The Annals of Probability
1986, Vol. 14, No. 4, 1418–1427

A FINITE FORM OF DE FINETTI'S THEOREM FOR STATIONARY MARKOV EXCHANGEABILITY

BY ARIF ZAMAN

Florida State University

De Finetti's theorem for stationary Markov exchangeability states that a sequence having a stationary and Markov exchangeable distribution is a mixture of Markov chains. A finite version of this theorem is given by considering a finite sequence $X_1, \dots, X_n$ which is stationary and Markov exchangeable. It is shown that any portion of k consecutive elements, say $X_1, \dots, X_k$ for $k < n$, is nearly a mixture of Markov chains (the distance measured in the variation norm).

1. Introduction. In de Finetti (1959; 1974, pages 217–220) it was suggested that all infinite length Markov exchangeable sequences are mixtures of Markov chains. Even though the assertion turned out to be false, it was nearly correct. Freedman (1962) showed that under the additional condition of stationarity, Markov exchangeability is equivalent to being a mixture of Markov chains. Diaconis and Freedman (1980a) later relaxed the condition of stationarity to that of recurrence. They showed that recurrence is both a necessary and a sufficient condition for the equivalence to hold, and categorized all the exceptions which can occur when recurrence does not hold. Both of these results are called de Finetti's theorems for Markov exchangeability (dFTME).

The above mentioned results of dFTME have the flavor of the well known de Finetti theorem (dFT) which states that infinite length exchangeable sequences are mixtures of i.i.d. sequences. All of these results are false for finite length exchangeable or Markov exchangeable sequences, but Diaconis and Freedman (1980b) were able to extract a finite version for dFT. Their method consisted of getting a mixture-of-urns representation for finite exchangeability, finding a bound for the "distance" (variation norm of projections) between an urn measure and an i.i.d. measure, and combining these to show that a finite exchangeable measure is "close" to a mixture of i.i.d. measures. The constructive nature of the proof and its simplicity provide insight into the workings of dFT, and yet the finite version is powerful enough that the most general known forms of dFT are simple consequences of it.

This same program of obtaining finite versions of asymptotic results is continued here for dFTME. Section 2 defines Markov exchangeability. Section 3 describes the finite version of dFT and some related results from Diaconis and Freedman (1980b). Section 4 describes a mixture-of-glued-urn-models representation from Zaman (1984a). The introduction of new notation allows that model to be useful in the next section. Section 5 finds an upper bound for the distance

Received November 1984.
AMS 1980 *subject classifications*. Primary 60J05; secondary 60G10.
Key words and phrases. de Finetti's theorem, Markov exchangeability, stationary processes, Markov chains.

between a glued-urn model and a Markov chain. Section 6 discusses why the bound of Section 5 is not enough and the further constraint of stationarity is needed. In Section 7, a bound for a ratio of stationary events is quoted from Zaman (1984b) and used to prove the finite form of dFTME under the assumption of stationarity (Theorem 6). This final theroem is strong enough to imply the infinite form of dFTME under the added constraints of stationarity and a finite state space.

2. Markov chains and Markov exchangeability. For a set C, the symbol C^n denotes the set of all sequences of length n taking values in C. For $X \in C^n$ and $k \le n$, the initial portion $(X_1, \dots, X_k)$ is denoted by $X^{(k)} \in C^k$. When $X \in C^n$ has distribution P, the distribution of $X^{(k)}$ is denoted by $P^{(k)}$.

All Markov chains here are assumed to have a finite state space C and stationary transition probabilities. Without loss of generality we take $C = \{1, 2, \dots, c\}$. A Markov chain probability P on $X \in C^n$ is parametrized by its transition probabilities $a_{ij} = P\{X_{k+1} = j \text{ given } X_k = i\}$ for $i, j \in C$ and $k = 1, \dots, n-1$ and its initial state probabilities $a_{0i} = P\{X_1 = i\}$ for $i \in C$. This $c+1$ by c matrix of parameters is denoted by $\tilde{a}$ and the Markov measure P is denoted by $R_{\tilde{a}}^{(n)}$. It is clear that a valid value for $\tilde{a}$ must lie in the set $A = \{\tilde{a}: \sum_{j \in C} a_{ij} = 1 \text{ for } i = 0, 1, \dots, c\}$.

If $\{P_\lambda\}_{\lambda \in \Lambda}$ is a parametric family of probability measures and μ is a probability on Λ then P_μ denotes the mixture measure given by $P_\mu = \int_\Lambda P_\lambda \mu \, d\lambda$. P_μ corresponds to the two stage procedure of picking a random $\lambda \in \Lambda$ according to μ, and then using P_λ.

The set of all mixtures of Markov chains of length n with state space C is denoted by $\mathscr{M}_C^n = \{R_\mu^{(n)}: \mu \text{ is a probability measure on } A\}$. The set of mixtures of infinite length Markov chains, denoted by $\mathscr{M}_C$, can be defined by $P \in \mathscr{M}_C$ iff $P^{(n)} = R_\mu^{(n)}$ for all n with the same mixing measure μ.

For $X \in C^n$, let $t_{ij}(X)$ be the number of i to j transitions in X, i.e.,

$$t_{ij}(X) = \sum_{k=1}^{n-1} I\{(X_k, X_{k+1}) = (i, j)\}. \tag{1}$$

Using this equation, the measure $R_\mu^{(n)}$ can be written out explicitly as

$$R_\mu^{(n)}\{X\} = \int_A a_{0X_1} \prod_{i, j \in C} a_{ij}^{t_{ij}(X)} \mu \, d\tilde{a}. \tag{2}$$

This shows that if $P \in \mathscr{M}_c^n$ then $P\{X\}$ depends upon X only through its initial state X_1 and the transition count matrix $t(X) = [t_{ij}(X)]_{i, j \in C}$.

Let $\mathscr{E}_C^n$ denote all measures on C^n with the above mentioned property, i.e., $P \in \mathscr{E}_C^n$ iff for all $X, Y \in C^n$

$$P\{X\} = P\{Y\} \quad \text{if } X_1 = Y_1 \text{ and } t(X) = t(Y). \tag{3}$$

The set $\mathscr{E}_C^n$ is the set of all Markov exchangeable probabilities on C^n, or as de Finetti calls them, partially exchangeable probabilities of the Markov type. The set of infinite length Markov exchangeable sequences, denoted by $\mathscr{E}_C$, can be defined by $P \in \mathscr{E}_C$ iff (3) holds for all $P^{(n)}$ with $n < \infty$.

3. Urns and the finite de Finetti theorem. The symbol $U = (u_1, \dots, u_c)$ will be used to denote an urn containing a total of $u_. = \sum_{i=1}^{c} u_i$ (the convention of using a dot subscript to indicate summation is used throughout) balls, with u_1 of them labelled with a $1, \dots,$ and u_c of them labelled with a c. H_U denotes the measure on sequences in $C^{U.}$ obtained by sampling without replacement from U. M_U is the measure when sampling with replacement (H stands for hypergeometric and M for multinomial). The same symbol u_i is also used as a function which counts the occurrences of i in a finite sequence $X \in C^n$, defined by

$$u_i(X) = \sum_{j=1}^{n} I\{X_j = i\}. \tag{4}$$

Thus if $X \in C^n$ has distribution H_U then $u_i(X) = u_i$ and $n = u$.

The following facts are borrowed from Diaconis and Freedman (1980b): Their equation (12) implies that for any sequence $X \in C^k$ and any urn U with $k \le u_.$,

$$\left[1 - \frac{M_U^{(k)}(X)}{H_U^{(k)}(X)}\right]^+ \le \sum_{i=1}^{c} \frac{u_i(X)}{u_i}, \tag{5}$$

where $[x]^+$ denotes $\max(0, x)$.

Given two probability measures P and Q on the same probability space $(\Omega, \mathscr{B})$, define the variation distance

$$\|P - Q\| = \sup_{B \in \mathscr{B}} |P(B) - Q(B)|.$$

If Ω is finite and all subsets of Ω measurable, then an equivalent form is

$$\|P - Q\| = 2 \sum_{w \in \Omega} [P(w) - Q(w)]^+. \tag{6}$$

With this norm, the finite form of dFT can be stated as:

THEOREM 1 (dFT). *For every exchangeable measure P on C^n there exists a mixture of i.i.d. measures, Q, such that*

$$\|P^{(k)} - Q^{(k)}\| \le 2ck/n.$$

4. Urn models for Markov exchangeability. Combining (1) and (4) for $X \in C^n$ gives

$$t_{i\circ}(X) = u_i(X^{(n-1)}) \tag{7}$$

so that

$$t_{i\circ}(X) + I\{X_n = i\} = u_i(X) = t_{\circ i}(X) + I\{X_1 = i\}. \tag{8}$$

Equation (8) is actually a necessary and sufficient condition for t to be a valid transition matrix, i.e., for $x, y \in C$ the set of all transition matrices of sequences starting at x and ending at y is

$$\begin{aligned} T_{x,y}^n &= \{t(X) \colon X \in C^n;\ X_1 = x;\ X_n = y\} \\ &= \{t \colon t_{i.} + I\{y = i\} = t_{.i} + I\{x = i\}\}. \end{aligned} \tag{9}$$

The notation developed above will be used to describe the urn model for Markov exchangeability developed in Zaman (1984a). Let $x, y \in C$ and $\tau \in T^n_{xy}$ be fixed (thought of as parameters for the distribution about to be described). Let C_y denote the set $C - \{y\}$. Let $f \in C^{C_y}$ be another fixed parameter, i.e., $f_i \in C$ for all $i \in C_y$.

Let $U_1, \dots, U_c$ be c separate urns with $U_i = (\tau_{i1}, \dots, \tau_{ic})$ for $i \in C$. From each urn U_i when $i \in C_y$, take a ball labeled f_i from U_i and "glue" to the bottom of U_i so that it can only be drawn after all other balls have been drawn. Let $X_1 = x$. For $i = 1, \dots, n-1$ let X_{i+1} be the label of a ball drawn without replacement from U_{X_i}. This sequence $X \in C^n$ is random and its distribution will be denoted by $G_{x, y, \tau, f}$ to indicate its parameters (G stands for glue).

A point which has been glossed over is that the above prescription may be impossible. It will suffice for our purposes here to know that there is a set $F \subset C^{C_y}$ depending upon x, y, and τ, such that if the parameter $f \in F$ then all of the instructions of the urn model can always be followed. The precise definition of F can be found in the original paper.

The importance of this urn measure derives from the following theorem.

THEOREM 2. *For every $x, y \in C$ and $\tau \in T^n_{x,y}$ there is a unique probability μ on F such that $G_{x,y,\tau,\mu}$ is an extreme point of the set of measures $\mathscr{E}^n_C$. Moreover any measure in $\mathscr{E}^n_C$ is a mixture of such extreme points.*

The measure $G_{x,y,\tau,f}$ is simply a combination of draws without replacement from urns, and so can be expressed in terms of the H_U notation with some work. For $X \in C^n$ define $s_i(X)$ as the subsequence of X containing the values immediately after the occurrences of the i's in X. For example:

$$\begin{aligned} X &= 2\,3\,2\,3\,1\,2\,1\,1\,3\\ s_1(X) &= \qquad\quad\; 2\;\; 1\,3\\ s_2(X) &= \;\;3\;\;\, 3 \qquad 1\\ s_3(X) &= \quad\;\; 2\;\;\, 1 \end{aligned}$$

(s_i stands for successors of i).

The following is a list of definitions (denoted by $\triangleq$) and properties that can easily be verified for $X \in C^n$ drawn by the urn model $G_{n,y,\tau,f}$ (these subscripts will not be repeated):

$$\begin{aligned} \tau(i,j) &\triangleq \tau_{ij},\\ \tau(i,j;k) &\triangleq \tau_{ij}(X^{(k)}) = u_j(s_i(X^{(k)})),\\ \tau(i,j;n) &= \tau(i,j), \qquad (10)\\ \tau(i;k) &\triangleq t_{i\circ}(X^{(k)}) = u_i(X^{(k-1)}) = \text{length of } s_i(X^{(k)}),\\ \tau(i) &\triangleq \tau(i;n). \end{aligned}$$

The key point in checking these is that $s_i(X)$ is exactly the order in which the balls are drawn from U_i in the urn model. It should be noted that while $\tau(i,j)$ and $\tau(i)$ are parameters, $\tau(i,j;k)$ and $\tau(i;k)$ are random variables.

If $i \in C_y$ and $\tau(i) > 1$ let $U_i \backslash f_i$ represent the urn U_i with one ball labeled f_i removed. If $i = y$ or $\tau(i) = 1$, then $U_i \backslash f_i$ is taken to be the same as U_i (to avoid dealing with empty urns). Then for $\sigma \in C^{\tau(i)}$

$$(11)\quad G\{s_i(X) = \sigma\} = \begin{cases} H_{U_i \backslash f_i}\{\sigma\} & \text{if } i = y \text{ or } \tau(i) = 1, \\ H_{U_i \backslash f_i}\{\sigma^{(\tau(i)-1)}\} I\{\sigma_{\tau(i)} = f_i\} & \text{if } i \in C_y \text{ and } \tau(i) > 1. \end{cases}$$

For any $X, Y \in C^n$ it can be shown that $X = Y$ iff $X_1 = Y_1$ and $s_i(X) = S_i(Y)$ for all $i \in C$, so

$$(12)\qquad G(X) = I\{X_1 = x\} \prod_{i \in C} G\{s_i(X)\}.$$

5. Approximating extreme points. For every point in $\mathscr{E}_C^n$ we want to find a close point in $\mathscr{M}_C^n$. A first step is to find nearby points in $\mathscr{M}_C^n$ for every extreme point of $\mathscr{E}_C^n$. We will do even more by finding a point in $\mathscr{M}_C^n$ near any measure $G_{x, y, \tau, f}$. Equations (11) and (12) show that G is nearly a product of urn measures $H_{U_i \backslash f_i}$. Define the Markov chain $R_{\tilde{a}}$ by

$$a_{0i} = I\{i = x\},$$

$$(13)\qquad a_{ij} = \begin{cases} \dfrac{\tau(i, j)}{\tau(i)} & \text{if } i = y \text{ or } \tau(i) = 1, \\ \dfrac{\tau(i, j) - I\{j = f_i\}}{\tau(i) - 1} & \text{if } i \in C_y \text{ and } \tau(i) > 1. \end{cases}$$

This is exactly the Markov chain one would get by using the urn model with the glued balls removed (i.e., using the urns $U_i \backslash f_i$) and sampling with replacement. One could alternately write

$$(14)\qquad R_{\tilde{a}}^{(n)}\{X\} = I\{X_1 = u\} \prod_{i=1}^{c} \mathscr{M}_{U_i \backslash f_i}\{s_i(X)\}.$$

THEOREM 3. *For G and $R_{\tilde{a}}^{(n)}$ defined as above and $k \le n$*

$$\frac{1}{2}\|G^{(k)} - R_{\tilde{a}}^{(k)}\| \le E_{G^{(k)}}\left\{ \sum_{i \in C} \min\left[1, \sum_{\substack{j \in C \\ \tau(i, j) > 0}} \frac{\tau(i, j; k)}{\tau(i, j) - I\{j = f_i\}}\right]\right\},$$

where $E_{G^{(k)}}$ refers to the expectation under $G^{(k)}$.

PROOF. Let x, y, τ, f, $\tilde{a}$, and G be as before. Let $R = R_{\tilde{a}}$. For the duration of the proof, let the symbol G_i represent the distribution of $s_i(X) \in C^{\tau(i)}$ when X has distribution G given in (11), and let $R_i = M_{U_i \backslash f_i}$, so that (12) and (14) become

$$(15)\qquad \begin{aligned} G\{X\} &= I\{X_1 = x\} \prod_{i \in C} G_i\{s_i(X)\}, \\ R\{X\} &= I\{X_1 = x\} \prod_{i \in C} R_i\{s_i(X)\}. \end{aligned}$$

The marginal distributions are given by

$$G^{(k)}\{X^{(k)}\} = I\{X_1 = x\} \prod_{i \in C} G_i^{(\tau(i;\,k))}\{s_i(X^{(k)})\},$$
$$R^{(k)}\{X^{(k)}\} = I\{X_1 = x\} \prod_{i \in C} R_i^{(\tau(i;\,k))}\{s_i(X^{(k)})\}. \tag{16}$$

Note that even for a fixed k, the value $\tau(i;\,k)$ depends upon the random outcome of X.

Using the property given in (6) about the variation distance between two measures,

$$\begin{aligned}
\frac{1}{2}\|G^{(k)} - R^{(k)}\| &= \sum_{X \in C^k} [G^{(k)}(X) - R^{(k)}(X)]^+ \\
&= \sum_{X \in C^k} G^{(k)}\{X\}\left[1 - \frac{R^{(k)}\{X\}}{G^{(k)}\{X\}}\right]^+ \\
&= E_{G^{(k)}}\left\{\left[1 - \frac{R^{(k)}\{X\}}{G^{(k)}\{X\}}\right]^+\right\} \\
&= E_{G^{(k)}}\left\{\left[1 - \prod_{i \in C}\frac{R_i^{(\tau(i;\,k))}\{s_i(X)\}}{G_i^{(\tau(i;\,k))}\{s_i(X)\}}\right]^+\right\} \\
&\le E_{G^{(k)}}\left\{\sum_{i \in C}\left[1 - \frac{R_i^{(\tau(i;\,k))}\{s_i(X)\}}{G_i^{(\tau(i;\,k))}\{s_i(X)\}}\right]^+\right\}.
\end{aligned} \tag{17}$$

The term $[\cdots]^+$ in the final expression can be bounded by

$$\left[1 - \frac{R_i^{(\tau(i;\,k))}\{s_i(X)\}}{G_i^{(\tau(i;\,k))}\{s_i(X)\}}\right]^+ \le \min\left[1, \sum_{j \in C} \frac{\tau(i, j;\, k)}{\tau(i, j) - I\{j = f_i\}}\right]. \tag{18}$$

To verify this, first note that the l.h.s. cannot exceed 1, so the min[1,...] on the r.h.s. can be ignored.

We now consider three cases.

CASE 1. $i = y$ so that U_i has no glued ball.

Then $G_i = H_{U_i}$ and $R_i = M_{U_i}$ so by (5), the l.h.s. can be bounded by $\sum_{j=1}^{C} u_j(s_i(X^{(k)}))/\tau(i, j)$. Since f_i is not defined for $i = y$, the term $I\{j = f_i\}$ in (18) should be treated as equal to 0 for all j.

CASE 2. $i \in C_y$ and $\tau(i;\,k) < \tau(i)$ so that the glued ball in U_i is not drawn by $X^{(k)}$.

By (11), $G_i^{(\tau(i;\,k))} = H_{U_i \setminus f_i}^{(\tau(i;\,k))}$ as long as $\tau(i;\,k) < \tau(i)$ so (5) can be used once again with the fact that $U_i \setminus f_i$ contains $\tau(i, j) - I(j = f_i\}$ balls with label j for $j \in C$.

CASE 3. $i \in C_y$ and $\tau(i; k) = \tau(i)$ so that U_i is completely empty including the glued ball by the time $X^{(k)}$ has been drawn.

Since U_i is empty, $s_i(X^{(k)}) = s_i(X^{(n)})$ and $\tau(i, j; k) = \tau(i, j)$ and $\sum_{j \in C} \tau(i, j; k)/(\tau(i, j) - I\{j = f_i\})$ is bigger than 1, and hence is not a constraint.

These three cases are exhaustive, and (18) holds in each case, which when substituted in (17) provides the conclusion of the theorem. □

6. The stationarity constraint. The bound given in Theorem 3 is not enough to prove dFTME. We need a uniform bound which decreases to zero as n approaches infinity. On the other hand the bound in Theorem 3 can be very bad regardless of how large n is, as long as one urn U_i is small, i.e., $\tau(i)$ is small. There is no constraint to ensure that each urn gets larger as n increases, so that one urn may have just two balls regardless of n, and an urn with two balls is very different from i.i.d. sampling.

The problem is not because the inequality in Theorem 3 is crude, but instead is a fundamental problem related to the exceptions to dFTME found by Diaconis and Freedman (1980a). There are measures in $\mathscr{E}_C$ for which $\tau(i; n)$, the size of U_i, remains bounded even in the limit as n approaches infinity, and these measures are simply not in $\mathscr{M}_C$. Thus we cannot hope to get a uniform bound decreasing to zero in n.

To obtain their version of dFTME, Diaconis and Freedman (1980a) had to add the condition of recurrence. There is no finite dimensional analogue of recurrence, so instead we will impose the stricter condition of stationarity (which implies recurrence for infinite sequences), as was done for the original proof of dFTME by Freedman (1962). Stationarity is also defined in terms of measures on infinite sequences, but has a straightforward generalization for finite sequences described in Zaman (1983). A finite random sequence $X \in C^n$ is called stationary if the distribution of $(X_1, X_2, \ldots, X_{n-1})$ is the same as the distribution of $(X_2, \ldots, X_n)$. Among other things this is equivalent to the existence of an infinite stationary extension of X, i.e., a stationary sequence $Y_1, Y_2, \ldots,$ such that $Y^{(n)}$ has the same distribution as X.

The addition of the stationarity assumption gives a kind of "homogeneity" to the sequence X, so that the chance of observing an i to j transition is the same at any place along the sequence. This allows the conclusion that $\tau(i, j; k)/\tau(i, j; n)$ is somewhat like k/n using a property of stationarity given in Zaman (1984b), which in turn gets a uniform bound for Theorem 3.

7. dFTME (finite form under stationarity).

THEOREM 4. *For G and $R_{\tilde{a}}^{(n)}$ defined as in Theorem 3 and $k \le n$, if G is stationary*

$$\|G^{(k)} - R_{\tilde{a}}^{(k)}\| \le 2[c^2 + c + 1]\frac{k-1}{n-1}[1 + \log(n-2)].$$

The main tool for the proof is Theorem 7 in Zaman (1984b) which can be written as

THEOREM 5. *If P is a stationary measure on $\{0,1\}^n$ and $k \le n$,*

$$E_P\left\{\frac{X_1 + \cdots + X_k}{X_1 + \cdots X_n}\right\} \le \frac{k}{n}[1 + \log(n-1)].$$

PROOF OF THEOREM 4. From Theorem 3, we can write

$$\frac{1}{2}\|G^{(k)} - R_{\tilde{a}}^{(k)}\| \le \sum_{i \in C} \sum_{\substack{j \in C \\ j \ne f_i}} E_G\left\{\frac{\tau(i,j;k)}{\tau(i,j)}\right\} + \sum_{i \in C_y} E_G\left\{\frac{\tau(i,f_i;k)}{\max(1, \tau(i,f_i) - 1)}\right\}. \tag{19}$$

For any $i, j \in C$ construct a sequence $Y \in \{0,1\}^{n-1}$ by defining $Y_m = I\{(X_m, X_{m+1}) = (i,j)\}$. If the measure on the X's is stationary, then so are the y's and so Theorem 5 applies:

$$E_G\left\{\frac{\tau(i,j;k)}{\tau(i,j)}\right\} = E_G\left\{\frac{Y_1 + \cdots + Y_{k-1}}{Y_1 + \cdots + Y_{n-1}}\right\} \le \frac{k-1}{n-1}[1 + \log(n-2)]. \tag{20}$$

To deal with the $\tau(i, f_i) - 1$ in the denominator in (19), simply note that

$$\frac{\tau(i,f_i;k)}{\max(1,\tau(i,f_i)-1)} = \frac{\tau(i,f_i)}{\max(1,\tau(i,f_i)-1)} \frac{\tau(i,f_i;k)}{\tau(i,f_i)} \le 2\frac{\tau(i,f_i;k)}{\tau(i,f_i)} \tag{21}$$

because $x/\max(1, x-1)$ has a maximum value of 2 when $x = 2$.

Of the c^2 terms $i, j \in C$ in (19), the first $c^2 - (c-1)$ are bounded by (20), and the last $c - 1$ by (21), giving the bound

$$\frac{1}{2}\|G^{(k)} - R_{\tilde{a}}^{(k)}\| \le (c^2 + c - 1)\frac{k-1}{n-1}[1 + \log(n-2)]. \qquad \square$$

The statement of Theorem 4 gives a uniform bound for all extreme points of $\mathscr{M}_C^n$. Extending this to all of $\mathscr{M}_C^n$ is a simple matter, and proves the finite form of dFTME for stationary sequences:

THEOREM 6. *For every $P \in \mathscr{E}_C^n$ there is a $Q \in \mathscr{M}_C^n$ such that if $k \le n$ then*

$$\|P^{(k)} - Q^{(k)}\| \le 2(c^2 + c - 1)\frac{k-1}{n-1}[1 + \log(n-2)].$$

PROOF. Let $\Lambda = \{(x, y, t, f): x, y \in C,\ t \in T_{x,y},\ f \in F\}$ be the parameter space of the urn model. $P \in \mathscr{E}_C^n$ implies that $P = G_\mu$ for some probability μ on

Λ. For any $\lambda \in \Lambda$, let $a(\lambda)$ be the $\tilde{a}$ defined in (13), and let the measure $Q = R^{(n)}_{a(\mu)}$. Then

$$\begin{aligned}\|P^{(k)} - Q^{(k)}\| &= \sup_{B \in C^k} \left| \int P^{(k)}_{\lambda}(B) - R^{(k)}_{a(\lambda)}(B) \mu \, d\lambda \right| \\ &\leq \int \sup_{B \in C^k} \left| P^{(k)}_{\lambda}(B) - R^{(k)}_{a(\lambda)}(B) \right| \mu \, d\lambda \\ &= \int \| P^{(k)}_{\lambda} - R^{(k)}_{a(\lambda)} \| \mu \, d\lambda \\ &\leq 2(c^2 + c - 1) \frac{k-1}{n-1} [1 + \log(n-2)]. \qquad \square\end{aligned}$$

COROLLARY 7. *When C is finite and $P \in \mathscr{E}_C$, if P is stationary then $P \in \mathscr{M}_C$.*

Since this is a very weak form of dFTME, it is given only to illustrate that an infinite result is possible from the finite form. The proof relies on the compactness of the parameter space A of finite state Markov chains. It follows along the same lines as the extension of the finite dFT to its infinite version in Diaconis and Freedman (1980b) and so is omitted.

In analogy with the finite version of de Finetti's theorem, a c^2k/n type of bound could have been expected under the best of conditions. The $\log n$ term is somewhat surprising even though it causes no harm in the asymptotics. It is not clear whether that is simply an artifact of the method of proof, instead of a real feature of the rate.

An interesting way to avoid this $\log n$ term is to consider sequences in both directions. Let P be a measure on $X_{-n}, \dots, X_n$. Denote the middle portion $X_{-k}, \dots, X_k$ by $X^{(-k,k)}$ and so on. Then

COROLLARY 8. *If $P \in \mathscr{E}_C^{-n,n}$ then there is a $Q \in M_C^{-n,n}$ such that for $k \leq n$,*

$$\|P^{(-k,k)} - Q^{(-k,k)}\| \leq 4(c^2 + c - 1)k/(n + 1 - k).$$

PROOF. Theorem 5 (which is Theorem 7 in Zaman (1984b)) has another version for middle portions of stationary sequences. Using that, the proof follows exactly like the proof of Theorem 6. $\square$

REFERENCES

DE FINETTI, B. (1959). La probabilitá et la statistica nei con l'induzione, secundo 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.

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

DIACONIS, P. and FREEDMAN, D. (1980b). Finite exchangeable sequences. *Ann. Probab.* **8** 745–764.

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

ZAMAN, A. (1983). Stationarity on finite strings and shift register sequences. *Ann. Probab.* **11** 678–684.

ZAMAN, A. (1984a). Urn models for Markov exchangeability. *Ann. Probab.* **12** 223–229.

ZAMAN, A. (1984b). A non-clustering property of stationary sequences. *Ann. Probab.* **12** 193–203.

DEPARTMENT OF STATISTICS
FLORIDA STATE UNIVERSITY
TALLAHASSEE, FLORIDA 32306