example (19) here.

In Section 2 of this paper we show that for a recurrent process taking values in a countable state space (3) holds if and only if the process is partially exchangeable. The extreme points of partially exchangeable processes with finite state space are found in Section 3. Section 4 contains remarks and complements to the main results.

2. Recurrent chains. Let $X = (X_0, X_1, \cdots)$ be a stochastic process on the probability triple $(\Omega, \mathcal{F}, P)$, taking values in the countable set I, whose elements will be referred to as "states". de Finetti's theorem gives a necessary and sufficient condition for the distribution of X to be a mixture of the distributions of sequences of independent and identically distributed random variables. The condition is that X be exchangeable. To rephrase the definition, let σ be a finite string of states. Let A_σ be the event that X starts by running through σ. Thus,

$$A_{ijk} = \{X_0 = i \quad \text{and } X_1 = j \quad \text{and } X_2 = k\}.$$

Clearly, X is exchangeable iff $P(A_\sigma) = P(A_\tau)$ whenever τ is a permutation of σ.

Under what circumstances is the distribution of X a mixture of distributions of Markov chains? Our object is to answer this question. Throughout this paper, as usual, a Markov chain has stationary transitions by definition, but need not start from a stationary distribution. In this section we assume that X is recurrent:

(4) $$P\{X_n = X_0 \text{ for infinitely many } n\} = 1.$$

The necessary and sufficient condition that X be a mixture of Markov chains is in terms of a certain kind of symmetry. To state it, let σ and τ be finite strings of states (a "string" is a finite sequence). Then $\sigma \sim \tau$ iff σ and τ start from the same state and exhibit the same number of transitions from i to j for every pair of states i and j. For instance, $123 \nsim 132$, but $12132 \sim 13212$. The condition, then, is that $\sigma \sim \tau$ implies $P(A_\sigma) = P(A_\tau)$. Such probabilities are called partially exchangeable.

The next lemma records some simple properties of the equivalence $\sim$.

(5) LEMMA. *Suppose $\sigma \sim \tau$. Then σ and τ have the same length and end at the same state. Furthermore, σ and τ pass through i the same number of times, for any state i.*

A characterization of $\sim$ in terms of invariance under a group of transformations is given in Proposition (27). Lemma (5) is close to Lemma 6.1.1 of Martin (1967). See also Freedman (1962b).

To say that X is a mixture of Markov chains means the following. Consider the set of stochastic matrices $\mathcal{P}$ on $I \times I$ in the topology of coordinate convergence. The set $I \times \mathcal{P}$ is Polish. There is to be a probability μ on the Borel subsets of this space such that

(6) $$P\{X_m = i_m \text{ for } 0 \leqslant m \leqslant n\} = \int_{\mathcal{P}} \prod_{m=0}^{n-1} p(i_m, i_{m+1})\mu(i_0, dp).$$

In these terms, the first theorem can be stated as follows.

(7) THEOREM. *Suppose X is recurrent in the sense* (4). *Then X is a mixture of Markov chains in the sense* (6) *if and only if*

(8) $$\sigma \sim \tau \text{ implies } P(A_\sigma) = P(A_\tau).$$

Moreover, the mixing measure μ is uniquely determined.

The necessity of (8) is almost obvious, and the sufficiency will be proved here. Without real loss of generality, suppose X_0 is degenerate, say at the state 1:

(9) $$P(X_0 = 1) = 1.$$

Consider I with the discrete topology and let I^∞ be the space of infinite I-sequences, endowed with the product topology; so I^∞ is Polish. To avoid measure-theoretic difficulties, and without loss of generality, suppose that X_n is the coordinate process on I^∞:

$$X_n(\omega) = \omega_n.$$

By definition, a 1-block is a finite string of states which begins with 1 and contains no further 1's. The space B of 1-blocks is countable and is given the discrete topology. Let $Y_0, Y_1, \cdots$ be the successive 1-blocks in X. Thus, if X starts off $1213122114 \cdots$ then $Y_0 = 12$, $Y_1 = 13$, $Y_2 = 122$, and $Y_3 = 1$. The recurrence condition (4) implies that the Y's are almost surely well defined. It is clear they are measurable. Condition (8) implies

(10) The 1-blocks of X are exchangeable.

Indeed, permuting 1-blocks cannot affect transitions.

Let $\mathcal{F}^{(n)}$ be the σ-field spanned by $Y_n, Y_{n+1}, \cdots$ and let $\mathcal{F}^{(\infty)} = \bigcap_n \mathcal{F}^{(n)}$, the tail σ-field of the 1-blocks.

De Finetti's theorem for Polish space random variables says that given $\mathcal{F}^{(\infty)}$ the 1-blocks are independent and identically distributed. More precisely, there is a regular conditional probability $P_\omega(A)$ on the Borel subsets of I^∞ given $\mathcal{F}^{(\infty)}$.

(11) LEMMA. *For P-almost all ω, with respect to P_ω the 1-blocks $Y_0, Y_1, \cdots$ are independent and identically distributed.*

A proof of Lemma (11) is given in the Appendix. See also Olshen (1974).

The next thing to establish is that given $\mathcal{F}^{(\infty)}$, the process X still has the symmetry property (8).

(12) LEMMA. *For P-almost all ω, the relation $\sigma \sim \tau$ implies $P_\omega(A_\sigma) = P_\omega(A_\tau)$.*

PROOF. For the usual reasons, it is enough to show that almost surely,

$$P\{A_\sigma | Y_n, \cdots, Y_{n+m}\} = P\{A_\tau | Y_n, \cdots, Y_{n+m}\} \tag{13}$$

for large n and all m. This would follow from

$$\begin{aligned} P\{A_\sigma \quad \text{and} \quad Y_{n+v} = \beta_v \quad \text{for} \quad 0 \leqslant v \leqslant m\} \\ = P\{A_\tau \quad \text{and} \quad Y_{n+v} = \beta_v \quad \text{for} \quad 0 \leqslant v \leqslant m\}, \end{aligned} \tag{14}$$

valid for all m and all 1-blocks β_v, provided n is at least as large as the number of 1's in σ (or τ). To prove (14), let S_σ be the set of strings ψ such that the $n + v$th 1-block in $\sigma\psi$ is β_v for $0 \leqslant v \leqslant m$. As (5) shows, $S_\sigma = S_\tau$, and $\sigma\psi \sim \tau\psi$, so $P(A_{\sigma\psi}) = P(A_{\tau\psi})$. Summing out ψ gives (14). □

The next result is a partial converse to Doeblin's theorem (Theorem 31, page 15 of Freedman (1971)).

(15) PROPOSITION. *Suppose X satisfies the recurrence condition* (4) *and the symmetry condition* (8). *Suppose too that the 1-blocks of X are independent and identically distributed. Then X is a Markov chain.*

PROOF. Let σ and σ' be finite strings of states which start at 1 and end at i; however, do not assume that $\sigma \sim \sigma'$. The Markov property, which must be proved, is

$$P(A_{\sigma j} | A_\sigma) = P(A_{\sigma' j} | A_{\sigma'}).$$

To avoid division by 0, the form

$$(16) \qquad P(A_\sigma)P(A_{\sigma' j}) = P(A_{\sigma'})P(A_{\sigma j})$$

is preferred.

For any strings of states α and β

$$(17) \qquad P(A_{1\alpha 1\beta}) = P(A_{1\alpha 1})P(A_{1\beta}),$$

because the 1-blocks are independent and identically distributed.

Let ψ run through all the finite strings of states which do not pass through 1. Then recurrence (4) implies

$$(18) \qquad P(A_\sigma) = \Sigma_\psi P(A_{\sigma\psi 1}).$$

Because σ and σ' start and end at 1 and i,

$$\sigma\psi\sigma' j \sim \sigma'\psi\sigma j.$$

By symmetry (8),

$$P(A_{\sigma\psi\sigma' j}) = P(A_{\sigma'\psi\sigma j}).$$

By (17),

$$P(A_{\sigma\psi 1})P(A_{\sigma' j}) = P(A_{\sigma\psi\sigma' j}) = P(A_{\sigma'\psi\sigma j}) = P(A_{\sigma'\psi 1})P(A_{\sigma j}).$$

Sum out ψ and use (18) to get (16). □

The sufficiency part of Theorem (7) is now easily proved, starting from the identity

$$P\{A\} = \int P_\omega\{A\}P(d\omega),$$

where P_ω was defined for Lemma (11). In view of (11) and (12), it is enough to integrate only over those ω's for which, relative to P_ω, the process X has the recurrence property (4), the symmetry property (8), and independent, identically distributed 1-blocks. For such ω, however, X is a Markov chain relative to P_ω by (15). Letting $\tilde{P}_\omega(i, j)$ denote the corresponding transition probabilities, verify that $\omega \to \tilde{P}_\omega(i, j)$ is measurable and

$$P\{X_m = i_m \text{ for } 0 \leqslant m \leqslant n\} = \int \Pi_{m=0}^{n-1} \tilde{P}_\omega(i_m, i_{m+1})P(d\omega).$$

This is (6) when $X_0 = 1$ is given. The general case follows by conditioning on X_0.

The uniqueness of μ asserted in (6) follows from the strong law of large numbers for Markov chains, as described in Remark (25) of Section 4. Another proof of uniqueness is given by Freedman (1962a). This completes the proof of (7).

The argument proves something a bit sharper than (7): namely, under conditions (4) and (8), given the tail σ-field of the 1-blocks, the X process is conditionally Markov.

3. The general case; finite state space. Without the recurrence condition (4) it is not true that partially exchangeable processes are mixtures of Markov chains. In this section we characterize partially exchangeable processes when the state space I is finite, without assuming recurrence.

When I is finite, the space of all probability measures on the Borel sets of I^{∞} is a compact metrizable space in the weak * topology. The space $\mathfrak{M}$ of all probability measures which are partially exchangeable (that is, satisfy the symmetry condition (8)) is a closed convex subset of the probability measures. We will find the extreme points $\mathcal{E}$ of $\mathfrak{M}$ and show that any element of $\mathfrak{M}$ is a unique mixture of elements of $\mathcal{E}$. For motivation, we first describe the situation when $I = \{0, 1\}$. All assertions will be proved in Theorem (21).

(19) EXAMPLES. Suppose $I = \{0, 1\}$. The extreme points of the partially exchangeable processes break into four groups:

(a) The recurrent Markov chains starting at 0,

(b) The recurrent Markov chains starting at 1,

(c) The processes π_k^0 which run deterministically through a sequence of k zeros, make a single transition at time $k + 1$ to state one, and end with all ones.

(d) The processes π_k^1 which start with k ones, make a single transition to zero at time $k + 1$, and end with all zeros.

Consider π_2^0. This is the law of a process $X = (X_0, X_1, \cdots)$ with the single sample path $0011111 \cdots$ (two zeros followed by all ones). This law is partially exchangeable but not a mixture of Markov chains.

Indeed, X is partially exchangeable because there are no other sequences except $00111111 \cdots$ which have the same transition counts. To see that X cannot be a mixture of Markov chains, consider a process Y which is a mixture of Markov chains. Suppose Y starts at zero. Let $\theta_k(Y)$ be the probability that Y starts with k zeros and has ones in all other positions. From (6),

$$\theta_k(Y) = \int_{\mathcal{P}_0} p_{00}^{k-1}(1 - p_{00})\mu(0, dp_{00})$$

where $\mathcal{P}_0$ is the set of stochastic matrices with $p_{11} = 1$. In particular θ_k decreases with k. For the process X, $\theta_1(X) = 0$, $\theta_2(X) = 1$, $\theta_3(X) = 0$. This shows that X cannot be represented as a mixture of Markov chains.

Since π_2^0 is an extreme point of $\mathfrak{M}$, not all extreme points are Markov chains. Since Markov chains starting at zero and having zero as a transient state can be represented as mixtures of the π_k^0, not all Markov chains are extreme points of $\mathfrak{M}$. Note that nonrecurrent Markov chains are limits of recurrent Markov chains, for example, by taking limits of transition matrices

$$\begin{pmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{pmatrix} \text{ as } p_{00} \to 1.$$

So the set $\mathcal{E}$ of extreme points is not closed. It can be argued that $\mathcal{E}$ is the intersection of a closed set and an open set. Further discussion of this example can be found in Section 4.

We now return to the general, finite alphabet situation and introduce notation for the extreme points. Consider first the special case when $I = \{1, 2, 3, 4, 5\}$. An extreme point of the set of partially exchangeable measures will begin by running through a finite string of transient states chosen from a subset $T \subset I$. Suppose

$T = \{1, 2\}$, so the process begins with a string like 121221. The number of transitions (and so the length) of the initial string is fixed, but all permutations of the string of 1's and 2's which have the same number of transitions are equally likely. The process then goes to one of the recurrent states, say 3, and thereafter continues as a recurrent Markov chain on states 3, 4, and 5.

In the case of a general finite state space an extreme point $\pi_\theta \in \mathcal{E}$ can be indexed by the parameter

(20) $$\theta = (T, i, M, j, p),$$

where T is a subset of I representing the class of transient states under π_θ, $i \in T$ is the initial state of the process, M is a matrix of transition counts for the transient states determined by $\pi_\theta, j \notin T$ is the first recurrent state, and p is a matrix of transition probabilities for the recurrent states (those in $I - T$) of the process. We require the states in $I - T$ to form a recurrent class; thus p^n has all entries positive for some integer n (Feller (1968) Section XV.4). The process determined by π_θ starts in state i, then runs through a finite string of transient states consistent with the matrix of counts M. The process then moves to state j, and then continues as a Markov chain on the states in $I - T$ with transition matrix p. We will also allow T to be the empty set and M to be zero. In this case we write $\theta = (j, p)$ and the process π_θ is a recurrent Markov chain starting at state j.

It must be argued that π_θ and mixtures of the π_θ are partially exchangeable. This is a consequence of the following combinatorial fact. The proof is straightforward and omitted.

(21) LEMMA. *Let σ be a string of states in T and ρ a string of states in $I - T$. If $\sigma\rho \sim \psi$ then $\psi = \sigma'\rho'$ with $\sigma' \sim \sigma$ and $\rho' \sim \rho$.*

We can now state the main result of this section.

(22) THEOREM. *For θ defined by* (20), *the measures π_θ are the extreme points of the class $\mathfrak{M}$ of partially exchangeable probabilities on a finite state space I. For any $P \in \mathfrak{M}$ there is a unique measure μ on the extreme points $\mathcal{E}$ of $\mathfrak{M}$ such that for any Borel set $A \subset I^\infty$,*

(23) $$P(A) = \int_{\mathcal{E}} \pi_\theta(A)\mu(d\theta).$$

PROOF. Let $P \in \mathfrak{M}$ be fixed. We first prove (23) and then argue uniqueness. The proof of (23) works by successively conditioning on the set of transient states T, the initial state i, the transition matrix M of the transient state, and the first recurrent state j. At each stage of conditioning we must argue that the conditional distribution is still partially exchangeable. We now condition on the transient states.

Let $T \subset I$. Let B be the event that the states in T occur finitely often in X, while the states in T^c occur infinitely often. If $P(B) > 0$, we claim $P(\cdot|B)$ is partially exchangeable. To simplify the writing, suppose $T = \{1, 2, \cdots, t\}$ and $T^c = \{t+1, \cdots, N\}$. Let $\sigma \sim \tau$. Define $B(\sigma)$ to be the set of points in I^∞ which start with σ

and in which states in T occur finitely often and states in T^c occur infinitely often. Similarly define $B(\tau)$. We must show that $P(B(\sigma)) = P(B(\tau))$. This requires some notation. Let $B(\sigma; n_1, n_2, \cdots, n_t; n_{t+1}, \cdots, n_N)$ be the set of points I^∞ which start with the string σ, have exactly n_k symbols k for $1 \leqslant k \leqslant t$ and more than n_k symbols k for $t+1 \leqslant k \leqslant N$. Similarly define $B(\tau; n_1, \cdots, n_t; n_{t+1}, \cdots, n_N)$. Since $\sigma \sim \tau$ we have $\sigma\rho \sim \tau\rho$ for any string ρ. Thus $P(A_{\sigma\rho}) = P(A_{\tau\rho})$. Take the sum as ρ runs through strings with exactly n_k symbols k for $1 \leqslant k \leqslant t$ and more than n_k symbols k for $t+1 \leqslant k \leqslant N$. We see that

$$P[B(\sigma; n_1, \cdots, n_t; n_{t+1}, \cdots, n_N)] = P[B(\tau; n_1, \cdots, n_t; n_{t+1}, \cdots, n_N)].$$

Next, write $B(\sigma; n_1, \cdots, n_t)$ for the points in I^∞ which start with σ, then have exactly n_k symbols for $1 \leqslant k \leqslant t$, and the symbols k with $t+1 \leqslant k \leqslant N$ occurring infinitely often. The sets $B(\sigma; n_1, \cdots, n_t; n_{t+1}, \cdots, n_N)$ decrease to the set $B(\sigma; n_1, \cdots, n_t)$ as $n_{t+1}, \cdots, n_N$ increase to infinity. Thus

$$P[B(\sigma; n_1, \cdots, n_t)] = P[B(\tau; n_1, \cdots, n_t)].$$

Finally, summing over all possible indices $n_1, n_2, \cdots, n_t$, we have $P[B(\sigma)] = P[B(\tau)]$ as was to be shown.

It is clear that we may also assume $P(X_0 = i) = 1$ and by the above argument that with P probability 1 all states in T occur finitely often and all states in T^c occur infinitely often. We now argue that P cannot have a transition from a state $k \in T^c$ to a state $k' \in T$. To see this, consider the k block process $\{Y_n\}$. This process is exchangeable. Let Z_n be one or zero as the nth k block has a k to k' transition or not. The Z_n's are exchangeable. If there is positive probability of a k to k' transition, then $P(Z_1 = 1) > 0$; and de Finetti's theorem (or the Poincaré recurrence theorem) shows that $P(Z_n = 1 \text{ infinitely often}) > 0$. This would imply that with positive probability k' occurred infinitely often, contrary to assumption. We have thus shown that any sample path of the probability P conditioned on the event B must start with a (possibly empty) string of states in T and continue with states in T^c.

The next stage of the argument is to show that conditioning P on the transition matrix of the transient states and the first recurrent state does not destroy partial exchangeability. Let M be a transition count matrix for transient states. We are assuming the processes starts with $i \in T$. Let j be a state in T^c. There are only a finite number of strings say $\sigma_1, \sigma_2, \cdots, \sigma_n$ which start with i, and have M as transition matrix. These strings all have the same length, the same final state, and contain only transient states. Clearly all these strings are equivalent and so $P(A_{\sigma_m j})$ does not depend on m. Conditioning on the matrix M and first recurrent state j is the same as conditioning on the event $C = \cup_{m=1}^n A_{\sigma_m j}$. To show that conditioning on C does not destroy partial exchangeability, let $\sigma \sim \tau$. We must show

$$P(A_\sigma \cap C) = P(A_\tau \cap C). \tag{24}$$

We may assume without loss of generality that σ (and hence τ) has no transition from T^c to T: otherwise both sides of (24) vanish.

There are two cases:

CASE 1. *Length* $\sigma \geqslant$ *length* $\sigma_1 j$. If either A_σ or A_τ meets C, then (21) shows that $\sigma = \sigma_{ij\rho}$ and $\tau = \sigma_{i'j\rho'}$ with $\rho \sim \rho'$, so (24) follows from (8). Otherwise, both sides of (24) are zero.

CASE 2. *Length* $\sigma <$ *length* $\sigma_1 j$. Let σ_m' be the typical σ_m which begin with σ. Write $\sigma_m' = \sigma\rho_m'$. Let $\sigma_m'', \cdots, \sigma_s''$ be the typical σ_m which begin with τ. Write $\sigma_m'' = \tau\rho_m''$. The sets $\{\rho_m'\}$ and $\{\rho_m''\}$ are in one-to-one correspondence because any ρ_m'' gives an admissible completion for σ and any ρ_m' gives an admissible completion for τ. The left side of equation (24) equals $\Sigma_m P(A_{\sigma\rho_m'})$. The right side of equation (24) equals $\Sigma_m P(A_{\tau\rho_m''})$. These sums are equal term by term, so (24) is true.

Combining arguments, we have shown that we may successively condition on the transient states, initial state, first recurrent state and matrix of transition counts for the transient states without destroying partial exchangeability. The conditional process, starting from the first recurrent state, is recurrent and partially exchangeable. It follows from Theorem 7 that the conditional processes is a mixture of Markov chains. This completes the proof of the representation (23). We now argue that the representation is unique.

Indeed, with θ as defined in (20), let

$$B = \{\theta: \theta_1 = T, \theta_2 = i, \theta_3 = M, \theta_3 = j\}$$

and define an event A for the coordinate process X_i on I^∞ by $A = \{X$ has T as transient states and T^c as recurrent states, $X_0 = i$, X is consistent with M and has first recurrent state $j\}$. Then $\pi_\theta(A) = 1_B(\theta)$ where $1_B(\cdot)$ is the indicator of the set B. Hence the representation (23) implies

$$P(A) = \int \pi_\theta(A)\mu(d\theta) = \mu(B).$$

So the μ-distribution of the first four coordinates of θ is uniquely determined by P. Now given T, i, M, j the process starting at j is a recurrent partially exchangeable process and Theorem 7 implies that the process is a uniquely determined mixture of Markov chains. Thus the μ-distribution of the matrix of transition probabilities (the last coordinate of θ) given T, i, M and j is uniquely determined, and so μ is uniquely determined. Uniqueness in (23) implies that the π_θ are extreme points of $\mathfrak{M}$ as in Proposition 1.4 of Phelps (1966). $\square$

4. Complements to the main results.

Limit theorems and zero-one laws. Just as for exchangeability, there is a collection of easily proved limit theorems for partially exchangeable processes. We will only discuss the recurrent case. The argument for Theorem (7) shows that the process conditioned on the initial state i and the tail σ-field of the i-blocks is Markov. Thus, known results for Markov chains can be used to yield results for recurrent partially exchangeable processes. Here is one consequence of this argument.

(25) REMARK. Let $T^{(n)}$ be the doubly infinite matrix with i, j entry the number of transitions from state i to state j divided by the number of transitions from state i up to time n. Then $T^{(n)}$ converges almost surely to a stochastic matrix T in the topology of coordinate-wise convergence. Write $\mathcal{P}$ for the set of stochastic matrices. We construct a probability law μ on $I \times \mathcal{P}$ as follows: the projection of μ on I coincides with the starting distribution of X; the conditional distribution on $\mathcal{P}$ given the starting state coincides with the probability law of T given the starting state of X. The measure μ is the mixing measure in Theorem (7).

To state the zero-one laws, we define several σ-fields for the processes X. Let $\mathcal{I}$ be the invariant field for the shift operating on I^∞. Let $\mathcal{T}$ be the tail field. Let $\mathcal{E}$ be the exchangeable field. Further, let $\mathcal{F}^{(n)}$ be the σ-field generated by X_0 and the matrices $T^{(n)}, T^{(n+1)}, \cdots$ defined in (25). It is easily seen that $\mathcal{F}^{(n)}$ is also the σ-field generated by X_0, $T^{(n)}$ and $X_{n+1}, X_{n+2}, \cdots$. Let $\mathcal{F} = \cap \mathcal{F}^{(n)}$. The σ-field $\mathcal{F}$ is called the *partially exchangeable σ-field*. It is straightforward to argue that

$$\mathcal{I} \subset \mathcal{T} \subset \mathcal{E} \subset \mathcal{F} \tag{26}$$

A corollary to Theorem (7) is that any pair of the σ-fields in (26) agrèe up to null sets under a recurrent partially exchangeable probability P. For instance, if $A \in \mathcal{F}$ there is $A' \in \mathcal{I}$ with $P(A \Delta A') = 0$. A second consequence of Theorem (7) is the following zero-one law: any recurrent Markov chain assigns probability zero or one to events in $\mathcal{F}$. These results are related to results for exchangeable processes given by Olshen (1971). Zero-one laws for Markov chains are discussed by Blackwell and Freedman (1964).

Infinite state space—general case. Without the recurrence condition (4), the description of the extreme points of the partially exchangeable processes with infinite state space is not as neat as Theorems (7) and (22). It can be shown that there are three types of extreme points:

(1) Recurrent Markov chains,
(2) Processes which start with a fixed length string of transient states and continue as recurrent Markov chains,
(3) Totally transient processes, where each state occurs only finitely often.

Type 2 extreme points are basically the same as the measures π_θ introduced in (20). For type 3 extreme points consider the doubly infinite random matrix M with ij entry equal to the total number of i to j transitions in X. Since every state occurs finitely often, M is almost surely finite. The extreme points of type 3 can be characterized as follows. Let $M_n(i, j)$ be the number of transitions from i to j in X, up to time n. So $M_n \uparrow M$. Then X is extreme if and only if $\mathcal{F} = \cap_n \sigma(M_n, X_{n+1}, X_{n+2}, \cdots)$ is trivial. We conjecture, but cannot prove, that $\mathcal{F}$ is spanned by M.

Group invariance. The equivalence relation $\sim$ is connected to invariance under the group of block-switch transformations. Informally, for $\sigma \in I^n$, a typical block-switch transformation T focuses on two disjoint blocks in σ. If the blocks begin with the same symbols, and end with the same symbols, then $T(\sigma)$ is σ with the two blocks interchanged. Formally, for $\sigma \in I^n$, let $0 \leqslant a \leqslant b \leqslant c \leqslant d \leqslant n$ and $\theta = (a, b, c, d)$. If $\sigma_a \neq \sigma_c$ or $\sigma_b \neq \sigma_d$ then $T_\theta(\sigma) = \sigma$. If $\sigma_a = \sigma_c$ and $\sigma_b = \sigma_d$ then

$$\sigma = \sigma_0\sigma_1 \cdots \sigma_{a-1} \underbrace{\sigma_a \cdots \sigma_b}_{\text{block 1}} \sigma_{b+1} \cdots \sigma_{c-1} \underbrace{\sigma_c \cdots \sigma_d}_{\text{block 2}} \sigma_{d+1} \cdots \sigma_n$$

and

$$T_\theta(\sigma) = \sigma_0\sigma_1 \cdots \sigma_{a-1} \underbrace{\sigma_c\sigma_{c+1} \cdots \sigma_d}_{\text{block 2}} \sigma_{b+1} \cdots \sigma_{c-1} \underbrace{\sigma_a \cdots \sigma_b}_{\text{block 1}} \sigma_{d+1} \cdots \sigma_n.$$

The T_θ are plainly $1-1$ maps from I^n onto I^n. If $\sigma \in I^n$, then $\tau = T_{\theta_1} \circ T_{\theta_2} \circ \cdots \circ T_{\theta_k}(\sigma)$ implies $\sigma \sim \tau$ since T_{θ_i} does not change the initial state nor transitions. The next result gives a converse.

(27) PROPOSITION. *If $\sigma \sim \tau \in I^n$, then there exist $T_{\theta_1}, \cdots, T_{\theta_k}$ on I^n such that $\sigma = T_{\theta_1} \circ \cdots \circ T_{\theta_k}(\tau)$.*

PROOF. Since $\sigma \sim \tau$, we have $\sigma_1 = \tau_1$ and $\sigma_n = \tau_n$. Without loss of generality suppose $\sigma_2 \neq \tau_2$. Then τ must have a $\tau_1\sigma_2$ transition at some later place. Let S_1 be the initial string of symbols in τ before the $\tau_1\sigma_2$; and let S_2 be the final string of symbols in τ after the $\tau_1\sigma_2$ transition, where S_2 starts with σ_2. Thus, the two strings appear as:

$$\tau = \overbrace{\tau_1\tau_2 \cdots}^{S_1} \tau_1 \underbrace{\sigma_2 \cdots \tau_n}_{S_2}$$

$$\sigma = \sigma_1\sigma_2, \cdots, \sigma_n \qquad \sigma_1 = \tau_1, \sigma_n = \tau_n.$$

We will argue that some symbol in S_1 occurs in S_2. By way of contradiction, suppose S_1 and S_2 have no common symbols. Then we show that S_1 only contains the symbol τ_1. Indeed, σ_2 does not appear in S_1, since σ_2 appears in S_2. Then the $\sigma_2\sigma_3$ transition in τ must appear in S_2, so that σ_3 does not appear in S_1. Note that neither σ_2 nor σ_3 can equal σ_1 because τ_1 equals σ_1. Similarly, $\sigma_4, \cdots, \sigma_n$ do not appear in S_1 and none of these symbols equals σ_1. Thus, if S_1 and S_2 have no common symbols, then σ contains at most one symbol σ_1 and τ must appear as:

$$\tau_1\tau_1 \cdots \tau_1 \underbrace{\sigma_2 \cdots \tau_n},$$

with no symbol τ_1 appearing in S_2. But τ cannot be $\tau_1\sigma_2 \cdots$ because $\tau_2 \neq \sigma_2$. Now τ has a $\tau_1\tau_1$ transition, so σ must too. In particular σ contains at least two symbols σ_1. This is a contradiction. Thus we have shown that some symbol (say β) in S_1 occurs in S_2.

Now the block-switch T which switches the first τ_1 to β block with the block that starts $\tau_1\sigma_1$ to β insures that $T(\tau)$ matches σ in its first two places. Continue inductively to complete the proof. □

Proposition (27) has been formulated in terms of finite strings, but it is clear that block-switch transformations that operate on I^n also operate on I^∞.

(28) COROLLARY. *A measure P on I^∞ is partially exchangeable if and only if P is invariant under all block-switch transformations.*

This corollary means that some results from ergodic theory apply to the present problem. In particular, either (Farrell (1962)) or (Chapter 10 of Phelps (1966)) implies that the partially exchangeable measures are mixtures of ergodic measures ("ergodic" means the measures are zero or one on the partially exchangeable field $\mathcal{F}$ defined in (26)). As usual, a more detailed description of the extreme points requires detailed arguments such as those presented in Sections 2 and 3 here.

There are two other approaches to the proof of Theorem (7) we want to mention. The first is to carry out de Finetti's original suggestion of proving a version of (7) for finite partially exchangeable sequences and then passing to the limit. We can do this for $I = \{0, 1\}$, but the argument does not generalize to other finite I. Some details are in Diaconis and Freedman (1978a). The second approach is to use the machinery associated with Gibbs states. We give this argument in some detail in Diaconis and Freedman (1978c). The special case of stationary processes with $I = \{0, 1\}$ is treated by Georgii (1975).

Mixtures of Markov chains. Without the recurrence condition (4) the class of partially exchangeable processes is larger than the class of mixtures of Markov chains. A natural problem is to find a condition to add to partial exchangeability which characterizes mixtures of Markov chains. When $I = \{0, 1\}$, a condition that works can be described as follows: let μ_i be the probability that the process starts with i or more zeros and ends with all ones. Let ν_i be the probability that the process starts with i or more ones and ends with all zeros.

(29) *A probability P on the Borel sets of $\{0, 1\}^\infty$ is a mixture of Markov chains if and only if*

(30) *P is partially exchangeable, and the sequences $\{\mu_i\}$, $\{\nu_i\}$ are completely monotone.*

PROOF. It is clear that mixtures of Markov chains satisfy (30). Conversely, suppose P satisfies (30). An easy argument shows that conditioning P on the algebra generated by the two events:

$$A = \{\text{The path of } X \text{ ends in all zeros}\}$$
$$B = \{\text{The path of } X \text{ ends in all ones}\}$$

leaves P partially exchangeable. Thus,

$$P(\cdot) = P(\cdot|A)P(A) + P(\cdot|B)P(B) + P(\cdot|(A \cup B)^c)P((A \cup B)^c). \tag{31}$$

The first two conditional probabilities on the right side of (31) are mixtures of Markov chains by the Hausdorff moment theorem, using the second condition of (30). The third conditional probability in (31) is a recurrent partially exchangeable probability, which is a mixture of Markov chains by Theorem (7). Thus, P is a mixture of Markov chains.

We have not been able to formulate a neat analog of (29) for state spaces I containing three elements. As noted by Freedman ((1962a), page 116) even when I has three elements, the mixing measure need not be uniquely defined. Consider the following example.

(32) EXAMPLE. Consider mixtures of Markov chains with three states 0, 1, 2, starting from 0, with only the following transitions permitted:

$$0 \to 1 \text{ or } 2, \quad 1 \to 0 \text{ or } 2, \quad \text{and } 2 \to 2.$$

(The other transitions are required to have probability 0.) There are thus two kinds of sample paths: paths containing a 02 transition and paths containing a 12 transition. The probabilities of the two kinds of paths are given by:

$$\int_S (p_{01}p_{10})^m (1 - p_{01})\mu(dp_{01}, dp_{10}) \qquad m = 0, 1, 2, \cdots,$$

$$\int_S (p_{01}p_{10})^n p_{01}(1 - p_{10})\mu(dp_{01}, dp_{10}) \qquad n = 0, 1, 2, \cdots,$$

where S is the open unit square. Suppose now that μ is concentrated on the curve $p_{01}p_{10} = \frac{1}{4}$. Then only two moments remain to separate such μ's from one another: $\int_S (1 - p_{01})\mu(dp_{01})$ and $\int_S p_{01}(1 - p_{10})\mu(dp_{01}, dp_{10})$. There are clearly many μ's with these moments so μ is not uniquely defined.

Stationary partially exchangeable processes. Freedman (1962b) showed that any stationary partially exchangeable process is a mixture of stationary Markov chains —chains which start with their stationary distribution. We will now derive this from Theorem (7).

We need some notation: if α is a subprobability distribution on I and p is a transition matrix, write $\langle \alpha, p \rangle$ for the law of the process which picks a starting state i from α and then evolves according to p. For a stationary partially exchangeable probability P let $\theta_k = P(X_0 = k)$. Suppose $\theta_k > 0$. By the Poincaré recurrence theorem, given $\{X_0 = k\}$, the process $\{X_n\}$ is recurrent in the sense of (4) and is therefore a mixture of Markov chains by Theorem 7:

$$P(A \mid X_0 = k) = \int \langle \delta_{k,p} \rangle (A) \mu_k(dp) \tag{33}$$

where δ_k is a point mass at k. Clearly, μ_k-almost all matrices p have k as a recurrent state. Put $\{X_n = k \text{ i.o.}\}$ for A in (33). From (33),

$$P = \Sigma_k \theta_k \int \langle \delta_{k,p} \rangle \mu_k(dp). \tag{34}$$

We will show that μ_k assigns mass 1 to matrices p for which state k is positive recurrent and that δ_k in (34) can be replaced by a stationary distribution for p. Define $\pi_k(p)(j) = C_1 \lim_{n\to\infty} p^n(k, j)$, where "$C_1$ lim" stands for the first Cesaro limit. As is well known (see, for example, chapter 1 of Freedman (1971)) this limit always exists. Three cases must be considered in what follows:

Case 1. If k is transient for p, then the description of $\pi_k(p)$ is complicated.

Case 2. If k is null recurrent for p, then $\pi_k(p) \equiv 0$.

Case 3. If k is positive recurrent, then $\pi_k(p)$ is the unique stationary distribution for the recurrence class containing k.

To simplify notation, let A be the set

$$\{X_0 = i_0, X_1 = i_1, \cdots, X_m = i_m\}.$$

Then by (33),

$$\begin{aligned} P(A) &= P(X_n = i_0, X_{n+1} = i_1, \cdots, X_{n+m} = i_m) \\ &= \Sigma_k \theta_k \int p^n(k, i_0) p(i_0, i_1) \cdots p(i_{m-1}, i_m) \mu_k(dp). \end{aligned}$$

Take C_1 limits of both sides to get

$$P(A) = \Sigma \theta_k \int \langle \pi_k(p), p \rangle (A) \mu_k(dp). \tag{35}$$

If $\theta_k > 0$ we claim that with μ_k-probability 1, the state k cannot be null recurrent for p. Otherwise, $\mu_k\{p: \pi_k(p) \equiv 0\} > 0$ and P has total mass less than 1. As a result, if $\theta_k > 0$, then μ_k assigns mass 1 to the matrices p for which k is positive recurrent, and $\pi_k(p)$ is a stationary distribution. This completes the argument.

APPENDIX

De Finetti's theorem in Polish spaces. The form of de Finetti's theorem used in Theorem (7) is quite widely known; but as we cannot supply a reference for precisely what we need, a proof will be given instead. To state the result in a bit more generality, let $Y_0, Y_1, \cdots$ be a stochastic process on the probability triple $(\Omega, \mathfrak{F}, P)$—where $(\Omega, \mathfrak{F})$ is assumed Polish. Suppose the Y_i take values in a Polish space S, equipped with the Borel σ-field $\mathfrak{B}$. Now exchangeability means that $(Y_{\pi(0)}, Y_{\pi(1)}, \cdots)$ is distributed as $(Y_0, Y_1, \cdots)$ for any finite permutation π of the nonnegative integers. As before, let $\mathfrak{F}^{(n)}$ be the σ-field spanned by $(Y_n, Y_{n+1}, \cdots)$ and let $\mathfrak{F}^{(\infty)} = \cap_n \mathfrak{F}^{(n)}$, the tail σ-field.

(36) THEOREM. *Let $Y_0, Y_1, \cdots$ be an exchangeable stochastic process on $(\Omega, \mathfrak{F}, P)$, with $(\Omega, \mathfrak{F})$ Polish and the Y_i taking values in a compact metric space S. Let $P_\omega(A)$ be a regular conditional probability on $\mathfrak{F}$ given the tail σ-field $\mathfrak{F}^{(\infty)}$ of the Y_i. For P-almost all ω, relative to P_ω, the Y_i are independent and identically distributed.*

Some lemmas will be helpful.

(37) LEMMA. *Let $A \in \mathfrak{B}$. Then* $\frac{1}{n}\Sigma_{v=0}^{n-1}1_A(Y_v(\omega)) \to P_\omega(Y_0 \in A)$ *for P-almost all* ω, *the exceptional null set depending on* A.

PROOF. Let $S_n = \Sigma_{v=0}^{n-1}1_A(Y_v)$, let $\hat{\mathfrak{F}}^{(n)}$ be the σ-field spanned by S_n and $Y_n, Y_{n+1}, \cdots$. Clearly, $\hat{\mathfrak{F}}^{(n+1)} \subset \hat{\mathfrak{F}}^{(n)}$; let $\hat{\mathfrak{F}}^{(\infty)} = \cap_n \hat{\mathfrak{F}}^{(n)}$. For $0 \leqslant v \leqslant n-1$, by exchangeability,

$$P\{Y_0 \in A \quad \text{and} \quad S_n > y \quad \text{and} \quad Y_n \in A_n, \cdots, Y_{n+m} \in A_{n+m}\}$$
$$= P\{Y_v \in A \quad \text{and} \quad S_n > y \quad \text{and} \quad Y_n \in A_n, \cdots, Y_{n+m} \in A_{n+m}\}.$$

Thus,

$$P\{Y_0 \in A | \hat{\mathfrak{F}}^{(n)}\} = P\{Y_v \in A | \hat{\mathfrak{F}}^{(n)}\}$$
$$= E\left\{\frac{1}{n}S_n | \hat{\mathfrak{F}}^{(n)}\right\} = \frac{1}{n}S_n.$$

Now $\frac{1}{n}S_n \to P\{Y_0 \in A | \hat{\mathfrak{F}}^{(\infty)}\}$ by the backwards martingale theorem. But $\lim \frac{1}{n}S_n$ is $\mathfrak{F}^{(\infty)}$-measurable, and $\mathfrak{F}^{(\infty)} \subset \hat{\mathfrak{F}}^{(\infty)}$. □

(38) LEMMA. *Let $A_r \in \mathfrak{B}$ for $r = 0, \cdots, R$. Let*

$$T_n(\omega) = \frac{1}{n(n-1)\cdots(n-R+1)}\Sigma_{i_0,\ldots,i_R}\Pi_{r=0}^R 1_{A_r}(Y_{i_r}(\omega)),$$

where $i_0, \cdots, i_R$ are all distinct and range from 0 *to* n. *Then* $T_n(\omega) \to P_\omega(Y_r \in A_r$ *for* $0 \leqslant r \leqslant R)$ *for P-almost all* ω, *the exceptional null set depending on* $A_0, \cdots, A_R$.

PROOF. As in (37), which is the case $R = 0$. □

(39) COROLLARY. *Let $A_r \in \mathfrak{B}$ for $0 \leqslant r \leqslant R$. Then*

(40) $$P_\omega\{Y_r \in A_r \text{ for } 0 \leqslant r \leqslant R\} = \Pi_{r=0}^R P_\omega\{Y_0 = A_r\}$$

for P-almost all ω, *the exceptional null set depending on* $A_0, \cdots, A_R$.

PROOF. Clearly,

$$\frac{n(n-1)\cdots(n-R+1)}{n^R}T_n = \Pi_{r=0}^R\left[\frac{1}{n}\Sigma_{v=0}^{n-1}1_{A_r}(Y_v)\right] + 0\left(\frac{1}{n}\right).$$

Now use (37) and (38). □

PROOF OF THEOREM (36). Let $\mathfrak{B}_0$ be a countable generating algebra of $\mathfrak{B}$. By countable additivity, there is a P-null set N, such that $\omega \notin N$ makes (40) hold simultaneously for all R and all $A_r \in \mathfrak{B}_0$. Then, by the monotone class argument, for $\omega \notin N$, relation (40) holds simultaneously for all R and all $A_r \in \mathfrak{B}$. □

REFERENCES

BLACKWELL, D. AND FREEDMAN, D. (1964). The tail σ-field of a Markov chain and a theorem of Orey. *Ann. Math. Statist.* **35** 1291–1295.

DIACONIS, P. (1977). Finite forms of de Finetti's theorem on exchangeability. *Synthese* **36** 271–281.

DIACONIS, P. and FREEDMAN, D. (1978a). de Finetti's generalizations of exchangeability. To appear in *Studies in Inductive Logic and Probability*. (R. Jeffrey, ed.)

DIACONIS, P. and FREEDMAN, D. (1978b). Finite exchangeable sequences. To appear in *Ann. Probability*.

DIACONIS, P. and FREEDMAN, D. (1978c). Exchangeability and sufficiency. Unpublished manuscript.

FARRELL, R. H. (1962). Representation of invariant measures. *Illinois J. Math.* **6** 447–467.

FELLER, W. (1971). *An Introduction to Probability Theory and Its Applications Vol.* 2. (2nd ed.) Wiley, New York.

DE FINETTI, B. (1959). La probabilità e la statistica nei rapporti con induzione; secondo i diversi punti di vista. In *Corso C. I. M. E. su Induzione e Statistica*, Cremonese, Rome (translated in Chapter 9 of de Finetti 1974).

DE FINETTI, B. (1974). *Probability, Induction, and Statistics*. Wiley, New York.

FREEDMAN, D. (1962a). Mixtures of Markov Processes. *Ann. Math. Statist.* **33** 114–118.

FREEDMAN, D. (1962b). Invariants under mixing which generalize de Finetti's theorem. *Ann. Math. Statist.* **33** 916–923.

FREEDMAN, D. (1971). *Markov Chains*. Holden-Day, San Francisco.

GEORGII, H. O. (1975). Canonical Gibbs states, their relation to Gibbs states and applications to two-valued Markov chains. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **32** 272–300.

HEWITT, E. and SAVAGE, L. J. (1955). Symmetric measures on Cartesian products. *Trans. Amer. Math. Soc.* **80** 470–501.

KINGMAN, J. F. C. (1978). Uses of exchangeability. *Ann. Probability* **6** 183–197.

MARTIN, J. J. (1967). *Bayesian Decision Problems and Markov Chains*, Wiley, New York.

OLSHEN, R. (1971). The coincidence of measure algebras under an exchangeable probability. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **18** 153–158.

OLSHEN, R. (1974). A note on exchangeable sequences. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **28** 317–321.

PHELPS, R. (1966). *Lectures on Choquet's Theorem*. Van-Nostrand, New York.

BELL LABORATORIES
600 MOUNTAIN AVENUE
MURRAY HILL, NEW JERSEY 07974

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA AT BERKELEY
BERKELEY, CALIFORNIA 94720