Hence (b) follows.

$$1 = \frac{m_{jl}+m_{lj}}{m_{lj}+m_{jl}} = \frac{n^{(l)}{}_{jj}/a_j}{n^{(j)}{}_{ll}/a_l}.$$

Hence (c) follows.

If in the Land of Oz example we make **R** absorbing, we obtain (see § 6.1),

$$N = \begin{pmatrix} 4/3 & 4/3 \\ 2/3 & 8/3 \end{pmatrix}.$$

$$\begin{aligned} A+(I-A)N^*(I-A) &= \begin{pmatrix} 2/5 & 1/5 & 2/5 \\ 2/5 & 1/5 & 2/5 \\ 2/5 & 1/5 & 2/5 \end{pmatrix} \\ &+ \begin{pmatrix} 3/5 & -1/5 & -2/5 \\ -2/5 & 4/5 & -2/5 \\ -2/5 & -1/5 & 3/5 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 \\ 0 & 4/3 & 4/3 \\ 0 & 2/3 & 8/3 \end{pmatrix} \begin{pmatrix} 3/5 & -1/5 & -2/5 \\ -2/5 & 4/5 & -2/5 \\ -2/5 & -1/5 & 3/5 \end{pmatrix} \\ &= 1/75 \begin{pmatrix} 86 & 3 & -14 \\ 6 & 63 & 6 \\ -14 & 3 & 86 \end{pmatrix} = Z \end{aligned}$$

as we saw in § 4.3. Using results from § 4.4 and from § 6.1, we can illustrate Corollaries 6.2.6 and 6.2.7.

$$\begin{aligned} m_{NS} &= (1/a_S)(n_{SS}-n_{NS})+t_N-t_S \\ &= (5/2)(8/3-4/3)+8/3-10/3 = 8/3 \end{aligned}$$

$$m_{SR}+m_{RS} = \frac{n_{SS}}{a_S} \quad \text{or} \quad 10/3+10/3 = (8/3)/(2/5).$$

§ 6.3 Combining states. Assume that we are given an r-state Markov chain with transition matrix P and initial vector π. Let $\mathbf{A}=\{\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_t\}$ be a partition of the set of states. We form a new process as follows. The outcome of the j-th experiment in the new process is the set $\mathbf{A}_k$ that contains the outcome of the j-th step in the original chain. We define branch probabilities as follows: At the zero level we assign

$$\mathbf{Pr}_\pi[\mathbf{f}_0 \in \mathbf{A}_i]. \tag{1}$$

At the first level we assign

$$\mathbf{Pr}_\pi[\mathbf{f}_1 \in \mathbf{A}_j | \mathbf{f}_0 \in \mathbf{A}_i].$$

In general, at the n-th level we assign branch probabilities,

$$\mathbf{Pr}_\pi[\mathbf{f}_n \in \mathbf{A}_t | \mathbf{f}_{n-1} \in \mathbf{A}_s \wedge \cdots \wedge \mathbf{f}_1 \in \mathbf{A}_j \wedge \mathbf{f}_0 \in \mathbf{A}_i]. \tag{2}$$

The above procedure could be used to reduce a process with a very large number of states to a process with a smaller number of states. We call this process a *lumped process*. It is also often the case in applications that we are only interested in questions which relate to this coarser analysis of the possibilities. Thus it is important to be able to determine whether the new process can be treated by Markov chain methods.

6.3.1 DEFINITION. *We shall say that a Markov chain is* lumpable *with respect to a partition* $\mathbf{A} = \{\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_r\}$ *if for every starting vector* π *the lumped process defined by* (1) *and* (2) *is a Markov chain and the transition probabilities do not depend on the choice of* π.

We shall see in the next section that, at least for regular chains, the condition that the transition probabilities do not depend on π follows from the requirement that every starting vector give a Markov chain.

Let $p_{iA_j} = \sum_{\mathbf{s}_k \in \mathbf{A}_j} p_{ik}$. Then p_{iA_j} represents the probability of moving from state $\mathbf{s}_i$ into set $\mathbf{A}_j$ in one step for the original Markov chain.

6.3.2 THEOREM. *A necessary and sufficient condition for a Markov chain to be lumpable with respect to a partition* $\mathbf{A} = \{\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_s\}$ *is that for every pair of sets* $\mathbf{A}_i$ *and* $\mathbf{A}_j$, p_{kA_j} *have the same value for every* $\mathbf{s}_k$ *in* $\mathbf{A}_i$. *These common values* $\{\hat{p}_{ij}\}$ *form the transition matrix for the lumped chain.*

PROOF. For the chain to be lumpable it is clearly necessary that

$$\mathbf{Pr}_\pi[\mathbf{f}_1 \in \mathbf{A}_j | \mathbf{f}_0 \in \mathbf{A}_i]$$

be the same for every π for which it is defined. Call this common value $\hat{p}_{ij}$. In particular this must be the same for π having a 1 in its k-th component, for state $\mathbf{s}_k$ in $\mathbf{A}_i$. Hence $p_{kA_j} = \mathbf{Pr}_k[\mathbf{f}_1 \in \mathbf{A}_j] = \hat{p}_{ij}$ for every $\mathbf{s}_k$ in $\mathbf{A}_i$. Thus the condition given is necessary. To prove it is sufficient, we must show that if the condition is satisfied the probability (2) depends only on $\mathbf{A}_s$ and $\mathbf{A}_t$. The probability (2) may be written in the form

$$\mathbf{Pr}_{\pi'}[\mathbf{f}_1 \in \mathbf{A}_t]$$

where π' is a vector with non-zero components only on the states of $\mathbf{A}_s$. It depends on π and on the first n outcomes. However, if $\mathbf{Pr}_k[\mathbf{f}_1 \in \mathbf{A}_t] = \hat{p}_{st}$ for all $\mathbf{s}_k$ in $\mathbf{A}_s$, then it is clear also that $\mathbf{Pr}_{\pi'}[\mathbf{f}_1 \in \mathbf{A}_t] = \hat{p}_{st}$. Thus the probability in (2) depends only on $\mathbf{A}_s$ and $\mathbf{A}_t$.

6.3.3 EXAMPLE. Let us consider the Land of Oz example. Recall that P is given by

$$P = \begin{array}{c} \\ \mathbf{R} \\ \mathbf{N} \\ \mathbf{S} \end{array} \begin{array}{c} \begin{array}{ccc} \mathbf{R} & \mathbf{N} & \mathbf{S} \end{array} \\ \begin{pmatrix} 1/2 & 1/4 & 1/4 \\ 1/2 & 0 & 1/2 \\ 1/4 & 1/4 & 1/2 \end{pmatrix} \end{array}.$$

Assume now that we are interested only in "good" and "bad" weather. This suggests lumping **R** and **S**. We note that the probability of moving from either of these states to **N** is the same. Hence if we choose for our partition $\mathbf{A} = (\{\mathbf{N}\}, \{\mathbf{R},\mathbf{S}\}) = (\mathbf{G}, \mathbf{B})$, the condition for lumpability is satisfied. The new transition matrix is

$$P = \begin{array}{c} \\ \mathbf{G} \\ \mathbf{B} \end{array} \begin{array}{c} \begin{array}{cc} \mathbf{G} & \mathbf{B} \end{array} \\ \begin{pmatrix} 0 & 1 \\ 1/4 & 3/4 \end{pmatrix} \end{array}.$$

Note that the condition for lumpability is not satisfied for the partition $\mathbf{A} = (\{\mathbf{R}\}, \{\mathbf{N},\mathbf{S}\})$ since $p_{NA_1} = p_{NR} = 1/2$ and $p_{SA_1} = p_{SR} = 1/4$.

Assume now that we have a Markov chain which is lumpable with respect to a partition $\mathbf{A} = \{\mathbf{A}_1, \ldots, \mathbf{A}_s\}$. We assume that the original chain had r states and the lumped chain has s states. Let U be the $s \times r$ matrix whose i-th row is the probability vector having equal components for states in $\mathbf{A}_i$ and 0 for the remaining states. Let V be the $r \times s$ matrix with the j-th column a vector with 1's in the components corresponding to states in $\mathbf{A}_j$ and 0's otherwise. Then the lumped transition matrix is given by

$$\hat{P} = UPV.$$

In the Land of Oz example this is

$$\hat{P} = \underset{U}{\begin{pmatrix} 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \end{pmatrix}} \underset{P}{\begin{pmatrix} 1/2 & 1/4 & 1/4 \\ 1/2 & 0 & 1/2 \\ 1/4 & 1/4 & 1/2 \end{pmatrix}} \underset{V}{\begin{pmatrix} 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}}$$

$$= \underset{U}{\begin{pmatrix} 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \end{pmatrix}} \underset{PV}{\begin{pmatrix} 1/4 & 3/4 \\ 0 & 1 \\ 1/4 & 3/4 \end{pmatrix}} = \begin{pmatrix} 0 & 1 \\ 1/4 & 3/4 \end{pmatrix}.$$

Note that the rows of PV corresponding to the elements in the same set of the partition are the same. This will be true in general for a chain which satisfies the condition for lumpability. The matrix U then simply removes this duplication of rows. The choice of U is by no means unique. In fact, all that is needed is that the i-th row should be a probability vector with non-zero components only for states in $\mathbf{A}_i$. We have chosen, for convenience, the vector with equal components for these states. Also it is convenient for proofs to assume that the states are numbered so that those in $\mathbf{A}_1$ come first, those in $\mathbf{A}_2$ come next, etc. In all proofs we shall understand that this had been done.

The following result will be useful in deriving formulas for lumped chains.

6.3.4 THEOREM. *If P is the transition matrix of a chain lumpable with respect to the partition* $\mathbf{A}$, *and if the matrices U and V are defined —as above—with respect to this partition, then*

$$VUPV = PV. \tag{3}$$

PROOF. The matrix VU has the form

$$VU = \left(\begin{array}{c|c|c} W_1 & 0 & 0 \\ \hline 0 & W_2 & 0 \\ \hline 0 & 0 & W_3 \end{array}\right),$$

where W_1, W_2, and W_3 are probability matrices. Condition (3) states that the columns of PV are fixed vectors of VU. But since the chain is lumpable, the probability of moving from a state of $\mathbf{A}_i$ to the set $\mathbf{A}_j$ is the same for all states in $\mathbf{A}_i$, hence the components of a column of PV corresponding to $\mathbf{A}_j$ are all the same. Therefore they form a fixed vector for W_j. This proves (3).

6.3.5 THEOREM. *If P,* $\mathbf{A}$, *U, and V are as in Theorem* 6.3.4, *then condition* (3) *is equivalent to lumpability.*

PROOF. We have already seen that (3) is implied by lumpability. Conversely, let us suppose that (3) holds. Then the columns of PV are fixed vectors for VU. But each W_j is the transition matrix of an ergodic chain, hence its only fixed column vectors are of the form $c\xi$. Hence all the components of a column of PV corresponding to one set $\mathbf{A}_j$ must be the same. That is, the chain is lumpable by § 6.3.2.

Note that from (3)

$$\begin{aligned}\hat{P}^2 &= UPVUPV \\ &= UP^2V\end{aligned}$$

and in general

$$\hat{P}^n = UP^nV.$$

This last fact could also be verified directly from the process.

Assume now that P is an absorbing chain. We shall restrict our discussion to the case where we lump only states of the same kind. That is, any subset of our partition will contain only absorbing states or only non-absorbing states. We recall that the standard form for an absorbing chain is

$$P = \left(\begin{array}{c|c} I & O \\ \hline R & Q \end{array}\right).$$

We shall write U in the form

$$U = \left(\begin{array}{c|c} U_1 & O \\ \hline O & U_2 \end{array}\right),$$

where entries of U_1 refer to absorbing states and entries of U_2 to non-absorbing states. Similarly we write V in the form

$$V = \left(\begin{array}{c|c} V_1 & O \\ \hline O & V_2 \end{array}\right).$$

Then, if we consider the condition for lumpability, $VUPV = PV$, we obtain in terms of the above matrices the equivalent set of conditions:

$$V_1U_1V_1 = V_1 \tag{4a}$$

$$V_2U_2RV_1 = RV_1 \tag{4b}$$

$$V_2U_2QV_2 = QV_2. \tag{4c}$$

Since $U_1V_1 = I$, the first condition is automatically satisfied.

The standard form for the transition matrix $\hat{P}$ is obtained from

$$\hat{P} = UPV$$

$$\hat{P} = \left(\begin{array}{c|c} U_1 & O \\ \hline O & U_2 \end{array}\right)\left(\begin{array}{c|c} I & O \\ \hline R & Q \end{array}\right)\left(\begin{array}{c|c} V_1 & O \\ \hline O & V_2 \end{array}\right).$$

Multiplying this out we obtain

$$\hat{P} = \left(\begin{array}{c|c} I & O \\ \hline U_2RV_1 & U_2QV_2 \end{array}\right).$$

Hence we have

$$\begin{aligned} \hat{R} &= U_2RV_1 \\ \hat{Q} &= U_2QV_2. \end{aligned}$$

From condition (4c) we obtain

$$\begin{aligned} \hat{Q}^2 &= U_2QV_2U_2QV_2 \\ &= U_2Q^2V_2. \end{aligned}$$

More generally we have

$$\hat{Q}^n = U_2Q^nV_2.$$

From the infinite series representation for the fundamental matrix N we have

$$\begin{aligned} \hat{N} &= I+\hat{Q}+\hat{Q}^2+\cdots \\ &= U_2IV_2+U_2QV_2+\cdots \\ &= U_2(I+Q+Q^2+\cdots)V_2 \\ \hat{N} &= U_2NV_2. \end{aligned}$$

From this we obtain

$$\begin{aligned} \hat{\tau} &= U_2NV_2\xi \\ \hat{\tau} &= U_2N\xi \\ \hat{\tau} &= U_2\tau \end{aligned}$$

and

$$\begin{aligned} \hat{B} = \hat{N}\hat{R} &= U_2NV_2U_2RV_1 \\ \hat{B} &= U_2NRV_1 \\ \hat{B} &= U_2BV_1. \end{aligned}$$

Hence all three of the quantities N, τ, and B are easily obtained for the lumped chain from the corresponding quantities for the original chain.

An important consequence of our result $\hat{\tau}=U_2\tau$ is the following. Let $\mathbf{A}_i$ be any non-absorbing set, and $\mathbf{s}_k$ be a state in $\mathbf{A}_i$. We can choose the i-th row of U_2 to be a probability vector with 1 in the $\mathbf{s}_k$ component. But this means that $t_i=t_k$ for all $\mathbf{s}_k$ in $\mathbf{A}_i$. Hence when a chain is lumpable, the mean time to absorption must be the same for all starting states $\mathbf{s}_k$ in the same set $\mathbf{A}_i$

As an example of the above, let us consider the random walk example

with transition matrix

$$P = \begin{array}{c} \\ s_1 \\ s_5 \\ s_2 \\ s_3 \\ s_4 \end{array} \begin{array}{c} \begin{array}{ccccc} s_1 & s_5 & s_2 & s_3 & s_4 \end{array} \\ \left(\begin{array}{cc|ccc} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ \hline 1/2 & 0 & 0 & 1/2 & 0 \\ 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 1/2 & 0 & 1/2 & 0 \end{array}\right) \end{array}.$$

We take the partition $\mathbf{A} = (\{s_1, s_5\}, \{s_2, s_4\}, \{s_3\})$. For this partition the condition for lumpability is satisfied. Notice that this would not have been the case if we have unequal probabilities for moving to the right or left.

From the original chain we found

$$N = \begin{array}{c} \\ s_2 \\ s_3 \\ s_4 \end{array} \begin{array}{c} \begin{array}{ccc} s_2 & s_3 & s_4 \end{array} \\ \begin{pmatrix} 3/2 & 1 & 1/2 \\ 1 & 2 & 1 \\ 1/2 & 1 & 3/2 \end{pmatrix} \end{array}$$

$$\tau = \begin{pmatrix} 3 \\ 4 \\ 3 \end{pmatrix}$$

$$B = \begin{array}{c} \\ s_2 \\ s_3 \\ s_4 \end{array} \begin{array}{c} \begin{array}{cc} s_1 & s_5 \end{array} \\ \begin{pmatrix} 3/4 & 1/4 \\ 1/2 & 1/2 \\ 1/4 & 3/4 \end{pmatrix} \end{array}.$$

The corresponding quantities for the lumped process are

$$\hat{P} = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 1/2 & 0 & 0 & 1/2 & 0 \\ 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 1/2 & 0 & 1/2 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

$$= \begin{array}{c} \\ A_1 \\ A_2 \\ A_3 \end{array} \begin{array}{c} \begin{array}{ccc} A_1 & A_2 & A_3 \end{array} \\ \begin{pmatrix} 1 & 0 & 0 \\ 1/2 & 0 & 1/2 \\ 0 & 1 & 0 \end{pmatrix} \end{array}$$

$$\hat{N} = \begin{pmatrix} 1/2 & 0 & 1/2 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 3/2 & 1 & 1/2 \\ 1 & 2 & 1 \\ 1/2 & 1 & 3/2 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \end{pmatrix}$$

$$= \begin{array}{c} \\ A_2 \\ A_3 \end{array} \begin{array}{c} \begin{array}{cc} A_2 & A_3 \end{array} \\ \begin{pmatrix} 2 & 1 \\ 2 & 2 \end{pmatrix} \end{array}$$

$$\hat{\tau} = \begin{pmatrix} 1/2 & 0 & 1/2 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 3 \\ 4 \\ 3 \end{pmatrix} = \begin{array}{c} A_2 \\ A_3 \end{array} \begin{pmatrix} 3 \\ 4 \end{pmatrix}$$

$$\hat{B} = \begin{pmatrix} 1/2 & 0 & 1/2 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 3/4 & 1/4 \\ 1/2 & 1/2 \\ 1/4 & 3/4 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

$$= \begin{array}{c} A_2 \\ A_3 \end{array} \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Assume now that we have an ergodic chain which satisfies the condition for lumpability for a partition **A**. The resulting chain will be ergodic. Let $\hat{A}$ be the limiting matrix for the lumped chain. Then we know that

$$\hat{A} = \lim_{n \to \infty} \frac{\hat{P} + \hat{P}^2 + \cdots + \hat{P}^n}{n}$$

$$\hat{A} = \lim_{n \to \infty} \frac{UPV + UP^2V + \cdots + UP^nV}{n}$$

$$\hat{A} = UAV.$$

In particular, this states that the components of $\hat{\alpha}$ are obtained from α by simply adding components in a given set. Similarly from the infinite series representation for the fundamental matrix $\hat{Z}$ we have

$$\hat{Z} = UZV.$$

There is in general no simple relation between M and $\hat{M}$. However, the mean time to go from a state in $\mathbf{A}_i$ to the set $\mathbf{A}_j$, in the original process, is the same for all states in $\mathbf{A}_i$. To see this we need only make the states of $\mathbf{A}_j$ absorbing. We know that the mean time to absorption is the same for all starting states chosen from a given set. If, in

addition, $\mathbf{A}_j$ happens to consist of a single state, then $\hat{m}_{ij}$ may be found from M.

We can also compute the covariance matrix of the lumped process. As a matter of fact we know (see § 4.6.7) that the covariances are easily obtainable from C even if the original process is not lumpable with respect to the partition, that is, if the lumped process is not a Markov chain. In any case

$$\hat{C} = \sum_{\substack{s_k \text{ in } \mathbf{A}_i \\ s_l \text{ in } \mathbf{A}_j}} c_{kl}.$$

Let us carry out these computations for the Land of Oz example. For $\hat{A}$ we have

$$\hat{A} = \begin{pmatrix} 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \end{pmatrix} \begin{pmatrix} 2/5 & 1/5 & 2/5 \\ 2/5 & 1/5 & 2/5 \\ 2/5 & 1/5 & 2/5 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$$= \begin{pmatrix} 1/5 & 4/5 \\ 1/5 & 4/5 \end{pmatrix}$$

$$Z = \begin{pmatrix} 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \end{pmatrix} \begin{pmatrix} 86/75 & 3/75 & -14/75 \\ 6/75 & 63/75 & 6/75 \\ -14/75 & 3/75 & 86/75 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$$= \begin{pmatrix} 63/75 & 12/75 \\ 3/75 & 72/75 \end{pmatrix}$$

$$\hat{C} = \begin{pmatrix} 12/125 & -12/125 \\ -12/125 & 12/125 \end{pmatrix}$$

$$M = \begin{array}{c} \\ \mathbf{R} \\ \mathbf{N} \\ \mathbf{S} \end{array} \begin{array}{c} \begin{array}{ccc} \mathbf{R} & \mathbf{N} & \mathbf{S} \end{array} \\ \begin{pmatrix} 5/2 & 4 & 10/3 \\ 8/3 & 5 & 8/3 \\ 10/3 & 4 & 5/2 \end{pmatrix} \end{array}.$$

From the fundamental matrix $\hat{Z}$ we find,

$$\hat{M} = \begin{array}{c} \\ \mathbf{N} \\ \mathbf{B} \end{array} \begin{array}{c} \begin{array}{cc} \mathbf{N} & \mathbf{B} \end{array} \\ \begin{pmatrix} 5 & 1 \\ 4 & 5/4 \end{pmatrix} \end{array}.$$

Note that the mean time to reach **N** from either **R** or **S** is 4. Here **N** in the lumped process is a single element set. This common value is the mean time in the lumped chain to go from **B** to **N**. Similarly, the value 5 is obtainable from M. We observe that the mean time to go from **N** to **B** is considerably less than the mean time to go from **N** to either of the states in **B** in the original process.

§ **6.4 Weak lumpability.** In practice if one wanted to apply Markov chain ideas to a process for which the states have been combined, with respect to a partition $\mathbf{A} = \{\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_n\}$, it is most natural to require that the resulting process be a Markov chain no matter what choice is made for the starting vector. However, there are some interesting theoretical considerations when we require only that at least one starting vector lead to a Markov chain. When this is the case we shall say that the process is *weakly lumpable with respect to the partition* **A**. We shall investigate the consequences of this weaker assumption in this section. We restrict the discussion to regular chains. The results of this section are based in part on results of C. K. Burke and M. Rosenblatt.†

For a given starting vector π, to determine whether or not the process is a Markov chain we must examine probabilities of the form

$$\mathbf{Pr}_\pi[\mathbf{f}_{n+1} \in \mathbf{A}_t | \mathbf{f}_n \in \mathbf{A}_s \wedge \cdots \wedge \mathbf{f}_1 \in \mathbf{A}_j \wedge \mathbf{f}_0 \in \mathbf{A}_i]. \tag{1}$$

For a given π the process will be a Markov chain if these probabilities do not depend upon the outcomes before the n-th.

We must find conditions under which the knowledge of the outcomes before the last one does not affect the probability (1). Let us see how such knowledge could affect it. Given the information in (1), we know that after n steps the underlying chain is in a state in $\mathbf{A}_s$, but we do not know in which state it is. We can, however, assign probabilities for its being in each state of $\mathbf{A}_s$. We do this as follows: For any probability vector β, we denote by β^j the probability vector formed by making all components corresponding to states not in $\mathbf{A}_j$ equal to 0 and the remaining components proportional to those of β. We shall say that β^j is β *restricted to* $\mathbf{A}_j$. (If β has all components 0 in $\mathbf{A}_j$ we do not define β^j.) Consider now the information given in (1). The fact that $\mathbf{f}_0 \in \mathbf{A}_i$ may be interpreted as changing our initial vector to π^i. Learning then that $\mathbf{f}_1 \in \mathbf{A}_j$ may be interpreted as changing this vector to $(\pi^i P)^j$. We continue this process until we have taken into account all of the information given in (1). We are led to a certain assignment of probabilities for the states in $\mathbf{A}_s$. From these probabilities we can

† C. K. Burke and M. Rosenblatt, "A Markovian function of a Markov chain," *Annals of Mathematical Statistics*, **29**: 1112–1122, 1958.

easily compute the probability of a transition to $\mathbf{A}_t$ on the next step. But note that this probability may be quite different for different kinds of information. For example, our information may place high probability for being in a state from which it is certain that we move to $\mathbf{A}_t$. A different history of the process may place low probability on this state. These considerations give us a clue as to when we could expect that the past could be ignored. Two different cases are suggested. First would be the case where the information gained from the past would not do us any good. For example, assume that the probability for moving to the set $\mathbf{A}_t$ from a state in $\mathbf{A}_s$ is the same for all states in $\mathbf{A}_s$. Then clearly the probabilities for being in each state of $\mathbf{A}_s$ would not affect our predictions for the next outcome in the lumped process. This is the condition we found for lumpability in § 6.3.2. A second condition is suggested by the following: Assume that no matter what the past information is, we always end up with the same assignment of probabilities for being in each of the states in $\mathbf{A}_s$. Then again the past can have no influence on our predictions. We shall see that this case can also arise.

We have indicated above that the information given in (1) can be represented by a probability vector restricted to $\mathbf{A}_s$. This vector is obtained from the initial vector π by a sequence of transformations, each time taking into account one more bit of information. That is, we form the sequence

$$\left.\begin{aligned} \pi_1 &= \pi^i \\ \pi_2 &= (\pi_1 P)^j \\ \pi_3 &= (\pi_2 P)^k \\ &\vdots \\ \pi_m &= (\pi_{m-1} P)^s \end{aligned}\right\} \tag{2}$$

We denote by $\mathbf{Y}_s$ the totality of vectors obtained by considering all finite sequences $\mathbf{A}_i, \mathbf{A}_j, \ldots, \mathbf{A}_s$, ending in $\mathbf{A}_s$.

6.4.1 THEOREM. *The lumped chain is a Markov chain for the initial vector π if and only if for every s and t the probability $\mathbf{Pr}_\beta[\mathbf{f}_1 \in \mathbf{A}_t]$ is the same for every β in $\mathbf{Y}_s$. This common value is the transition probability for moving from set $\mathbf{A}_s$ to set $\mathbf{A}_t$ in the lumped process.*

PROOF. The probability (1) can be represented in the form $\mathbf{Pr}_\beta[\mathbf{f}_1 \in \mathbf{A}_t]$ for a suitable β in $\mathbf{Y}_s$. To do this we use the first n outcomes for the construction (2). By hypothesis this probability depends only on s and t as required. Hence the lumped process is a Markov chain. Conversely, assume that the lumped chain is a Markov chain for initial vector π. Let β be any vector in $\mathbf{Y}_s$. Then β is obtained from a possible sequence, say of length $\mathbf{n}$, $\mathbf{A}_i, \mathbf{A}_j, \ldots, \mathbf{A}_s$. Let these be the

given outcomes used to compute a probability of the form (1). This probability is $\mathbf{Pr}_\beta[\mathbf{f}_1 \in \mathbf{A}_t]$ and by the Markov property must not depend upon the outcomes before $\mathbf{A}_t$. Hence it has the same value for every β in $\mathbf{Y}_s$.

6.4.2 EXAMPLE. Consider a Markov chain with transition matrix

$$P = \begin{array}{c} \\ \mathbf{A}_1 \\ \\ \mathbf{A}_2 \end{array} \begin{array}{c} \begin{array}{cc} \mathbf{A}_1 & \mathbf{A}_2 \end{array} \\ \left(\begin{array}{c|cc} 1/4 & 1/4 & 1/2 \\ \hline 0 & 1/6 & 5/6 \\ 7/8 & 1/8 & 0 \end{array}\right) \end{array}.$$

Let $\mathbf{A} = (\{\mathbf{s}_1\}, \{\mathbf{s}_2, \mathbf{s}_3\})$. Consider any vector of the form $(1-3a, a, 2a)$. Any such vector multiplied by P will again be of this form. Also any such vector restricted to $\mathbf{A}_1$ or $\mathbf{A}_2$ will be such a vector. Hence for any such starting vector the set $\mathbf{Y}_1$ will contain the single element $(1, 0, 0)$ and $\mathbf{Y}_2$ the single element $(0, 1/3, 2/3)$. Thus the condition of § 6.4.1 is satisfied trivially for any such starting vector. On the other hand assume that our starting vector is $\pi = (0, 0, 1)$. Let $\pi_1 = (\pi P)^2 = (0, 1, 0)$ and $\pi_2 = (\pi_1 P)^2 = (0, 1/6, 5/6)$. Then π_1 and π_2 are in $\mathbf{Y}_2$ and $\mathbf{Pr}_{\pi_1}[\mathbf{f}_1 \in \mathbf{A}_1] = 0$ while $\mathbf{Pr}_{\pi_2}[\mathbf{f}_1 \in \mathbf{A}_1] = 35/48$. Hence this choice of starting vector does not lead to a Markov chain.

We see that it is possible for certain starting vectors to lead to Markov chains while others do not. We shall now prove that if there is any starting vector which gives a Markov chain, then the fixed vector α does.

6.4.3 THEOREM. *Assume that a regular Markov chain is weakly lumpable with respect to* $\mathbf{A} = \{\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_s\}$. *Then the starting vector* α *will give a Markov chain for the lumped process. The transition probabilities will be*

$$\hat{p}_{ij} = \mathbf{Pr}_{\alpha^i}[\mathbf{f}_1 \in \mathbf{A}_j].$$

Any other starting vector which yields a Markov chain for the lumped process will give the same transition probabilities.

PROOF. Since the chain is weakly lumpable there must be some starting vector π which leads to a Markov chain. Let its transition matrix be $\{\hat{p}_{ij}\}$. For this vector π

$$\mathbf{Pr}_\pi[\mathbf{f}_n \in \mathbf{A}_j | \mathbf{f}_{n-1} \in \mathbf{A}_i \wedge \mathbf{f}_{n-2} \in \mathbf{A}_k] = \hat{p}_{ij}$$

for all sets for which this probability is defined. But this may be written as

$$\mathbf{P}_{\pi P^{n-2}}[\mathbf{f}_2 \in \mathbf{A}_j | \mathbf{f}_1 \in \mathbf{A}_i \wedge \mathbf{f}_0 \in \mathbf{A}_k].$$

Letting n tend to infinity we have

$$\mathbf{Pr}_\alpha[\mathbf{f}_2 \in \mathbf{A}_j | \mathbf{f}_1 \in \mathbf{A}_i \wedge \mathbf{f}_0 \in \mathbf{A}_k] = \hat{p}_{ij}.$$

We have proved that the probability of the form (1), with α as starting vector, does not depend upon the past beyond the last outcome for the case $n=1$. The general case is similar. Therefore, for α as a starting vector, the lumped process is a Markov chain. In the course of the proof we showed that $\hat{p}_{ij}$ for a starting vector π is the same as for α, hence it will be the same for any starting vector which yields a Markov chain.

By the previous theorem, if we are testing for weak lumpability we may assume that the process is started with the initial vector α. In this case the transition matrix $\hat{P}$ can be written in the form

$$\hat{P} = UPV$$

where V is as before but U is a matrix with i-th row α^i. When we have lumpability there is a great deal of freedom in the choice of U and in that case we chose a more convenient U. We do not have this freedom for weak lumpability.

We consider now conditions for which we can expect to have weak lumpability. If the chain is to be a Markov chain when lumped then we can compute $\hat{P}^2$ in two ways. Computing it directly from the underlying chain we have $\hat{P}^2 = UP^2V$. By squaring $\hat{P}$ we have $UPVUPV$. Hence it must be true that

$$UPVUPV = UPPV.$$

One sufficient condition for this is

$$VUPV = PV. \tag{3}$$

This is the condition for lumpability expressed in terms of our new U. It is necessary and sufficient for lumpability, and hence sufficient for weak lumpability. A second condition which would be sufficient for the above is

$$UPVU = UP. \tag{4}$$

This condition states the rows of UP are fixed vectors for VU. The matrix VU is now of the form

$$VU = \left(\begin{array}{c|c|c} W_1 & 0 & 0 \\ \hline 0 & W_2 & 0 \\ \hline 0 & 0 & W_3 \end{array}\right),$$

where W_j is a transition matrix having all rows equal to α^j. To say that the i-th row of UP is a fixed vector for VU means that this vector, restricted to $\mathbf{A}_j$, is a fixed vector for W_j. But this means that the components of this vector must be proportional to α^j. Hence we have

$$(\alpha^i P)^j = \alpha^j. \tag{5}$$

This means that if we start with α, the set $\mathbf{Y}_i$, obtained by construction (2), consists, for each i, of a single element, namely α^i. Conversely, if each such set has only a single element, then (5) is satisfied and hence also (4). To say that $\mathbf{Y}_i$ has only one element for each i is to say that when the last outcome was $\mathbf{A}_i$ the knowledge of previous outcomes does not influence the assignment of the probabilities for being in each of the states of $\mathbf{A}_i$. Hence we have found that (4) is necessary and sufficient for the past beyond the last outcome to provide no new information, and is sufficient for weak lumpability.

Example 6.4.2 is a case where (4) is satisfied. Recall that we found that each $\mathbf{Y}_i$ had only one element.

We can summarize our findings as follows: We stated in the introduction that there are two obvious ways to make the information contained in the outcomes before the last one useless. One way is to require that even if we know the exact state of the original process our predictions would be unchanged. This is condition (3). The other is to require that we get no information at all from the past except the last step. This is condition (4). Each leads to weak lumpability. We have thus proved:

6.4.4. THEOREM. *Either condition* (3) *or condition* (4) *is sufficient for weak lumpability.*

There is an interesting connection between (3) and (4) in terms of the process and its associated reverse process (see § 5.3).

6.4.5 THEOREM. *A regular chain satisfies* (3) *if and only if the reverse chain satisfies* (4).

PROOF. Assume that a process satisfies (3). Then

$$VUPV = PV.$$

Let P_0 be the transition matrix for the reverse process, then $P = DP^T{}_0D^{-1}$. Hence

$$VUDP^T{}_0D^{-1}V = DP^T{}_0D^{-1}V$$

or, transposing,

$$V^TD^{-1}P_0DU^TV^T = V^TD^{-1}P_0D,$$

and

$$V^T D^{-1} P_0 D U^T V^T D^{-1} = V^T D^{-1} P_0.$$

We observe that $V^T D^{-1} = \hat{D}^{-1} U$. Furthermore, VUD is a symmetric matrix so that $VUD = DU^T V^T$ or $DU^T V^T D^{-1} = VU$. Using these two facts, our last equation becomes

$$\hat{D}^{-1} U P_0 V U = \hat{D}^{-1} U P_0.$$

Multiplying on the left by $\hat{D}$ gives condition (4) for P_0. The proof of the converse is similar.

6.4.6 THEOREM. *If a given process is weakly lumpable with respect to a partition* $\mathbf{A}$, *then so is the reverse process.*

PROOF. We must prove that all probabilities of the form

$$\mathbf{Pr}_a[\mathbf{f}_1 \in \mathbf{A}_i | \mathbf{f}_2 \in \mathbf{A}_j \wedge \mathbf{f}_3 \in \mathbf{A}_h \wedge \cdots \wedge \mathbf{f}_n \in \mathbf{A}_t]$$

depend only on $\mathbf{A}_i$ and $\mathbf{A}_j$. We can write this probability in the form

$$\frac{\mathbf{Pr}_a[\mathbf{f}_1 \in \mathbf{A}_i \wedge \mathbf{f}_2 \in \mathbf{A}_j \wedge \mathbf{f}_3 \in \mathbf{A}_h \wedge \cdots \wedge \mathbf{f}_n \in \mathbf{A}_t]}{\mathbf{Pr}_a[\mathbf{f}_2 \in \mathbf{A}_j \wedge \mathbf{f}_3 \in \mathbf{A}_h \wedge \cdots \wedge \mathbf{f}_n \in \mathbf{A}_t]}$$

$$= \frac{\mathbf{Pr}_a[\mathbf{f}_n \in \mathbf{A}_t \wedge \cdots \wedge \mathbf{f}_3 \in \mathbf{A}_h | \mathbf{f}_2 \in \mathbf{A}_j \wedge \mathbf{f}_1 \in \mathbf{A}_i]\ \mathbf{Pr}_a[\mathbf{f}_1 \in \mathbf{A}_i \wedge \mathbf{f}_2 \in \mathbf{A}_j]}{\mathbf{Pr}_a[\mathbf{f}_n \in \mathbf{A}_t \wedge \cdots \wedge \mathbf{f}_3 \in \mathbf{A}_h | \mathbf{f}_2 \in \mathbf{A}_j]\ \mathbf{Pr}_a[\mathbf{f}_2 \in \mathbf{A}_j]}.$$

By hypothesis the forward process is a Markov chain, so that the first term in the numerator does not depend on $\mathbf{A}_i$. Hence this whole expression is simply

$$\frac{\mathbf{Pr}_a[\mathbf{f}_1 \in \mathbf{A}_i \wedge \mathbf{f}_2 \in \mathbf{A}_j]}{\mathbf{Pr}_a[\mathbf{f}_2 \in \mathbf{A}_j]}$$

which depends only on $\mathbf{A}_i$ and $\mathbf{A}_j$.

6.4.7 THEOREM. *A reversible regular Markov chain is reversible when lumped.*

PROOF. By reversibility,

$$P = DP^T D^{-1}$$

and

$$\hat{P} = UPV.$$

Hence

$$\hat{P} = UDP^T D^{-1} V.$$

We have seen that $V^T D^{-1} = \hat{D}^{-1} U$. Hence $UD = \hat{D} V^T$. Also $D^{-1} V = U^T \hat{D}^{-1}$. Thus we have

$$\hat{P} = \hat{D} V^T P^T U^T \hat{D}^{-1},$$

and hence

$$\hat{P} = \hat{D} \hat{P}^T \hat{D}^{-1}.$$

This means that the lumped process is reversible.

6.4.8 THEOREM. *For a reversible regular Markov chain, weak lumpability implies lumpability.*

PROOF. Let P be the transition matrix for a regular reversible chain. Then, if the chain is weakly lumpable,

$$UPPV = UPVUPV$$

or

$$UP(I - VU)PV = 0.$$

Since $U = \hat{D} V^T D^{-1}$, we have

$$\hat{D} V^T D^{-1} P(I - VU)PV = 0,$$

or, multiplying through by $\hat{D}^{-1}$ and using the fact that for a reversible chain $D^{-1}P = P^T D^{-1}$, we have

$$V^T P^T D^{-1}(I - VU)PV = 0.$$

Let $W = D^{-1} - D^{-1}VU$. Then $W = D^{-1} - U^T \hat{D}^{-1} U$. We shall show that W is semi-definite. That is, for any vector β, $\beta^T W \beta$ is non-negative. It is sufficient to prove that

$$\sum_{k \text{ in } \mathbf{A}_i} a_k b^2_k \geqslant \hat{d}_i \Big(\sum_{k \text{ in } \mathbf{A}_i} a_k b_k \Big)^2$$

where $\hat{d}_i$ is the i-th diagonal entry of $\hat{D}$, or equivalently

$$\sum_{k \text{ in } \mathbf{A}_i} a_k \hat{d}_i b^2_k \geqslant \sum_{k \text{ in } \mathbf{A}_i} (a_k \hat{d}_i b_k)^2.$$

But since the coefficients $a_k \hat{d}_i$ are non-negative and have sum 1, this is a standard inequality of probability theory. It can be proved by considering a function f which takes on the value b_k with probability $a_k \hat{d}_i$. Then the inequality expresses that $\mathbf{M}[\mathbf{f}^2] \geqslant (\mathbf{M}[\mathbf{f}])^2$; and, by § 1.8.5, this simply asserts that the variance of f is non-negative.

Since W_i is semi-definite, $W = X^T X$ for some matrix X. Thus

$$V^T P^T X^T X P V = 0$$

or

$$(XPV)^T (XPV) = 0.$$

This can be true only if

$$XPV = 0.$$

Hence

$$X^T XPV = 0,$$

or

$$D^{-1}(I - VU)PV = 0,$$

or

$$(I - VU)PV = 0.$$

Hence

$$PV = VUPV.$$

Note that while we have given necessary and sufficient conditions for lumpability with respect to a partition A, we have not given necessary and sufficient conditions for weak lumpability. We have given two different sufficient conditions (3) and (4). It might be hoped that for weak lumpability one of the two conditions would have to be satisfied. It is, however, easy to get an example where neither is satisfied as follows: If we take a Markov chain and find a method of combining states to give a Markov chain, we can then ask whether the new chain can be combined. If so, the result can be considered a combining of states in the original chain. To get our counterexample, we take a chain for which we can combine states by condition (3) and then combine states in the new chain by condition (4); the result considered as a lumping of the original chain will obviously be a Markov chain, but it will satisfy neither (4) nor (3). Consider a Markov chain with transition matrix

$$P = \begin{array}{c} A_1 \\ A_2 \\ \\ A_3 \end{array} \left(\begin{array}{c|cc|c} 1/4 & 1/16 & 3/16 & 1/2 \\ \hline 0 & 1/12 & 1/12 & 5/6 \\ 0 & 1/12 & 1/12 & 5/6 \\ \hline 7/8 & 1/32 & 3/32 & 0 \end{array}\right)$$

For the partition $A = (\{s_1\}, \{s_2, s_3\}, \{s_4\})$ the strong condition (3) is satisfied. Hence we obtain a lumped chain with transition matrix

$$P = \begin{array}{c} \\ A_1 \\ A_2 \\ A_3 \end{array} \begin{array}{c} \begin{array}{ccc} A_1 & A_2 & A_3 \end{array} \\ \begin{pmatrix} 1/4 & 1/4 & 1/2 \\ 0 & 1/6 & 5/6 \\ 7/8 & 1/8 & 0 \end{pmatrix} \end{array}.$$

But this is Example 6.4.2, which satisfies (4). Hence we can lump it by $(\{A_1\}, \{A_2, A_3\})$. The result is a lumping of the original chain by

$\mathbf{A} = (\{s_1\}, \{s_2, s_3, s_4\})$. It is easily checked that neither (3) nor (4) is satisfied in the original process for this partition.

We conclude with some remarks about a lumped process when the condition for weak lumpability is not satisfied. We assume that P is regular. Then if the process is started in equilibrium

$$\hat{p}_{ij} = \mathbf{Pr}_\alpha[\mathbf{f}_{n+1} \in \mathbf{A}_j | \mathbf{f}_n \in \mathbf{A}_i]$$

is the same for every n. Hence the matrix $\hat{P} = \{\hat{p}_{ij}\}$ may still be interpreted as a one-step transition matrix. Also

$$\hat{a}_i = \mathbf{Pr}_\alpha[\mathbf{f}_n \in \mathbf{A}_i]$$

is the same for all n. The vector $\hat{a} = \{\hat{a}_i\}$ will be the unique fixed vector for $\hat{P}$. Its components may be obtained from α by simply adding the components corresponding to each set. Similarly we may define two-step transition probabilities by

$$\hat{p}^{(2)}{}_{ij} = \mathbf{Pr}_\alpha[\mathbf{f}_{n+2} \in \mathbf{A}_j | \mathbf{f}_n \in \mathbf{A}_i].$$

The two-step transition matrix will then be $\hat{P}^{(2)} = \{\hat{p}^{(2)}{}_{ij}\}$. It will no longer be true that $\hat{P}^2 = \hat{P}^{(2)}$.

We can also define the mean first passage matrix $\hat{M}$ for the lumped process. It cannot be obtained by our Markov chain formulas. To obtain $\hat{M}$ it is necessary first to find m_{i,A_j}, the mean time to go from state i to set $\mathbf{A}_j$ in the original process. We can do this by making all of the elements of $\mathbf{A}_j$ absorbing and find the mean time to absorption. (A slight modification is necessary if i is in $\mathbf{A}_j$.) From these we obtain the mean time to go from $\mathbf{A}_i$ to $\mathbf{A}_j$, by

$$\hat{m}_{ij} = \sum_{k \text{ in } \mathbf{A}_i} a^*{}_k m_{k,A_j}$$

where $a^*{}_k$ is the k-th component of α^i.

§ 6.5 Expanding a Markov chain. In the last two sections we showed that under certain conditions a Markov chain would, by lumping states together, be reduced to a smaller chain which gave interesting information about the original chain. By this process we obtained a more manageable chain at the sacrifice of obtaining less precise information. In this section we shall show that it is possible to go in the other direction. That is, to obtain from a Markov chain a larger chain which gives more detailed information about the process being considered. We shall base the presentation on results obtained by S. Hudson in his senior thesis at Dartmouth College.

Consider now a Markov chain with states $s_1, s_2, \ldots, s_r$. We form a new Markov chain, called the *expanded process*, as follows. A state