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

by

Arif Zaman

Florida State University
Department of Statistics
Tallahassee, Florida 32306

FSU Statistics Report M689

November, 1984

Key Words and phrases: de Finetti's Theorem, Markov exchangeability, stationary processes, Markov chains

AMS(1980) Subject Classifications. Primary 60J05; secondary 60G10
ABSTRACT

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 $X_1, \ldots, X_n$ which is stationary and Markov exchangeable. It is shown that any portion of $k$ consecutive elements, say $X_1, \ldots, X_k$ for $k<n$, is nearly a mixture of Markov chains (the distance measured in the variation norm).
Introduction

In de Finetti (1959), (1974 pp. 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 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's 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
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 dFTIME under the assumption of stationarity (theorem 6). This final theorem is strong enough to imply the infinite form of dFTIME 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 $k \leq n$, the initial portion $(X_1, \ldots, X_k)$ 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 $C = \{1, 2, \ldots, c\}$. A Markov chain probability $P$ on $X \in C^n$ is parametrized by its transition probabilities $a_{ij} = P\{X_k = i \text{ and } X_{k+1} = j\}$ for $i, j \in C$ and $k = 1, \ldots, n-1$ and its initial state probabilities $a_{oi} = P\{X_1 = i\}$ 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 $R^{(n)}_{\tilde{a}}$. It is clear that a valid value for $\tilde{a}$ must lie in the set $A = \{\tilde{a}: \Sigma_{j \in C} a_{ij} = 1 \text{ for } i=0, 1, \ldots, c\}$.

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

The set of all mixtures of Markov chains of length $n$ with state space $C$ is denoted by $M^n_C = \{R^{(n)}_{\mu}: \mu \text{ is a probability measure on } \Lambda\}$. The set of mixtures of infinite length Markov chains, denoted by $M_C$, can be defined by $P \in M_C$ iff $p^{(n)} = R^{(n)}_{\mu}$ for all $n$ with the same mixing measure $\mu$. 
For $X \in \mathbb{C}^n$, let $t_{ij}(X)$ be the number of $i$ to $j$ transitions in $X$, i.e.

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

(1)

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

$$\mathbb{R}_{\mu}^{(n)}(X) = \int_a^{a_0} \prod_{i,j \in \mathbb{C}} t_{ij}(X) u(\text{d}a)$$

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

Let $E_C^n$ denote all measures on $\mathbb{C}^n$ with the above mentioned property, i.e.

$P \in E_C^n$ iff for all $X, Y \in \mathbb{C}^n$

(3)

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

The set $E_C^n$ is the set of all Markov exchangeable probabilities on $\mathbb{C}^n$, or as de Finetti calls them, partially exchangeable probabilities of the Markov type.

The set of infinite length Markov exchangeable sequences, denoted by $E_C$, can be defined by $P \in E_C$ iff equation 3 holds for all $P^{(n)}$ with $n < \infty$.

3. Urns and the finite de Finetti's Theorem.

The symbol $U = (u_1, \ldots, u_C)$ will be used to denote an urn containing a total of $u = \sum_{i=1}^C u_i$ balls, with $u_1$ of them labelled with a "1", ..., and $u_C$ of them labelled with a "c". $H_U$ denotes the measure on sequences in $\mathbb{C}^{u}$ 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 $X \in \mathbb{C}^n$, defined by

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

(4)

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

The following facts are borrowed from Diaconis and Freedman (1980b). Equation 12 from there implies that for any sequence $X \in C^k$ and any urn $U$ with $k \leq u$.

$$\left[ \frac{M^{(k)}_U(X)}{H^{(k)}_U(X)} \right]^+ \leq \frac{c}{\sum_{i=1}^k \frac{u_i(X)}{u_i}} ,$$

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

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

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

If $\Omega$ is finite and all subsets of $\Omega$ measurable, then an equivalent form is

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

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

**Theorem (dFT)**

For every exchangeable measure $P$ on $C^n$ there exists a mixture of i.i.d. measures, $Q$, such that

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

4. **Urn models for Markov exchangeability.**

Combining equations 1 and 4 for $X \in C^n$ gives

$$t_{i,*}(X) = u_i(X^{(n-1)})$$

so that

$$t_{i,*}(X) + I\{X_n = i\} = u_i(X) = t_{i,*}(X) + I\{X_1 = i\} .$$

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 $y$ are
\[ \tau^n_{x,y} = \{ t(x) : X \in C^n ; X_1 = u ; X_n = y \} \]

(9)

\[ = \{ t : t^{*}_i + I(y=i) = t^{*}_i + I(x=i) \} \]

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 \( \tau \in T_{x,y} \) 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 \( i \neq y, i \in C \).

Let \( U_1, \ldots, U_c \) be \( c \) separate urns with \( U_i = (\tau_{i1}, \ldots, \tau_{ic}) \) 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 \( X_1 = u \). For \( i = 1,\ldots,n-1 \) 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 \( \tau \), 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 \( \tau \in T_{x,y} \) there is a unique probability \( \mu \) on \( F \) such that \( G_{x,y,\tau,\mu} \) is an extreme point of the set of measures \( E^C_n \). Moreover any measure in \( E^C_n \) 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:
\[ X = 2 \ 3 \ 2 \ 3 \ 1 \ 2 \ 1 \ 1 \ 3 \]

\[ s_1(X) = 2 \ 1 \ 3 \]

\[ s_2(X) = 3 \ 3 \ 1 \]

\[ s_3(X) = 2 \ 1 \]

(s₁ stands for successors of i).

The following is a list of definitions (denoted by \( \Delta \)) 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)

\[ \tau(i,j) \ \Delta \ \tau_{ij} \]

\[ \tau(i,j;k) \ \Delta \ \tau_{ij}(x^{(k)}) = u_j(s_i(x^{(k)})) \]

\[ \tau(i,j;n) = \tau(i,j) \]

\[ \tau(i;k) \ \Delta \ \tau_i(x^{(k)}) = u_i(x^{(k-1)}) \]

\[ \tau(i) \ \Delta \ \tau(i;n) = \text{length of } s_i(x^{(k)}) \]

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

\[ G(s_i(X) = \delta) = \begin{cases} H_{U_i \setminus f_i} \{ \delta \} & \text{if } i = y \text{ or } \tau(i) = 1 \\ H_{U_i \setminus f_i} \{ \delta \} \tau(i - 1) \} \{ \delta \} & \text{if } i \in C_y \text{ and } \tau(i) > 1. \end{cases} \]

For \( X, Y \) two sequences in \( C^n \), \( \forall_i \), \( s_i(X) = s_i(Y) \) and \( X_1 = Y_1 \), iff \( X = Y \), so

\[ G(X) = I\{X_1 = x\} \prod_{i \in C} G(s_i(X)) \]
5. Approximating extreme points. For every point in $E^n_C$ we want to find a
close point in $M^n_C$. A first step is to find nearby points in $M^n_C$ for every extreme
point of $E^n_C$. We will do even more by finding a point in $M^n_C$ near any measure
$G_{x,y,\tau,f}$. Equations 11 and 12 show that $G$ is nearly a product of urn measures
$H_{U_1\setminus f_i}$. Define the Markov chain $R^{(n)}_\tilde{a}$ by

$$ a_{oi} = I\{i = x\} $$

(13)$$ a_{ij} = \frac{\tau(i,j)/\tau(i)}{[\tau(i,j) - I\{j = f_i\}]/[\tau(i) - 1]} \quad \text{if } i = y \text{ or } \tau(i) = 1 $$

$$ \text{if } i \in C_y \text{ and } \tau(i) > 1 $$

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

(14)$$ R^{(n)}_\tilde{a}(X) = I\{X_1 = u\} \prod_{i=1}^{c} \prod_{U_i \setminus f_i} \{s_i(X)\} $$

Theorem 3 For $G$ and $R^{(n)}_\tilde{a}$ defined as above and $k \leq n$

$$ \frac{1}{2} \|G^{(k)} - R^{(k)}_\tilde{a}\| \leq E_G(k) \left\{ \sum_{i \in C} \min_{j \in C} \frac{\tau(i,j;k)/(\tau(i,j) - I\{j = f_i\})}{\tau(i,j) > 0} \right\} $$

Proof: Let $x,y,\tau,f,\tilde{a}$ and $G$ be as before. Let $R = R^{(n)}_\tilde{a}$. 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 equation 11, and let $R_i = M_{U_i \setminus f_i}$, so that equations 12
and 14 become

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

(15)$$ R(X) = I\{X_1 = x\} \prod_{i \in C} R_i\{s_i(X)\} $$
The marginal distributions are given by
\[ G^{(k)}\{X(k)\} = I\{X_1 = x\} \prod_{i \in C} G_i^{(\tau(i;k))}\{s_i(X(\vec{k}))\} \]
(16)
\[ R^{(k)}\{X(k)\} = I\{X_1 = x\} \prod_{i \in C} R_i^{(\tau(i;k))}\{s_i(X(\vec{k}))\}. \]
Note that even for a fixed \( k \), the value \( \tau(i;k) \) depends upon the random outcome of \( X \).

Using the property given in equation 7 about the variation distance between two measures,
\[ \frac{1}{2} \| G^{(k)} - R^{(k)} \| = \sum_{\vec{x} \in C^k} \left| G^{(k)}(\vec{x}) - R^{(k)}(\vec{x}) \right|^+ \]
\[ = \sum_{\vec{x} \in C^k} G^{(k)}\{\vec{x}\} [1 - R^{(k)}\{\vec{x}\}/G^{(k)}\{\vec{x}\}]^+ \]
\[ = E_{G^{(k)}}[[1-R^{(k)}\{\vec{x}\}/G^{(k)}\{\vec{x}\}]^+] \]
\[ = E_{G^{(k)}}[[1 - \prod_{i \in C} R_i^{(\tau(i;k))}\{s_i(X)\}/G_i^{(\tau(i;k))}\{s_i(X)\}]^+] \]
\[ \leq E_{G^{(k)}}\left[ \sum_{i \in C} [1 - (R_i^{(\tau(i;k))}\{s_i(X)\}/G_i^{(\tau(i;k))}\{s_i(X)\})]^+ \right]. \]
(17)

The term \([...]^+\) in the final expression can be bounded by
\[ [1 - (R_i^{(\tau(i;k))}\{s_i(X)\}/G_i^{(\tau(i;k))}\{s_i(X)\})]^+ \leq \min[1, \sum_{j \in C} \tau(i,j;k)/(\tau(i,j)-I(j=f_i))]. \]
(18)

To verify this, first note that the LHS cannot exceed 1, so the \( \min[1,...] \) on the RHS can be ignored.

We now consider three cases.

Case 1: \( i=y \) so that \( U_i \) has no glued ball.

Then \( G_i = U_i \) and \( R_i = U_i \) so by equation 5, the LHS 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 equation 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 equation 11, \( G_i^{(\tau(i;k))} = U_i \setminus f_i \) as long as \( \tau(i;k) < \tau(i) \) so equation 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 equation 18 holds in each case, which when substituted in equation 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 \( E_C \) 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 \( M_C \). 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 \( X \in \mathbb{C}^n \) 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 \( R^{(n)}(\underline{a}) \) defined as in theorem 3 and \( k \leq n \), if \( G \) is stationary

\[
\| G^{(k)} - R^{(k)}(\underline{a}) \| \leq 2[C^2 + C + 1] \frac{k-1}{n-1} [1 + \log(n-2)]
\]

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 \( k \leq n \)

\[
E_{ \underline{p} } \left\{ \frac{X_1 + \ldots + X_k}{X_1 + \ldots + X_u} \right\} \leq \frac{k}{n} [1 + \log(n - 1)]
\]

**Proof of theorem 4:**

From theorem 3, we can write

\[
\| G^{(k)} - R^{(k)}(\underline{a}) \| \leq \sum_{i \in C} \sum_{j \in C, j \neq f_i} E_{G} \{ \tau(i,j;k)/\tau(i,j) \} + \sum_{i \in C \setminus Y} E_{G} \{ \tau(i,f_i;k)/\max(1,\tau(i,f_i) - 1) \}
\]
For any $i, j \in \mathcal{C}$ 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 $X$'s is stationary, then so are the $y$'s and so theorem 5 applies.

\begin{equation}
E_G \{\tau(i, j, k)\} = E \{\frac{Y_1 + \ldots + Y_{k-1}}{\tau(i, j)}\} \leq \frac{k-1}{n-1} [1 + \log(n-2)].
\end{equation}

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

\begin{equation}
\frac{\tau(i, f_i ; k)}{\max(1, \tau(i, f_i ; 1) - 1)} = \frac{\tau(i, f_i)}{\max(1, \tau(i, f_i ; 1) - 1)} = \frac{\tau(i, f_i ; k)}{\tau(i, f_i)} \leq 2 \frac{\tau(i, f_i ; k)}{\tau(i, f_i)}
\end{equation}

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

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

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

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

**Theorem 6**

For every $P \in \mathcal{E}^n_C$ there is a $Q \in \mathcal{M}^n_C$ such that if $k \leq n$ then

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

**Proof:**

Let $\Lambda = \{(x, y, t, f): x, y \in \mathcal{C}, t \in T_x, y, f \in F\}$ be the parameter space of the urn model. $P \in \mathcal{E}^n_C$ implies that $P = G_{\mu}$ for some probability $\mu$ on $\Lambda$. For any $\lambda \in \Lambda$, let $a(\lambda)$ be the $\tilde{a}$ defined in equation 13, and let the measure $Q = R^{(n)}_{a(\mu)}$.

Then

$$\| p^{(k)} - Q^{(k)} \| = \sup_{B \in \mathcal{C}^k} \left| \int p^{(k)}_{\lambda} (B) - R^{(k)}_{a(\lambda)} (B) \mu(d\lambda) \right|$$

$$\leq \int \sup_{B \in \mathcal{C}^k} \left| p^{(k)}_{\lambda} (B) - R^{(k)}_{a(\lambda)} (B) \right| \mu(d\lambda)$$

$$= \int \| p^{(k)}_{\lambda} - R^{(k)}_{a(\lambda)} \| \mu(d\lambda)$$

$$\leq 2(c^2 + c - 1) \frac{k-1}{n-1} [1 + \log(n-2)].$$

\(\square\)
Corollary 7

When C is finite and $P \in E_C$, if $P$ is stationary then $P \in M_C$.

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 $X_{-n}, \ldots, X_n$. Denote the middle portion $X_{-k}, \ldots, X_k$ by $X^{(-k,k)}$ and so on. Then

Corollary 8: If $P \in \xi_{C}^{-n,n}$ 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.
REFERENCES


