韓國 軍事運營分析 學會誌
第11卷 第1號, 1985. 6.

Sensitivity of Conditions for Lumping Finite Markov Chains

Suh, Moon Taek*

Abstract

Markov chains with large transition probability matrices occur in many applications such as manpowr models. Under certain conditions the state space of a stationary discrete parameter finite Markov chain may be partitioned into subsets, each of which may be treated as a single state of a smaller chain that retains the Markov property. Such a chain is said to be "lumpable" and the resulting lumped chain is a special case of more general functions of Markov chains.

There are several reasons why one might wish to lump. First, there may be analytical benefits, including relative simplicity of the reduced model and development of a new model which inherits known or assumed strong properties of the original model (the Markov property). Second, there may be statistical benefits, such as increased robustness of the smaller chain as well as improved estimates of transition probabilities. Finally, the identification of lumps may provide new insights about the process under investigation.

I. INTRODUCTION

Markov chains with large transition probability matrices occur in many application such as manpower models. Under certain conditions the state space of a stationary discrete parameter finite Markov chain may be partitioned into subsets, each of which may be treated as a single state of a smaller chain that retains the Markov property. Such a chain is said to be "lumpable" and the resulting lumped chain is a special case of more general functions of Markov chains.

Consider a Markov chain $\{X:t = 0,1,2, \ldots\}$ with finite state space $S = \{1,2, \ldots, n\}$, stationary

* Rok Army Force

transition probability matrix $P = [p_{ij}]$, and a priori distribution of "initial states", $p^0 = (p_1{}^0, p_2{}^0, ..., p_n{}^0)$. Let $\tilde{S}$ denote a nontrivial partition of S into $m < n$ "lumps", say $\tilde{S} = \{I(1), L(2), ..., L(m)\}$. If $\{X_t\}$ is lumpable with respect to $\tilde{S}$, denote by $\{\tilde{X}_t\}$ the lumped chain with state space $\tilde{S}$ and transition probability matrix $\tilde{P}$.

A well-known characterization [Ref. 2] is that $\{X_t\}$ is lumpable to $\{\tilde{X}_t\}$ if and only if there exist matrices A and B such that

$$BAPB = PB \tag{1.1}$$

where B consists of m nonzero orthogonal n-dimensional column vectors whose components are zeros or ones, and A is B' with rows normalized to probability vectors (i.e, $A = (B'B)^{-1} B'$). The positions of the 1's in each column of B correspond to states in S that together form a lump in $\tilde{S}$. It follows that if BAPB = PB is satisfied, then $\tilde{P} = APB$ as is shown in Chapter 2.

Many of the mathematical quantities associated with $\{X_t\}$ can be transformed directly to corresponding quantities for $\{\tilde{X}_t\}$, using the lumping matrix B. In Chapter 2, for example, we show that if an original Markov chain $\{X_t\}$ is lumpable to $\{\tilde{X}_t\}$ and $\{\tilde{X}_t\}$ is further lumpable to $\{\tilde{\tilde{X}}_t\}$, then $\{X_t\}$ is directly lumpable to $\{\tilde{\tilde{X}}_t\}$, and we give the lumping matrix for $\{\tilde{\tilde{X}}_t\}$ in terms of the underlying two lumpings.

There are several reasons why one might wish to lump [Ref. 1]. First, there may be analytical benefits, including relative simplicity of the reduced model and development of a new model which inherits known or assumed strong properties of the original model (the Markov property). Second, there may be statistical benefits, such as increased robustness of the smaller chain as well as improved estimates of transition probabilities. Finally, the identification of lumps may provide new insights about the process under investigation.

However, a problem that arises in connection with practical applications of Markov chain models is to determine whether the chain is lumpable. For cahins with large state spaces S, it is practically impossible to use an exhaustive search to determine whether lumpability conditions such as those given in equation (1.1) are met for some matrices B, because of the large number of ways partitioning S, i.e, the large number of candidate B matrices. For example, if S has 10 elements, there are 115,975 partitions of S.

Another problem is to estimate the matrix $P = \{p_{ij}\}$ of transition probabilities and to find bounds on Δ, the largest error of $\rho_{ij} - p_{ij}$ for all i and j. We shall investigate the sensitivity of the lumping conditions in equation (1.1) for varying Δ. If $\{X_t\}$ is lumpable with lumping matrix B, is condition (1.1) satisfied with P replaced by the estimate $\hat{P}$?

This thesis will attempt to examine the sensitivity of the lumping conditions based on reasonable estimation errors Δ when P is not known and must use estimated by $\hat{P}$. We describe these facts about lumpability using eigenvalues and eigenvectors, including the theorem mentioned by D.R. Barr and M.U. Thomas [Ref. 3]. We do not review elementary concepts of Markov chains here; the reader may

wish to consult [Ref. 2] and [Ref. 4] for review of basic facts and specific terminologies such as lumpability, regular Markov chain, etc.

II. THEORY OF LUMPABILITY

This chapter will cover general facts about lumping such as, conditions for lumping, the number of partitions possible for any given size of state space S, and theorems associated with eigenvector conditions for Markov chain lumpability.

A. Conditions for Lumping

Consider a Markov chain $\{x:t = 0,1,2, ...\}$ with finite state space $S = \{1,2, ..., n\}$, stationary transition probability matrix $P = \{p_{ij}\}$, and a priori distribution of "initial states", $p^0 = (p_1{}^0, p_2{}^0, ..., p_n{}^0)$. Let $\tilde{S}$ denote a nontrivial partition of S into $m < n$ "lumps", that is $\tilde{S} = \{L(1), L(2), ..., l(m)\}$. If $\{X_t\}$ is lumpable with respect to $\tilde{S}$, denote by $\{\tilde{X}_t\}$ the lumped chain with state space $\tilde{S}$ and transition probability matrix $\tilde{p}$.

We now show that if the condition (1,1) for lumpability with respect to the lumping matrix B,

$$BAPB = PB \tag{2.1}$$

is satisfied, then the lumped transition matrix $\tilde{P}$ is given by

$$\tilde{P} = APB \tag{2.2}$$

Proof. $\tilde{P}_{ij}$ is the sum $\sum_{k \in L(j)} p_{ik}$, where L(j) is the partition subset containing $j \in S$ and i is any element of L(i). By the lumpabilility condition, this value is the same for any $i \in L(i)$. But the product PB sums the columns of P in accordance with the partition subsets indicated by the columns of B. Hence, PB is an n x m matrix with rows repeated in accordance with the partition setsL(1), L(2), ... , L(m); the effect of pre-multiplying by $A = (B'B)^{-1}B'$ is to "average" these common rows yielding an mx m matrix $\tilde{P}$ without the repeated rows. But such "averages" are just the common rows being averaged. Hence, $\tilde{P} = APB$ is the m x m transition matrix of the lumped chain with state space {L(1), L(2), ..., L(m)}.

Example 1. Consider a transition probability matrix P with 4 states which can be partitioned into $\tilde{S} = \{\{1\}, \{2,3\}, \{4\}\} = \{L(1), L(2), L(3)\}$. Let

$$P = \begin{bmatrix} 1/4 & 1/16 & 3/16 & 1/2 \\ 0 & 1/12 & 1/12 & 5/6 \\ 0 & 1/12 & 1/12 & 5/6 \\ 7/8 & 1/32 & 3/32 & 0 \end{bmatrix}$$

Then

$$B = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad A = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0.5 & 0.5 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

We know equation (2.1) is satisfied with partitioning $\tilde{S} = \{I(1), L(2), L(3)\}$. Thus, the lumped transition matrix is

$$APB = \tilde{P} = \begin{bmatrix} 0.25 & 0.25 & 0.5 \\ 0 & 0.167 & 0.833 \\ 0.875 & 0.125 & 0 \end{bmatrix}$$

Many of the mathematical quantities associated with $\{X_t\}$ can be transformed directly to corresponding quantities for $\{X_t\}$, using A and B of equation (2.1). For example, since AB is the m-dimensional identity matrix, it follows that for s a positive integer,

$$(\tilde{P})^s = (APB)^s = A(PB)\,A\,(PB) \ldots \; A\,(PB) = AP^s B \tag{2.3}$$

We now show that $(\tilde{P})^s = AP^sB$

$$\begin{aligned} (\tilde{P})^s &= (APB)\,(APB)\,(APB) \ldots (APB) \\ &= AP\,(BAPB)\,(APB) \ldots (APB) \\ &= AP\,(PB)\,(APB) \ldots (APB) \\ &= AP^2\,(BAPB) \ldots (APB) \\ &= AP^2PB \ldots (APB) \\ &= \ldots \\ &= AP^s\,B. \end{aligned}$$

But $AP^s\,B = \tilde{P}^s$, since

$$\begin{aligned} &BAPB = PB \\ &PBAPB = P^2B \\ &BAPBAPB = P^2B \\ &BAP^2B = P^2B \\ &\ldots\ldots\ldots \\ &BAP^sB = P^sB, \end{aligned}$$

so P^s is lumpable with the same matrix B and $\tilde{P}^s = AP^sB$. This implies in turn that if $\{X_t\}$ has steady state distribution π, then $\{X_t\}$ has steady state distribution $\tilde{\pi} = \pi B$.

Theorem 1. The steady state distribution $\tilde{\pi}$ of the lumped chain $\{X_t\}$ is πB where $\pi = \pi P$.

Proof.

$$\begin{aligned} \pi B &= (\pi P)B \\ &= \pi B\,(APB) \\ &= (\pi B)\tilde{P} \end{aligned}$$

Therefore, $\tilde{\pi} = \pi B$.

Similarly, the a priori distribution $\tilde{P^0}$ of the initial state of the lumped chain corresponding to that of the original chain P^0, is given by $\tilde{P^0} = P^0B$, since by equations (2.1) and (2.3),

$$\begin{aligned} P^0P^sB &= P^0P \ \ldots \ PB \\ &= P^0P \ \ldots \ PBAPB \\ &= P^0P \ \ldots \ PBP \\ &= \ \ldots \\ &= P^0BP^s. \end{aligned}$$

Note that P^0P^sB is the distribution of lumped states occupied by the lumped chain after s transitions. Since this equals $P^0B\tilde{P}^s = \tilde{P}^0\tilde{P}^s$, it follows that $\tilde{P^0} = P^0B$.

B. Partitions of a Set of States

The matrix B consists of m nonzero orthogonal n-dimensional column vectors whose components are zeros and ones which determine a specific partition of $S = \{1,2, \ldots, n\}$. Example 1 illustrates this, where the state space $S = \{1,2,3,4\}$ is partitioned into

$\tilde{S} = \{[1], \{2,3\}, \{4\}\} = \{L(1), L(2), L(3)\}$, and

$$B = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Permutations of these columns give a matrix which also lumps $\{X_t\}$. In order to see this, let B* be B with columns permuted in some order. Then $B^* = BI^*$, where I is the identity matrix with its columns permuted in the same order. Now if BAPB = PB, then

$$\begin{aligned} B^*A^*PB^* &= B^*(B^{*\prime}B^*)^{-1}B^{*\prime}PB^* \\ &= BI^*(I^{*\prime}B\,BI^*)^{-1}I^{*\prime}B'PBI^* \\ &= BI^*(I^*)^{-1}(B'B)^{-1}(I^{*\prime})^{-1}I^{*\prime}B'PBI^* \\ &= B(B'B^{-1})B'PBI^* \\ &= BAPBI^* \\ &= PBI^* \\ &= PB^*, \end{aligned}$$

so it follows that $\{X_t\}$ is also lumpable with respect to the matrix B^*.

Now, how many candidate lumping matrices are there? This would be the number of partitions of S. {Ref. 5} gives a recursion relation for the number A_N of ways of partitioning a set S = {1,2, ..., N}:

$$A_N = \sum_{k=0}^{N-1} \binom{N-1}{k} A_k \qquad (N \geq 1, A_0 = 1) \tag{2.4}$$

From this relation we find $A_1 = 1$, $A_2 = 2$, $A_3 = 5$, $A_4 = 15$, etc. The sizes of the entries in Table 1 show that it would be impossible to use a trial and error approach to finding lumping matrices B for lumping a chain with larger state spaces, say with 10 or more elements. Values of A_N for larger N are shown in Table 1.

Table 1. Partitions of a Set of N States

N	Partitions	N	Partitions
5	52	20	5.172415×10^{13}
6	203	30	$8.467490145 \times 10^{23}$
7	877	40	$1.574505884 \times 10^{35}$
8	4,140	50	$1.857242688 \times 10^{47}$
9	21,147	60	$9.769393075 \times 10^{59}$
10	115,975	70	$1.80750039 \times 10^{73}$

It is of interest to be able to systematically prescribe alternative lumpings by generating matrices B for a given transition matrix P, using some method other than trial and error. In the next section, we describe an approach to finding B matrices using the eigenvalues and eigenvectors of P.

C. An Eigenvector Condition for Markov Chain Lumpability

Many problems in science and mathematics deal with a linear operator $T: V \rightarrow V$, and it is of importance to determine these scalars for which the equation $Tx = \lambda x$ has nonzero solutions x. In this section we discuss this problem and its relationship with finding matrices B.

Theorem 2. The value 1 is always an eigenvalue for any Markov chain transition probability matrix.

Proof. Let P be any n x n transiton probability matrix of $\{X_t\}$, x be a left eigenvector in R^n, and λ be the corresponding eigenvalue of P. Then $xP = x\lambda$ which is equivalent to

$$x(P - \lambda I) = 0 \tag{2.5}$$

For λ to be an eigenvalue, there must be a nonzero solution x of equation (2.5). Equation (2.5) will have a nonzero solution if and only if

$$\det (P - \lambda I) = 0 \tag{2.6}$$

This is called the characteristic equation. To show that $\lambda = 1$ always satisfies equation (2.6), we need only show that the columns of the matrix in equation (2.6) are linearly dependent. Note that

$$(P - I) = \begin{bmatrix} P & P \ldots P & 1 & 0 \ldots 0 \\ P & P \ldots P & 0 & 1 \ldots 0 \\ p & \cdots\cdots p & 0 & 0 \cdots 1 \end{bmatrix}$$

$$= \begin{bmatrix} P-1 & P \ldots P \ldots p \\ p & p-1 \ldots p \ldots p \\ \cdots\cdots\cdots & \\ p & p \ldots\ldots p-1 \end{bmatrix} \tag{2.7}$$

Since $\sum_{j=1}^{n} p_{ij} = 1$ for Markov chains, it follows that the rows in equation (2.7) sum to zero, so the determinant in equation (2.6) is zero with $\lambda = 1$. It follows that $\lambda = 1$ is an eigenvalue of the Markov chain $\{X_t\}$. We'd next like to see properties of eigenvectors corresponding to the eigenvalue $\lambda = 1$.

Theorem 3. For any regular Markov chain, components of the eigenvector corresponding to $\lambda = 1$ are proportional to the steady state distribution of $\{X_t\}$.

Proof. Let x be a left eigenvector of P, and λ be the corresponding engenvalue of P, such that $xP = x\lambda$, and assume $\sum_{i=1}^{n} x_i = 1$. For given $\lambda = 1$, $xP = x$. The steady state distribution of $\{X_t\}$ is unique [Ref. 4]. Therefore, x must be the steady state distribution π since $\sum_{i=1}^{n} x_i = 1$.

The following example demonstrates Theorem 3.

Example 2. Let

$$P = \begin{bmatrix} 1/4 & 1/16 & 3/16 & 1/2 \\ 0 & 1/12 & 1/12 & 5/6 \\ 0 & 1/12 & 1/12 & 5/6 \\ 7/8 & 1/32 & 3/32 & 0 \end{bmatrix}$$

The eigenvectors corresponding to the eigenvalues of P are displayed as column vectors below:

$$\begin{array}{ll} \text{Eigenvalues} : & \begin{bmatrix} 1 & 0 & -0.25 & -0.3333 \end{bmatrix} \\ \text{Eigenvectors} : & \begin{bmatrix} 0.7367 & 0 & 10.5 & 0.8247 \\ 0.09209 & -0.7201 & -0.375 & -0.03436 \\ 0.2236 & 0.7201 & -4.125 & -0.2405 \\ 0.6315 & 0 & -6 & -0.5498 \end{bmatrix} \end{array}$$

Note that $\pi = (\pi_1, \pi_2, \pi_3, \pi_4)$

$= (0.4375, 0.0547, 0.1328, 0.375),$

where $\pi_1 = \dfrac{0.7367}{0.7367 + 0.09209 + 0.2236 + 0.6315}$, etc.

Theorem 4. Eigenvectors corresponding to eigenvalues other than 1 are orthogonal to e = (1,1, ..., 1).

Proof. $xe' = x(Pe') = (xP)e' = x\lambda e'$. Therefore, xe' must be zero for $\lambda \neq 1$.

We are also interested in finding the relationship between eigenvalues of P and those of lumped transition probability matrix $\widetilde{P}$, where $\widetilde{P} = APB$ as described above.

Theorem 5. Suppose $\{X_t\}$ *with transition matrix* P *is lumpable to* $\{\widetilde{X}_t\}$ with transition matrix $\widetilde{P}$. The eigenvalues of $\widetilde{P}$ are eigenvalues of P.

Proof. Let $\alpha(\lambda) = 0$ be the (n^{th} degree) characteristic equation of P. By the Cayley – Hamilton theorem [Ref. 6],

$$\alpha(P) = \alpha_n P^n + \alpha_{n-1} P^{n-1} + \ldots + \alpha_1 P + \alpha I = 0,$$

which together with equation (2.3) implies

$$\begin{aligned} A\alpha(P)B &= \alpha_n \widetilde{P^n} + \alpha_{n-1} \widetilde{P^{n-1}} + \ldots + \alpha I \\ &= \alpha(\widetilde{P}) \\ &= 0 \end{aligned}$$

Since $\widetilde{P}$ satisfies P's characteristic equation and since eigenvalues of $\alpha(\widetilde{P})$ are of the form $\alpha(\widetilde{\lambda})$, it follows that $\alpha(\widetilde{\lambda}) = 0$. Thus all eigenvalues $\widetilde{\lambda}$ of $\widetilde{P}$ are also eigenvalues of P.

We next examine the eigenvectors of P and $\widetilde{P}$, with the aim of identifying lumpings of $\{X_t\}$ directly in terms of the eigenvectors of P. We have seen that $\widetilde{p^0}$ is obtained directly as p^0B; a similar relationship holds with eigenvectors of P.

Theorem 6. Suppose x is a left eigenvector of P corresponding to eigenvalue λ, and suppose $\{X_t\}$ *is lumpable to a chain with transition matrix* $\widetilde{P} = APB$. Then xB satisifes the equation $(xB)\widetilde{P} = (xB)\lambda$.

Proof. By equation (2.1), $(xB)\widetilde{P} = xBAPB = xPB$. But $xP = x\lambda$, and the result follows.

We note that xB is not necessarily an eigenvector of $\widetilde{P}$ because it may be zero. In fact, it easily follows that $xB = 0$ if λ is not an eigenvalue of $\widetilde{P}$. But xB may be null even if λ is an eigenvalue of $\widetilde{P}$, in cases of where λ is a repeated eigenvalue of P more times than of $\widetilde{P}$.

[Ref. 7] pointed out some other useful properties associated with eigenvalues and eigenvectors such as: 1) if the matrix P is symmetric, then eigenvalues are real and eigenvectors are different for repeated eigenvalues, 2) if the matrix is not symmetric, then the eigenvectors are the same for repeated eigenvalues.

Theorem 7. If $\{X_t\}$ with transition matrix P is lumpable to $\{\widetilde{X}_t\}$ with transition matrix $\widetilde{P}$, and

$\{\tilde{X}_t\}$ is lumpable to $\{\tilde{\tilde{X}}_t\}$ with transiton matrix $\tilde{\tilde{P}}$, then $\{X_t\}$ is directly lumpable to $\{\tilde{\tilde{\tilde{X}}}_t\}$ where $\{\tilde{\tilde{\tilde{X}}}_t\}$ is the lumped chain of $\{X_t\}$.

Proof. Let $\{X_t\}$ be lumpable to $\{\tilde{X}_t\}$, and $\{\tilde{X}_t\}$ be lumpable to $\{\tilde{\tilde{X}}_t\}$ by matrices B_1 and B_2, where B_1 and B_2 are lumping matrices in which the dimension n x m of B_1 is greater than that of B_2. By equation (2.1), $\tilde{P} = A_1 P B_1$ and $\tilde{\tilde{P}} = A_2 \tilde{P} B_2$. Thus,

$$\tilde{\tilde{P}} = A_2 \tilde{P} B_2 = A_2 (A_1 P B_1) B_2 = (A_2 A_1) P (B_1 B_2)$$

To see that $B_1 B_2$ is a lumping matrix and $A_2 \cdot A_1$ is of the required form, we need to show that $(A_2 \cdot A_1), (B_1 \cdot B_2)$ is the identity matrix as mentioned in Section A. But

$$(A_2 \bullet A_1) \bullet (B_1 \bullet B_2) = A_2 \bullet (A_1 \bullet B_1) \bullet B_2 = A_2 \bullet I \bullet B_2 = A_2 \bullet B_2 = I.$$

Also, note that $B_1 \bullet B_2$ is B_1 lumped by B_2, so $B_1 \bullet B_2$ has columns of the required form. Therefore, $\{X_t\}$ is directly lumpable to $\{\tilde{\tilde{\tilde{X}}}_t\}$, by the lumping matrix $B = B_1 \bullet B_2$.

Example 3. Consider a Markov chain with 5 states, and transition probability matrix.

$$P = \begin{bmatrix} 0.3 & 0.1 & 0.2 & 0.1 & 0.3 \\ 0.1 & 0.3 & 0.1 & 0.3 & 0.2 \\ 0.5 & 0.1 & 0 & 0.1 & 0.3 \\ 0.1 & 0.5 & 0.2 & 0.1 & 0.1 \\ 0.5 & 0 & 0.1 & 0.2 & 0.2 \end{bmatrix}$$

First, consider $S = \{1,2,3,4,5\}$ which can be partitioned to $\tilde{S} = \{\{1\}, \{2,4\}, \{3,5\}\} = \{L(1), L(2), L(3)\}$. The corresponding lumping matrices are

$$B_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \text{ and } A_1 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0.5 & 0 & 0.5 & 0 \\ 0 & 0 & 0.5 & 0 & 0.5 \end{bmatrix},$$

and the lumped transition probability matrix is

$$\tilde{P} = A_1 P B_1 = \begin{bmatrix} 0.3 & 0.2 & 0.5 \\ 0.1 & 0.6 & 0.3 \\ 0.5 & 0.2 & 0.3 \end{bmatrix}.$$

Secondly, consider $\tilde{S}$ with 3 states which can be partitioned to $\tilde{\tilde{S}} = \{\{1.3\}, \{2\}\} = \{L'(1), L'(2)\}$ with matrices

$$B_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \text{ and } A_2 = \begin{bmatrix} 0.5 & 0 & 0.5 \\ 0 & 1 & 0 \end{bmatrix}$$

The corresponding lumped transition matrix is

$$\tilde{\tilde{P}} = A_2 \tilde{P} B_2 = \begin{bmatrix} 0.8 & 0.2 \\ 0.4 & 0.6 \end{bmatrix}$$

Finally, consider lumping the transiton probability matrix directly. For partitioning,

$$\tilde{\tilde{S}} = \{\{1,3,5\}, \{2,4\}\} = \{L''(1), L''(2)\}, \text{ and}$$

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} = B_1 \cdot B_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad A_2 \cdot A_1 = \begin{bmatrix} 1/3 & 0 & 1/3 & 0 & 1/3 \\ 0 & 1/2 & 0 & 1/2 & 0 \end{bmatrix},$$

and the directly lumped transiton probability matrix is

$$\tilde{\tilde{P}} = \begin{bmatrix} 0.8 & 0.2 \\ 0.4 & 0.6 \end{bmatrix}$$

Theorem 7 shows that lumping is "transitive", in the following sense. Define two transition matrices P and Q to be equivalent, ($P \doteq Q$), if $Q = \tilde{P}$ for a lumping matrix B whose columns are those of the identity matrix, in some permuted order. (Thus the chain $\{X_t\}$ and $\{Y_t\}$ differ only in the labels associated with their states). *Define a relation* "$\preccurlyeq$" between transition matrices as follows: $Q \preccurlyeq P$ if and only if $Q = \hat{P}$ for some lumping matrix B. Then theorem 7 shows that $Q \preccurlyeq P$, $R \preccurlyeq Q \Rightarrow R \preccurlyeq P$. This relation "$\preccurlyeq$" is reflexive, since $Q \preccurlyeq Q$ using the lumping matrix I (identity). Finally, "$\preccurlyeq$" is antisymmetric since $Q \preccurlyeq P$ and $P \preccurlyeq Q \Rightarrow P \doteq Q$. Thus, the set if all transition probability matrices is patially ordered by the "lumping" partial order, "$\preccurlyeq$".

III. BOUNDS ON THE LARGEST ERROR, Δ, IN P

In this chapter we consider three procedures to find bounds on Δ. First, we use the central limit theorem for given i and j. Secondly, we use a binomial approximation on the basis of the first procedure. Finally, we get the largest error Δ, using the asymptotic extreme value distribution. These three approximations are only designed to give a rough idea of the relationships between Δ and the number M of elements in the state space, the total number of observed transitons K..,and the probability α.

A. Approach Using Central Limit Theorem

We are interested in the sizes of the errors between the estimate $\hat{P}$ and the unknown P, where P is the

transition probability matrix of $\{X_t\}$. We assume the transition probability matrix P is of size M x M.

Let K_{ij} be the number of observed transitions of $\{X_t\}$ from state i to state j, and let $K_{i\cdot}$ be the number of observed transitions from state i. Similarly K·j is the number of observed transitions into state j.

Let p_{ij} be an unknown transition probability from state i to state j and $\hat{p}_{ij}$ be an estimate of p_{ij} based on K..observed transitions. Then the usual estimate $\hat{p}_{ij}$ of p_{ij} is the ratio of Kij to $K_{i\cdot}$. Now, as a rough approximation, imagine that Ki. is fixed, and the number of transitions from state i to state j, K_{ij}, is Binomial $(K_{i\cdot}, p_{ij})$. Then by the central limit theorem,

$$\hat{P}_{ij} \text{ is approximate by Normal } [p_{ij}, \frac{K_{i\cdot}\, p_{ij}\,(1-p_{ij})}{(K_{i\cdot})^2}$$

since

$$E[\hat{p}_{ij}] = E[\frac{K_{ij}}{K_{i\cdot}}] \approx p_{ij}, \text{ and}$$

$$\text{Var}[\hat{p}_{ij}] = \text{Var}[\frac{k_{ij}}{k_{i\cdot}}]$$

$$\approx \frac{\text{Var}[K_{ij}]}{(K_{i\cdot})^2}$$

$$\frac{K_{i\cdot}\, p_{ij}\,(1-p_{ij})}{(K_{i\cdot})^2} \tag{3.1}$$

We want to find a bound Δ on the estimation error $|\hat{p}_{ij} - p_{ij}|$ which occurs with probability at least α ; that is, the largest Δ for which

$$P[|\hat{p}_{ij} - p_{ij}| \geq \Delta] \geq \alpha.$$

Now

$$P[|\hat{p}_{ij} - p_{ij}| \geq \Delta] = P\left[\left|\frac{p_{ij} - p_{ij}}{\sqrt{\frac{K_{i\cdot} p_{ij}(1-p_{ij})}{(K_{i\cdot})^2}}}\right| \geq \frac{\Delta}{\sqrt{\frac{K_{i\cdot}\, p_{ij}\,(1-p_{ij})}{(K_{i\cdot})^2}}}\right] . \tag{3.2}$$

Let $Z = \dfrac{\hat{p}_{ij} - p_{ij}}{\sqrt{\dfrac{k_{i\cdot}\, p_{ij}\,(1-p_{ij})}{(K_{i\cdot})^2}}}$, then

Z is approximate by standard Normal. Rewrite equation (3.2) as

$$P\left[\,|Z| \geq \frac{\Delta}{\sqrt{\frac{K_{i\cdot}\, p_{ij}(1-p_{ij})}{(K_{i\cdot})^2}}}\,\right] \geq \alpha \;;\; 0 < \alpha < 1.$$

Equation (3.2) is approximately

$$P\left[\,Z \geq \frac{\Delta}{\sqrt{\frac{K_{i\cdot}\, p_{ij}(1-p_{ij})}{(K_{i\cdot})^2}}}\,\right] \geq \frac{\alpha}{2}$$

since the Normal distribution is symmetric. Solving for Δ, we have

$$\Delta \leq N^{-1}\left(1 - \frac{\alpha}{2}\right) \sqrt{p_{ij}(1 - p_{ij})}\; \frac{1}{\sqrt{K_{i\cdot}}}$$

where $N^{-1}\left(1 - \frac{\alpha}{2}\right)$ is the $\left(1 - \frac{\alpha}{2}\right)^{th}$ quantile of the standard Normal distribution. Suppose the steady state probability π_i of state i is $\frac{1}{M}$ based on the equally likely case, and suppose the worst case in which

$$\sqrt{p_{ij}(1 - p_{ij})} = \frac{1}{2}$$

Then an approximate value for Δ is given by

$$\begin{aligned}
\Delta &= N^{-1}\left(1 - \frac{\alpha}{2}\right)(0.5)\; \frac{1}{\sqrt{K_{i\cdot}}} \\
&= N^{-1}\left(1 - \frac{\alpha}{2}\right)(0.5)\; \frac{1}{\sqrt{\pi_i K_{\cdot\cdot}}} \\
&= N^{-1}\left(1 - \frac{\alpha}{2}\right)(0.5)\; \frac{1}{\sqrt{\frac{K_{\cdot\cdot}}{M}}} \\
&= N^{-1}\left(1 - \frac{\alpha}{2}\right)(0.5)\; \sqrt{\frac{M}{K_{\cdot\cdot}}}. \qquad (3.3)
\end{aligned}$$

Equation (3.3) concerns the eror $|\hat{p}_{ij} - p_{ij}|$ for fixed i and j. We'd now like to find an error bound Δ overall i and j. That is, we wish to find the largest Δ for which

$$P\left[\,|\hat{p}_{ij} - p_{ij}| \geq \Delta \text{ for some i and j}\right] \geq \alpha,$$

which is roughly the same as

$$P\,[\,\hat{p}_{ij} - p_{ij} \geq \Delta \text{ for some i and j}\,] \geq \frac{\alpha}{2} \tag{3.4}$$

We apply the binomial approximation in equation (3.4), so that

$$\begin{aligned} P\,[\,\hat{p}_{ij} - p_{ij} \geq \Delta \text{ for some i and j}\,] \\ &= 1 - P\,[\hat{p}_{ij} - p_{ij} < \Delta \text{ for all i and j}] \\ &= 1 - (1 - \frac{\alpha}{2})^{M^2} \quad . \end{aligned} \tag{3.5}$$

Let $1 - (1 - \frac{\alpha}{2})^{M^2} = \beta$ for some $0 < \beta < 1$. Solve for α, which gives

$$\alpha = 2 - 2\sqrt[M^2]{1-\beta} \quad . \tag{3.6}$$

Substitute the value of α in equation (3.6) into equation (3.3). Finally we get the approximate bound Δ for all i and j:

$$\Delta \approx N^{-1}\left(\sqrt[M^2]{1-\beta}\right)\,(0.5)\sqrt{\frac{M}{K..}} \tag{3.7}$$

Equation (3.7) gives an approximate expression for Δ, using binomial approximation.

B. Approach Using Order Statistics

Assume $Z_1, Z_2, \ldots, Z_M$ are independent continuous random variables, each with density function $f_Z(z)$ and distribution $F_Z(z)$. Now let $Z_{(1)}, Z_{(2)}, \ldots, Z_{(M)}$ denote their ordered values, from smallest $Z_{(1)}$ to largest $Z_{(M)}$; these are called the order statistics of $Z_1, Z_2, \ldots, Z_M$. We now consider the probability law for $Z_{(M)}$ [Ref. 8], the largest or maximum value.

The event $[Z_{(M)} \leq z]$ occurs if and only if the event $[Z_1 \leq z, Z_2 \leq z, \ldots, Z_M \leq z]$ occurs, since if the largest Z is smaller than z, all M of the random variables must be smaller than z, where z is any fixed real number. The distribution function for $Z_{(M)}$ is

$$\begin{aligned} F_{Z_{(M)}}(z) &= P\,[Z_{(M)} \leq z\,] \\ &= P\,[Z_1 \leq z, Z_2 \leq z, \ldots, Z_M \leq z] \\ &= P\,[Z_1 \leq z]\;P\,[Z_2 \leq z]\;\ldots P\,[Z_M \leq z] \end{aligned}$$

since $Z_1, Z_2, \ldots, Z_M$ are assumed independent. But each of $Z_1, Z_2, \ldots, Z_M$ has the same distribution $F_Z(z)$. So

$$F_{Z_{(M)}}(z) = [F_Z(z)]^M$$

The density function for Z then is

$$f_{Z_{(M)}}(z) = \frac{d}{dz} F_{Z_{(M)}}(z)$$

$$= \frac{d}{dz}[F_Z(z)]^M$$

$$= M\,[F_Z(z)]^{M-1} f_Z(z)$$

where $f_Z(z) = \dfrac{dF_Z(z)}{d_z}$

Consider the limiting distribution function of the maximum $Z_{(M)}$ as M tends to infinity. [Ref. 9, 10] show this distribution is

$$\lim_{M\to\infty} [F_Z(z)]^M = e^{-e^{-\sqrt{2\log M}\ (z\ -\ \sqrt{2\log M})}} \tag{3.8}$$

if $Z_1, Z_2, \ldots, Z_M$ is a random sample from standard Normal population. We want find a bound Δ on the largest of M^2 errors between estimates in $\hat{P}$ and the unknown components of P. The random variables $\hat{p}_{ij} - p_{ij}$ are very roughly Normal with mean 0 and variance $\frac{M}{4K..}$ which is derived from equation (3.1) for i, j = 1,2, ..., M. Recall that K.. is the total number of transitions observed.

Let $X_\ell = \hat{p}_{ij} - p_{ij}$ where $\ell = 1,2, \ldots, M^2$. Then we know the random variable X is approximately equal to $\frac{1}{2}\sqrt{\frac{M}{K..}}\,Z$, where Z has a standard Normal distribution. Let the random variable $X_{(M^2)}$ be equal to $\max |\hat{p}_{ij} - p_{ij}|$. Then

$$\lim_{M\to\infty} P\,[X_{(M^2)} \leq \Delta\,] = \lim_{M\to\infty} P\,[\max |\hat{p}_{ij} - p_{ij}| \leq \Delta]$$

$$= P\,\left[\text{smallest of } \hat{p}_{ij} - p_{ij} \geq -\Delta \text{ and largest of } \hat{p}_{ij} - p_{ij} \leq \Delta\right]$$

Now $X_{(1)}$ and $X_{(M^2)}$ are asymptotically independent, so for large M,

$$\lim_{M\to\infty} P\,[X_{(M^2)} \leq \Delta]$$

$$= \{P\,[X_1 \geq -\Delta] \ldots P[X_{M^2} \geq -\Delta]\} \cdot [F_X(\Delta)]^{M^2}$$

$$= [1 - F_X(-\Delta)]^{M^2} \cdot [F_X(\Delta)]^{M^2}$$

$$= [F_X(\Delta)]^{2M^2}. \tag{3.9}$$

From equation (3.9) we derive an expression for Δ as follows. Let Δ be the largest value for which

$$P\,[X_{(M^2)} \leqslant \Delta] \leqslant 1-\alpha\,.$$

This is the complementary probability because we wish to have $P\,[|\,\hat{p}_{ij} - p_{ij}\,| \geqslant \Delta$ for some i and j$] \geqslant \alpha$, as in the previous section. The limiting distribution function of the maximum $X_{(M^2)}$ is the same as

$$\lim_{M\to\infty} [F_X(\Delta)]^{2M^2} = \lim_{M\to\infty} [F_Z(2\Delta\sqrt{\frac{K..}{M}}\,]^{2M^2}$$

$$= e^{-e^{-\sqrt{2\log 2M^2}\left(2\Delta\sqrt{\frac{K..}{M}} - \sqrt{2\log 2M^2}\right)}}.$$

Then, approximately,

$$\log(1-\alpha) \approx -e^{-\sqrt{2\log 2M^2}\left(2\Delta\sqrt{\frac{K..}{M}} - \sqrt{2\log 2M^2}\right)}$$

and

$$\log\{-\log(1-\alpha)\} \approx -\sqrt{2\log 2M^2}\left(2\Delta\sqrt{\frac{K..}{M}} - \sqrt{2\log 2M^2}\right).$$

Finally,

$$\Delta \approx (0.5)\sqrt{\frac{M}{K..}}\left(\sqrt{2\log 2M^2} - \frac{\log\log\frac{1}{1-\alpha}}{\sqrt{2\log 2M^2}}\right) \tag{3.10}$$

Equation (3.10) is an approximate expression for Δ based on the asymptotic distribution of the extreme order statistic. We will compare the central limit theorem Δ's with those obtained with the extreme value distribution, in the next section.

C. Comparison of the Three Expressions

The three expressions for Δ obtained using the central limit theorem and order statistics have been developed under aporoximations such as: 1) the steady state distribution of $\{X_t\}$ is $\frac{1}{M}$ (equally likely), 2) the variances of $|\hat{p}_{ij} - p_{ij}\,|$ have $\frac{M}{4K..}$ as a maximum value (worst case), and 3) all transitions are independent. Information about $\{X_t\}$ is from the estimate $\hat{P}$ because we don't have information about the unknown P. In a view of the above approximations and computations, our expressions for Δ are very rough. However they do provide some insight into the occurrences and sizes of estimation errors in $\hat{P}$.

Figure 3.1 contains 3 graphs showing Δ as a function of K.. and M for fixed $\alpha = 0.90$ based on the

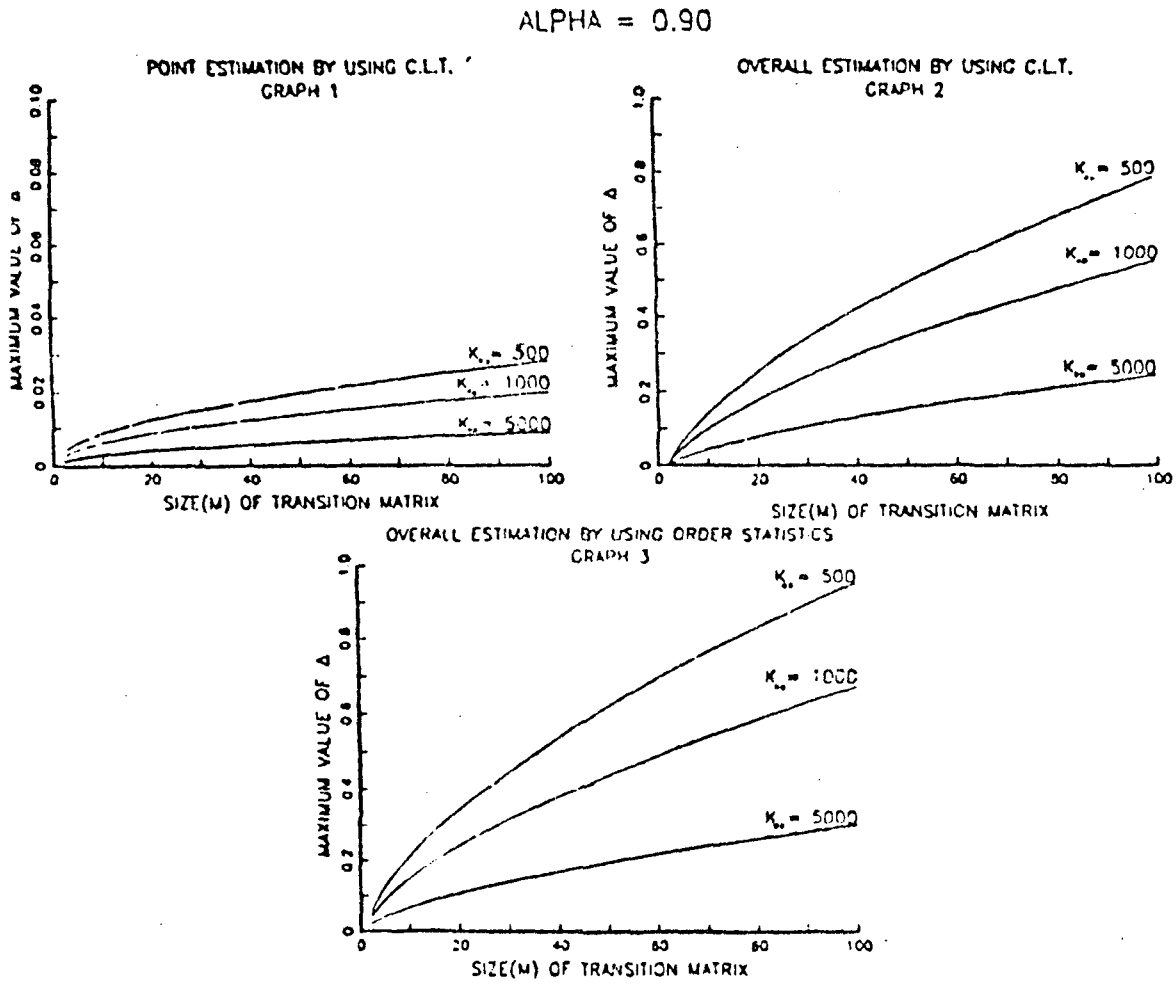

Figure 3.1 Variation of Δ for varying K.. and M.

three expressions (3.3), (3.7) and (3.10).

The first graph shows error bounds using the central limit theorem on $\hat{p}_{ij} - p_{ij}$ for fixed i and j. The second graph is given by the same approach as the first graph, except overall estimation errors are considered, for all i and j. The third graph is based on the asymptotic distribution of the largest value of $|\hat{p}_{ij} - p_{ij}|$ over all i and j.

From Figure 3.1 we see that the largest estimation error depends very much on the number of transition observations and matrix size, but not so much on the α value as seen from Figure 3.2. Graphs 2 and 3 in Figure 3.1 are very similar even though they use different approaches. They give an idea of how large likely values of Δ are for given K.. and M, in the "worst case".

If we consider a Markov chain $\{X_t\}$ with M = 20 or 30 states, and we have observed K = 5000

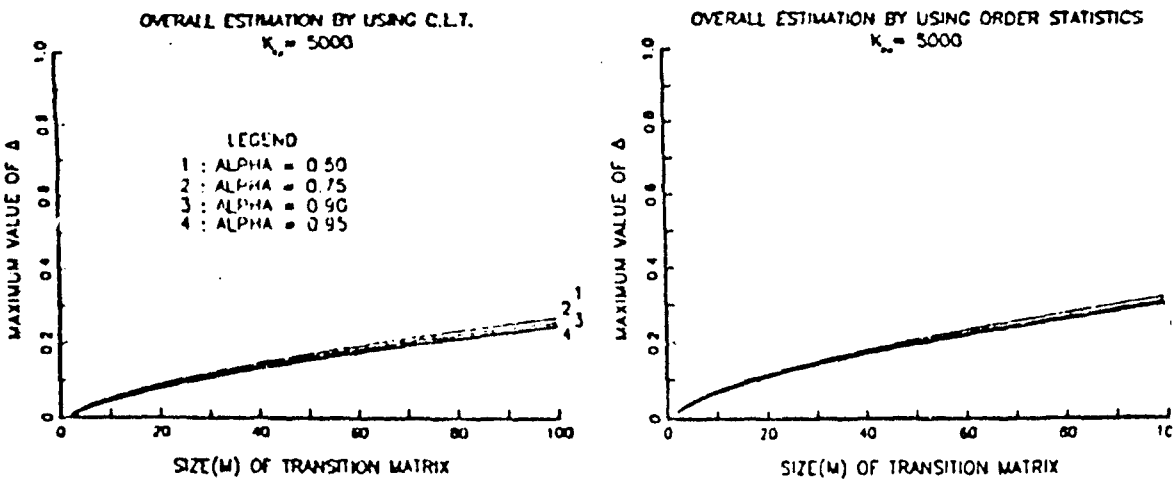

Figure 3.2 Variation of Δ by Changing Alpha (α).

transitions then, roughly, it is likely (prob = 0.90) that at least one element of $\hat{P}$ is in error by at least 0.1. In general, expressions (3.7) and (3.10) may be useful for Markov chains $\{X_t\}$ with M = 20 or 30 states and large numbers of observed transitions.

D. Sensitivity of Lumping Conditions

We have developed expressions for Δ, using the central limit theorem and order statistics. We want to examine the sensitivity of the lumping conditions applied to $\hat{P}$, the estimate of P. If equation (2.1), which is a necessary condition for lumping a Markov chain with transition matrix $\hat{P}$, is satisfied, then the lumped transition matrix $\tilde{\hat{P}}$ is given by equation (2.2). However, even though P satisfies the lumping conditions, it is extremely unlikely that its estimate $\hat{P}$ will also satisfy these conditions, as we shall now demonstrate.

In order to simulate the difference between $\hat{P}$ and P, consider a matrix of errors $R\Delta$, where R is a random matrix with dimension the same as P, whose components are 1's, -1's, and O's where the sum of each row is zero. Now consider the lumpability of the simulated estimate P*, which is constructed by taking P plus the random matrix R times Δ, that is, $P^* = P + R\cdot\Delta$.

To show the sensitivity of the lumping conditions, we assume the unknown P is lumpable with lumping matrix B, and consider the difference (BAP*B – P*B). If equation (2.1) is satisfied by P* then all of these components must be zero.

Theorem 8. The difference (BAP*B – P*B) is a linear function of Δ.

Proof. Let R be the random matrix as defined above and let $P^* = P + R\cdot\Delta$. Then (BAP*B – P*B) is given by

$$
\begin{aligned}
\{\, BA(P+R\cdot\Delta)\,B - (P+R\cdot\Delta)B \,\} &= (BAPB + BAR\,\Delta\,B - PB - R\,\Delta\,B) \\
&= (BAPB - PB + (BARB - RB)\,\Delta) \\
&= (BARB - RB)\;\Delta \\
&= c\cdot\Delta
\end{aligned}
$$

Therefore the difference of BAP*B – P*B is linearly dependent on Δ and P* is not lumpable unless BARB = RB (i.e., R is "lumpable"), which is not likely to occur.

Since $\hat{P}$ is likely to have elements differing appreciably from the corresponding elements in P (errors of size Δ), it can be seen that the lumpability conditions will not be satisfied (not even nearly so) by $\hat{P}$, even though $\{ X_t \}$ is lumpable. We conclude that attempting to check the lumpability of the estimate $\hat{P}$ when P is not known is not useful.

IV. SUMMARY AND CONCLUSIONS

We have given several theorems associated with eigenvalues and eigenvectors for lumpable Markov chains $\{ X_t \}$ with finite state spaces. We have derived rough, approximate mathematical expressions for the largest error made in estimating P by $\hat{P}$ based on transition data.

Both expressions (3.7) and (3.10) are very similar even though the estimated Δ's for the first expression are slightly less than those in the second expression. These expressions show that the largest estimation errors depend very much on the number of transition observations and on the matrix size, but not so much on the α value.

Since $\hat{P}$ is likely to have elements differing appreciably from the corresponding elements in P, it is of interest to examine whether the equation BAPB = PB is likely to be nearly satisfied with $\hat{P}$, i.e., will $(B\hat{A}\hat{P}B - \hat{P}B)$ be nearly zero? This is examined by simulation of "estimates" p* of P, using random perturbations of elements of P of sizes Δ which are likely to occur as errors in $\hat{P}$.

This shows that the classical lumping conditions are extremely sensitive to estimation errors which can be expected to occur even when a large number of transitions have been observed. Thus, the classifical lumping conditions may be of limited value in many actual applications.

As further research, it is recommended that some constructive approach to finding matrices B for lumping a lumpable Markov chain $\{ X_t \}$ be developed, perhaps along the lines of the theorems mentioned in Chapter 2. It is hoped that the present study will be useful to those who might otherwise have endeavored to check the classical condition for lumpability of a Markov chain $\{ X_t \}$ when the transition matrix P has been estimated.

REFERENCES

1. D.R. Barr and M.U. Thomas, "An Approximation Test of Markov Chain Lumpability", J.Am. Statistical Association, vol. 72, pp. 175-179, 1977.
2. J.G. Kemeny and J.L. Snell, Finite Markov Chains, Van Nostrand, Prinston, N.J., 1967.
3. D.R. Barr and M.U. Thomas, "An Eigenvector Condition for Markov Chain Lumpability", Operations Research, vol. 25, no. 6, pp. 1028-1031, November, 1977.
4. S.M. Ross, Introduction to Probability Models, Academic Press, 2nd edition, p. 124, 1980.
5. Albert Nijenhuis and Herbert S. Wilf, Combinatorial Algorithms for Computers and Calculators, Academic Press, 2nd edition, pp. 93-98.
6. Marvin Marcus and Henryk Minc. Introduction to Linear Algebra, Macmillan Company, pp. 163-164, 1965.
7. W.E. Boyce and R.C. Diprima, Elementary Differential Equations and Boundary Value Problems, John Wiley & Sons, 3rd edition, pp. 287-297.
8. H.J. Larson, Introduction to Probability Theory and Interference, John Wiley & Sons, 3rd edition, pp. 318-320, 1982.
9. A.E. Sarhan and B.G. Greenberg, Contributions to Order Statistics, John Wiley & Sons, pp. 15-19.
10. M.G. Kendall and Alan Stuart, The Advanced Theory of Statistics, Griffith, pp. 333-334, 1958.
11. Howard Anton, Elementary Linear Algebra, John Wiley & Sons, 3rd edition.
12. C.J. Burke and M. Rosenblatt, "A Markovian Function of a Markov Chain", Ann. Math. Statist., vol. 29, pp. 112-122, 1958.
13. Volker Abel, "Conditions for Lumping the Grades of a Hierarchical Manpower System", Operations Research, vol. 17, no. 4, pp. 379-383, November, 1983.
14. Volker Abel, "Test on Lumpability for Markovian Manpower Models", Operations Research Proceedings