8.6 STOCHASTIC COMPLEMENTATION

When the concept of Perron complementation is applied to irreducible stochastic matrices as discussed in §8.4 (p. 687), some interesting and useful aspects of the theory of Markov chains are produced. In particular, the Perron complementation idea applied to Markov chains results in a technique for reducing a chain with a large number of states to a smaller chain without losing important characteristics.

Consider an n-state irreducible Markov chain, and let

$$\mathbf{P} = \begin{pmatrix} \mathbf{P}_{11} & \mathbf{P}_{12} & \cdots & \mathbf{P}_{1k} \\ \mathbf{P}_{21} & \mathbf{P}_{22} & \cdots & \mathbf{P}_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{P}_{k1} & \mathbf{P}_{k2} & \cdots & \mathbf{P}_{kk} \end{pmatrix} \quad \text{(with square diagonal blocks)} \tag{8.6.1}$$

be a partition of the associated transition probability matrix. We know that $\mathbf{P}$ is an irreducible stochastic matrix with $\rho(\mathbf{P}) = 1$ (p. 689), so the associated Perron complements are given by

$$\mathbf{S}_i = \mathbf{P}_{ii} + \mathbf{P}_{i\star}(\mathbf{I} - \mathbf{P}_i^\star)^{-1}\mathbf{P}_{\star i}.$$

As we will see, these complements $\mathbf{S}_i$ have additional stochastic properties, so they are alternately referred to as ***stochastic complements*** in the context of Markov chains. Properties (8.5.6)–(8.5.8) on p. 707 guarantee that each $\mathbf{S}_i$ is also a nonnegative irreducible matrix with $\rho(\mathbf{S}_i) = 1$. Furthermore, if $\mathbf{e}$ denotes a column of 1's whose size is determined by the context, then

$$\begin{aligned} \mathbf{Pe} = \mathbf{e} \quad &\Longrightarrow \quad \mathbf{P}_{ii}\mathbf{e} + \mathbf{P}_{i\star}\mathbf{e} = \mathbf{e} \quad \text{and} \quad \mathbf{P}_{\star i}\mathbf{e} + \mathbf{P}_i^\star\mathbf{e} = \mathbf{e} \\ &\Longrightarrow \quad \mathbf{P}_{ii}\mathbf{e} + \mathbf{P}_{i\star}\mathbf{e} = \mathbf{e} \quad \text{and} \quad \mathbf{e} = (\mathbf{I} - \mathbf{P}_i^\star)^{-1}\mathbf{P}_{\star i}\mathbf{e} \\ &\Longrightarrow \quad \mathbf{S}_i\mathbf{e} = \mathbf{e}. \end{aligned}$$

In other words, *every stochastic complement* $\mathbf{S}_i$ *is itself the transition probability matrix of some smaller Markov chain.* To understand the relationship between the smaller chain defined by $\mathbf{S}_i$ and the parent chain associated with $\mathbf{P}$, consider the simpler (but equivalent) situation where the set of states $\{1, 2, \ldots, n\}$ is partitioned into two clusters,

$$\mathcal{S}_1 = \{1, 2, \ldots, r\} \quad \text{and} \quad \mathcal{S}_2 = \{r+1, r+2, \ldots, n\},$$

so that

$$\mathbf{P} = \begin{array}{c} 1 \\ \vdots \\ r \\ r+1 \\ \vdots \\ n \end{array} \overset{\begin{array}{cc} 1 \cdots r & r+1 \cdots n \end{array}}{\left(\begin{array}{c|c} \mathbf{P}_{11} & \mathbf{P}_{12} \\ \hline \mathbf{P}_{21} & \mathbf{P}_{22} \end{array}\right)}, \quad \text{and} \quad \begin{aligned} \mathbf{S}_1 &= \mathbf{P}_{11} + \mathbf{P}_{12}(\mathbf{I} - \mathbf{P}_{22})^{-1}\mathbf{P}_{21}, \\ \mathbf{S}_2 &= \mathbf{P}_{22} + \mathbf{P}_{21}(\mathbf{I} - \mathbf{P}_{11})^{-1}\mathbf{P}_{12}. \end{aligned} \tag{8.6.2}$$

Let's focus on one of these complements—say the second one—and let's interpret the (i,j)-entry $[\mathbf{S}_2]_{ij} = [\mathbf{P}_{22}]_{ij} + [\mathbf{P}_{21}(\mathbf{I}-\mathbf{P}_{11})^{-1}\mathbf{P}_{12}]_{ij}$. Notice first that $[\mathbf{P}_{22}]_{ij}$ is simply the probability of moving from state $r+i \in \mathcal{S}_2$ to state $r+j \in \mathcal{S}_2$ in one step, while

$$[\mathbf{P}_{21}(\mathbf{I}-\mathbf{P}_{11})^{-1}\mathbf{P}_{12}]_{ij} = \sum_{k=1}^{r} [\mathbf{P}_{21}]_{ik}[(\mathbf{I}-\mathbf{P}_{11})^{-1}\mathbf{P}_{12}]_{kj}.$$

The term $[\mathbf{P}_{21}]_{ik}$ is the probability of moving from $r+i \in \mathcal{S}_2$ to $k \in \mathcal{S}_1$ in one step, while $[(\mathbf{I}-\mathbf{P}_{11})^{-1}\mathbf{P}_{12}]_{kj}$ is the probability of hitting state $r+j \in \mathcal{S}_2$ the first time the chain enters $\mathcal{S}_2$ when the process starts from $k \in \mathcal{S}_1$. This can be seen by considering the states in $\mathcal{S}_2$ to be absorbing so as to artificially force the process to stop as soon as the chain enters $\mathcal{S}_2$. The result in Example 8.4.4 (p. 700) says that $[(\mathbf{I}-\mathbf{P}_{11})^{-1}\mathbf{P}_{12}]_{ij}$ is the probability of entering $\mathcal{S}_2$ at state $r+j$ when the chain starts in $k \in \mathcal{S}_1$. Consequently, $[\mathbf{P}_{21}]_{ik}[(\mathbf{I}-\mathbf{P}_{11})^{-1}\mathbf{P}_{12}]_{kj}$ is the probability of moving directly from $r+i \in \mathcal{S}_2$ to $k \in \mathcal{S}_1$ and then, perhaps after several steps inside of $\mathcal{S}_1$, reentering $\mathcal{S}_2$ at state $r+j$ (without regard to what happened while the process was in $\mathcal{S}_1$). Therefore,

$$[\mathbf{S}_2]_{ij} = [\mathbf{P}_{22}]_{ij} + \sum_{k=1}^{r} [\mathbf{P}_{21}]_{ik}[(\mathbf{I}-\mathbf{P}_{11})^{-1}\mathbf{P}_{12}]_{kj}$$

is probability of moving from $r+i \in \mathcal{S}_2$ to $r+j \in \mathcal{S}_2$ in a single step or else by moving directly from $r+i \in \mathcal{S}_2$ to somewhere inside of $\mathcal{S}_1$ (perhaps staying there for awhile) and then hitting state $r+j$ upon first reentry into $\mathcal{S}_2$. In other words, $\mathbf{S}_2$ is the transition probability matrix for a chain that records the location of the process only when the process is visiting states in $\mathcal{S}_2$, and visits to states in $\mathcal{S}_1$ are simply ignored or *censored out*. This explains the following terminology.

Censored Markov Chains

For an n-state irreducible Markov chain with transition probability matrix $\mathbf{P}$ that is partitioned as in (8.6.1), let $\mathcal{S}$ denote the collection of states that correspond to the row (and column) indices of the i^{th} diagonal block $\mathbf{P}_{ii}$. and let $\overline{\mathcal{S}}$ denote the complementary set of states. The ***censored Markov chain*** associated with $\mathcal{S}$ is defined to be the Markov chain that records the location of the parent chain defined by $\mathbf{P}$ only when the parent chain visits states in $\mathcal{S}$. Visits to states in $\overline{\mathcal{S}}$ are ignored. The transition probability matrix for this censored chain is the stochastic complement

$$\mathbf{S}_i = \mathbf{P}_{ii} + \mathbf{P}_{i\star}(\mathbf{I}-\mathbf{P}_i^{\star})^{-1}\mathbf{P}_{\star i}. \tag{8.6.3}$$

Property (8.5.7) guarantees that every stochastic complement $\mathbf{S}_i$ is an irreducible matrix, so every censored chain is an irreducible Markov chain. Consequently each censored chain has an associated stationary probability distribution, $\mathbf{s}_i^T$, such that

$$\mathbf{s}_i^T\mathbf{S}_i = \mathbf{s}_i^T, \qquad \mathbf{s}_i^T > \mathbf{0}, \quad \text{and} \quad \mathbf{s}_i^T\mathbf{e} = 1 \quad \text{(as summarized on p. 693).}$$

In the language of matrix theory $\mathbf{s}_i^T$ is the left-hand Perron vector for $\mathbf{S}_i$, but in the context of Markov chains $\mathbf{s}_i^T$ is called a ***censored probability distribution***.

To interpret the meaning of a censored distribution, suppose that the state space for an n-state Markov chain is partitioned into clusters as

$$\{1,2,\ldots,n\} = \mathcal{S}_1 \cup \mathcal{S}_2 \cup \cdots \cup \mathcal{S}_k, \quad \text{where} \quad \mathcal{S}_i = \{\sigma_{i1}, \sigma_{i2}, \ldots, \sigma_{in_i}\}, \tag{8.6.4}$$

and partition the t^{th} step distribution and the stationary distribution in accord with (8.6.4) as

$$\mathbf{p}^T(t) = \big(\mathbf{p}_1^T(t)\,|\,\mathbf{p}_2^T(t)\,|\,\cdots\,|\,\mathbf{p}_k^T(t)\big) \quad \text{and} \quad \boldsymbol{\pi}^T = \big(\boldsymbol{\pi}_1^T\,|\,\boldsymbol{\pi}_2^T\,|\,\cdots\,|\,\boldsymbol{\pi}_k^T\big). \tag{8.6.5}$$

To ensure that limits exist assume the chain is primitive (p. 693). Let X_t be the state of the chain after the t^{th} step, and let Y_t be the cluster that contains X_t after the t^{th} step. The probability of being in state σ_{ij} (the j^{th} state of the i^{th} cluster) after t steps is

$$P(X_t = \sigma_{ij}) = \big[\mathbf{p}_i^T(t)\big]_j \quad \text{(the } j^{th} \text{ component of } \mathbf{p}_i^T(t)\text{),}$$

and the limiting probability of being in σ_{ij} is

$$\lim_{t\to\infty} P(X_t = \sigma_{ij}) = \lim_{t\to\infty}\big[\mathbf{p}_i^T(t)\big]_j = \big[\boldsymbol{\pi}_i^T\big]_j \quad \text{(the } j^{th} \text{ component of } \boldsymbol{\pi}_i^T\text{).}$$

Similarly, the probability of being inside cluster $\mathcal{S}_i$ after t steps is

$$P(Y_t = i) = \mathbf{p}_i^T(t)\mathbf{e},$$

and the limiting probability of being somewhere in $\mathcal{S}_i$ is

$$\lim_{t\to\infty} P(Y_t = i) = \lim_{t\to\infty}\mathbf{p}_i^T(t)\mathbf{e} = \boldsymbol{\pi}_i^T\mathbf{e}. \tag{8.6.6}$$

Since $\boldsymbol{\pi}^T$ the left-hand Perron vector for the transition probability matrix $\mathbf{P}$, it follows from the left-hand interpretation of (8.5.10) that the j^{th} component of i^{th} censored distribution $\mathbf{s}_i^T$ with respect to the partition (8.6.4) is

$$\big[\mathbf{s}_i^T\big]_j = \frac{\big[\boldsymbol{\pi}_i^T\big]_j}{\boldsymbol{\pi}_i^T\mathbf{e}} = \lim_{t\to\infty}\frac{\big[\mathbf{p}_i^T(t)\big]_j}{\mathbf{p}_i^T(t)\mathbf{e}} = \lim_{t\to\infty}\frac{P(X_t = \sigma_{ij})}{P(Y_t = i)} = \lim_{t\to\infty} P(X_t = \sigma_{ij}\,|\,Y_t = i).$$

In other words, $\big[\mathbf{s}_i^T\big]_j$ is the limiting conditional probability of being in σ_{ij} given that the process is somewhere in $\mathcal{S}_i$. Below is a summary.

Censored Probability Distributions

For an n-state irreducible Markov chain whose transition probability matrix $\mathbf{P}$, stationary distribution $\boldsymbol{\pi}^T = (\boldsymbol{\pi}_1^T \,|\, \boldsymbol{\pi}_2^T \,|\, \cdots \,|\, \boldsymbol{\pi}_k^T)$, and state space are partitioned according to

$$\{1, 2, \ldots, n\} = \mathcal{S}_1 \cup \mathcal{S}_2 \cup \cdots \cup \mathcal{S}_k \quad \text{where} \quad \mathcal{S}_i = \{\sigma_{i1}, \sigma_{i2}, \ldots, \sigma_{in_i}\},$$

the ***censored probability distributions*** are the stationary distributions $\mathbf{s}_i^T$ of the censored Markov chains defined by the stochastic complements $\mathbf{S}_i$ given in (8.6.3) so that $\mathbf{s}_i^T\mathbf{S}_i = \mathbf{s}_i^T$, where $\mathbf{s}_i^T > \mathbf{0}$ and $\mathbf{s}_i^T\mathbf{e} = 1$. Censored distributions have the following additional properties.

- $\mathbf{s}_i^T = \boldsymbol{\pi}_i^T / \boldsymbol{\pi}_i^T\mathbf{e}$ for each $i = 1, 2, \ldots k$. (8.6.7)
- If $\mathbf{P}$ is primitive, then the j^{th} component of $\mathbf{s}_i^T$ is the limiting conditional probability of being in the j^{th} state of cluster $\mathcal{S}_i$ given that the process is somewhere in $\mathcal{S}_i$. In other words,

$$\left[\mathbf{s}_i^T\right]_j = \lim_{t\to\infty} P(X_t = \sigma_{ij} \,|\, Y_t = i),$$

where X_t and Y_t are the respective state and cluster number of the chain after the t^{th} step.

Now let's specialize the coupling theorem for Perron complements given on p. 709 to Markov chains. Vectors are on the left-hand side of matrices for Markov chain applications, so, for the partition of $\mathbf{P}$ in (8.6.1) that corresponds to the partition of the state space in (8.6.4), the coupling matrix on p. 709 takes the form

$$\mathbf{A} = \begin{pmatrix} \mathbf{s}_1^T\mathbf{P}_{11}\mathbf{e} & \cdots & \mathbf{s}_1^T\mathbf{P}_{1k}\mathbf{e} \\ \vdots & \ddots & \vdots \\ \mathbf{s}_k^T\mathbf{P}_{k1}\mathbf{e} & \cdots & \mathbf{s}_k^T\mathbf{P}_{kk}\mathbf{e} \end{pmatrix} = \begin{pmatrix} \mathbf{s}_1^T & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{s}_k^T \end{pmatrix} \begin{pmatrix} \mathbf{P}_{11} & \cdots & \mathbf{P}_{1k} \\ \vdots & \ddots & \vdots \\ \mathbf{P}_{k1} & \cdots & \mathbf{P}_{kk} \end{pmatrix} \begin{pmatrix} \mathbf{e} & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{e} \end{pmatrix}$$

$$= \mathbf{L}_{k\times n}\mathbf{P}_{n\times n}\mathbf{R}_{n\times k}, \tag{8.6.8}$$

where the $\mathbf{s}_i^T$'s in $\mathbf{L}$ are the censored distributions, and the $\mathbf{e}$'s in $\mathbf{R}$ are columns of 1's of appropriate size. (We switched the notation for the coupling matrix from $\mathbf{C}$ to $\mathbf{A}$ for reasons that soon will be apparent.) Remarkably, $\mathbf{A}$ also defines an irreducible Markov chain—but this chain has only k states. The nonnegativity and irreducibility of $\mathbf{A}$ are guaranteed by the coupling theorem on p. 709, and $\mathbf{A}$ is stochastic because

$$\mathbf{Ae} = \mathbf{LPRe} = \mathbf{LPe} = \mathbf{Le} = \mathbf{e}.$$

To understand the nature of the chain defined by $\mathbf{A}$ along with its stationary distribution $\boldsymbol{\alpha}^T$, let's interpret the individual entries $a_{ij} = \mathbf{s}_i^T \mathbf{P}_{ij}\mathbf{e}$ in $\mathbf{A}$ as probabilities. As before, let X_t and Y_t be the respective state and cluster number of the chain after the t^{th} step, and let $\wedge$ and $\vee$ denote *AND* and *OR*, respectively.

Given that the process is in cluster $\mathcal{S}_i$ after t steps, consider the the probability of moving to cluster $\mathcal{S}_j$ on the next step. In other words, consider

$$P(Y_{t+1} = j \,|\, Y_t = i) = \frac{P(Y_t = i \wedge Y_{t+1} = j)}{P(Y_t = i)}. \tag{8.6.9}$$

To determine this conditional probability, suppose that

$$\mathbf{P} = \begin{pmatrix} \mathbf{P}_{11} & \cdots & \mathbf{P}_{1k} \\ \vdots & \ddots & \vdots \\ \mathbf{P}_{k1} & \cdots & \mathbf{P}_{kk} \end{pmatrix}, \quad \mathbf{p}^T(t) = (\, \mathbf{p}_1^T(t) \,|\, \cdots \,|\, \mathbf{p}_k^T(t)\,), \quad \text{and} \quad \boldsymbol{\pi}^T = (\, \boldsymbol{\pi}_1^T \,|\, \cdots \,|\, \boldsymbol{\pi}_k^T \,)$$

are partitioned in accord with (8.6.4), and compute the numerator in (8.6.9) as

$$\begin{aligned} &P(Y_t = i \wedge Y_{t+1} = j) \\ &\qquad = P\big([X_t = \sigma_{i1} \vee \cdots \vee X_t = \sigma_{in_i}] \wedge [X_{t+1} = \sigma_{j1} \vee \cdots \vee X_{t+1} = \sigma_{jn_j}]\big) \\ &\qquad = P\big([X_t = \sigma_{i1} \wedge X_{t+1} = \sigma_{j1}] \vee \cdots \vee [X_t = \sigma_{in_i} \wedge X_{t+1} = \sigma_{jn_j}]\big) \\ &\qquad = \sum_{g=1}^{n_i}\sum_{h=1}^{n_j} P(X_t = \sigma_{ig} \wedge X_{t+1} = \sigma_{jh}) \\ &\qquad = \sum_{g=1}^{n_i}\sum_{h=1}^{n_j} P(X_t = \sigma_{ig})\, P(X_{t+1} = \sigma_{jh} \,|\, X_t = \sigma_{ig}) \\ &\qquad = \sum_{g=1}^{n_i} \big[\mathbf{p}_i^T(t)\big]_g \sum_{h=1}^{n_j} [\mathbf{P}_{ij}]_{gh} = \sum_{g=1}^{n_i} \big[\mathbf{p}_i^T(t)\big]_g [\mathbf{P}_{ij}\mathbf{e}]_g \\ &\qquad = \mathbf{p}_i^T(t)\mathbf{P}_{ij}\mathbf{e}. \end{aligned}$$

The denominator in (8.6.9) is $P(Y_t = i) = \mathbf{p}_i^T(t)\mathbf{e}$, and thus

$$P(Y_{t+1} = j \,|\, Y_t = i) = \frac{\mathbf{p}_i^T(t)\mathbf{P}_{ij}\mathbf{e}}{\mathbf{p}_i^T(t)\mathbf{e}}. \tag{8.6.10}$$

It follows from (8.6.7) on p. 715 that[93]

$$\mathbf{s}_i^T = \frac{\boldsymbol{\pi}_i^T}{\boldsymbol{\pi}_i^T\mathbf{e}} = \lim_{t\to\infty} \frac{\mathbf{p}_i^T(t)}{\mathbf{p}_i^T(t)\mathbf{e}},$$

[93] For these limits to exist, $\mathbf{P}$ must be assumed to be primitive.

and therefore, by (8.6.10), the entries in $\mathbf{A}$ are given by

$$a_{ij} = \mathbf{s}_i^T \mathbf{P}_{ij}\mathbf{e} = \lim_{t\to\infty} \frac{\mathbf{p}_i^T(t)\mathbf{P}_{ij}\mathbf{e}}{\mathbf{p}_i^T(t)\mathbf{e}} = \lim_{t\to\infty} P(Y_{t+1} = j \mid Y_t = i). \tag{8.6.11}$$

An irreducible chain is said to be in ***equilibrium*** at time (step) t if the process is at steady state in the sense that $\mathbf{p}^T(t) = \boldsymbol{\pi}^T$. Consequently, (8.6.11) means that a_{ij} *is the transition probability of moving from cluster* $\mathcal{S}_i$ *to cluster* $\mathcal{S}_j$ *when the chain is in equilibrium.* Below is a summary.

Aggregation Theorem for Markov Chains

An irreducible Markov chain whose states are partitioned into k clusters

$$\{1, 2, \ldots, n\} = \mathcal{S}_1 \cup \mathcal{S}_2 \cup \cdots \cup \mathcal{S}_k$$

can be compressed into a smaller k-state ***aggregated chain*** whose states are the individual clusters $\mathcal{S}_i$.

- The transition probability matrix $\mathbf{A}$ of the aggregated chain is the coupling matrix described on p. 709. That is,

$$\mathbf{A} = \begin{pmatrix} \mathbf{s}_1^T\mathbf{P}_{11}\mathbf{e} & \cdots & \mathbf{s}_1^T\mathbf{P}_{1k}\mathbf{e} \\ \vdots & \ddots & \vdots \\ \mathbf{s}_k^T\mathbf{P}_{k1}\mathbf{e} & \cdots & \mathbf{s}_k^T\mathbf{P}_{kk}\mathbf{e} \end{pmatrix}_{k\times k}, \tag{8.6.12}$$

  where $\mathbf{P}_{ij}$ is the (i, j) block in the partitioned transition matrix $\mathbf{P}$ of the unaggregated chain, and $\mathbf{s}_i^T$ is the censored distribution of the i^{th} stochastic complement derived from $\mathbf{P}$.

- If Y_t is the cluster that the unaggregated chain occupies after t steps, then, for primitive chains, the ***aggregated transition probability*** $a_{ij} = \mathbf{s}_i^T\mathbf{P}_{ij}\mathbf{e}$ can be expressed as

$$a_{ij} = \lim_{t\to\infty} P(Y_{t+1} = j \mid Y_t = i).$$

In other words, *transitions between states in the aggregated chain correspond to transitions between clusters in the unaggregated chain when the unaggregated chain is in equilibrium.*

The utility of aggregation in Markov chains is illustrated in the following example.

Example 8.6.1

Limiting Aggregation Probabilities. Consider a large primitive chain that is partitioned into k clusters $\{1, 2, \ldots, n\} = \mathcal{S}_1 \cup \mathcal{S}_2 \cup \cdots \cup \mathcal{S}_k$, and let the associated transition probability matrix $\mathbf{P}$ be given by (8.6.1).

Problem: Without directly computing stationary probabilities for the chain, determine the eventual probability that the chain is somewhere inside cluster $\mathcal{S}_i$ (the individual state in $\mathcal{S}_i$ that the process might eventually occupy is irrelevant). In other words, if Y_t is the cluster in which the process resides after t steps, the problem is to determine $\lim_{t\to\infty} P(Y_t = i)$.

Solution: This is trivial if the stationary probabilities are known because, as pointed out in (8.6.6), if $\mathbf{p}^T(t) = \left(\mathbf{p}_1^T(t) \mid \cdots \mid \mathbf{p}_k^T(t)\right)$ and $\boldsymbol{\pi}^T = \left(\boldsymbol{\pi}_1^T \mid \cdots \mid \boldsymbol{\pi}_k^T\right)$, then the limiting probability of being somewhere in $\mathcal{S}_i$ is

$$\alpha_i = \lim_{t\to\infty} P(Y_t = i) = \lim_{t\to\infty} \mathbf{p}_i^T(t)\mathbf{e} = \boldsymbol{\pi}_i^T \mathbf{e}. \tag{8.6.13}$$

But computing all of $\boldsymbol{\pi}^T$ just to find $\boldsymbol{\pi}_i^T$ can be wasted effort. Since transitions in the aggregated chain correspond to transitions between the clusters $\mathcal{S}_i$ in the unaggregated chain at equilibrium, we expect the i^{th} component of stationary distribution for the aggregated chain to be the limiting probability of being in $\mathcal{S}_i$, and this is true.

- *The stationary distribution of the aggregated chain defined by* $\mathbf{A}$ *in* (8.6.12) *is* $\boldsymbol{\alpha}^T = (\alpha_1, \alpha_2, \ldots, \alpha_k)$, *where* $\alpha_i = \boldsymbol{\pi}_i^T\mathbf{e} = \lim_{t\to\infty} P(Y_t = i)$. *Consequently, the solution to the problem posed above is to form* $\mathbf{A}$ *and compute its* i^{th} *stationary probability.*

Proof. Clearly, $\boldsymbol{\alpha}^T > \mathbf{0}$ and $\boldsymbol{\alpha}^T\mathbf{e} = 1$. To see that $\boldsymbol{\alpha}^T\mathbf{A} = \boldsymbol{\alpha}^T$, write

$$\mathbf{A} = \begin{pmatrix} \mathbf{s}_1^T & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{s}_k^T \end{pmatrix} \begin{pmatrix} \mathbf{P}_{11} & \cdots & \mathbf{P}_{1k} \\ \vdots & \ddots & \vdots \\ \mathbf{P}_{k1} & \cdots & \mathbf{P}_{kk} \end{pmatrix} \begin{pmatrix} \mathbf{e} & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{e} \end{pmatrix} = \mathbf{LPR},$$

and recall from p. 715 that

$$\mathbf{s}_i^T = \frac{\boldsymbol{\pi}_i^T}{\boldsymbol{\pi}_i^T\mathbf{e}} = \frac{\boldsymbol{\pi}_i^T}{\alpha_i}$$

to conclude that $\boldsymbol{\alpha}^T\mathbf{L} = \boldsymbol{\pi}^T$, and thus

$$\boldsymbol{\alpha}^T\mathbf{A} = \boldsymbol{\alpha}^T\mathbf{LPR} = \boldsymbol{\pi}^T\mathbf{PR} = \boldsymbol{\pi}^T\mathbf{R} = \boldsymbol{\alpha}^T. \quad \blacksquare$$

Note: The primitivity of $\mathbf{P}$ implies the primitivity of $\mathbf{A}$ (Exercise 8.5.1), so limiting probabilities in the aggregated chain exist. Consequently, if the t^{th} step distribution vector for the aggregated chain is $\mathbf{a}^T(t) = (a_1(t), a_2(t), \ldots, a_k(t))$, then

$$\lim_{t\to\infty} a_i(t) = \alpha_i = \lim_{t\to\infty} P(Y_t = i).$$

When interpreted in the context of Markov chains, the coupling theorem on p. 709 represents an expansion or *disaggregation process.* Below is the formal statement of the disaggregation theorem.

Disaggregation in Markov Chains

Consider an irreducible Markov chain along with the associated aggregated chain for which the respective transition probability matrices are

$$\mathbf{P} = \begin{pmatrix} \mathbf{P}_{11} & \mathbf{P}_{12} & \cdots & \mathbf{P}_{1k} \\ \mathbf{P}_{21} & \mathbf{P}_{22} & \cdots & \mathbf{P}_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{P}_{k1} & \mathbf{P}_{k2} & \cdots & \mathbf{P}_{kk} \end{pmatrix}_{n\times n} \quad \text{and} \quad \mathbf{A} = \begin{pmatrix} \mathbf{s}_1^T\mathbf{P}_{11}\mathbf{e} & \cdots & \mathbf{s}_1^T\mathbf{P}_{1k}\mathbf{e} \\ \vdots & \ddots & \vdots \\ \mathbf{s}_k^T\mathbf{P}_{k1}\mathbf{e} & \cdots & \mathbf{s}_k^T\mathbf{P}_{kk}\mathbf{e} \end{pmatrix}_{k\times k},$$

where $\mathbf{s}_i^T$ is the i^{th} censored distribution (the stationary distribution of the i^{th} stochastic complement $\mathbf{S}_i = \mathbf{P}_{ii} + \mathbf{P}_{i\star}(\mathbf{I} - \mathbf{P}_i^\star)^{-1}\mathbf{P}_{\star i}$). If $\boldsymbol{\alpha}^T = (\alpha_1, \alpha_2, \ldots, \alpha_k)$ is the stationary distribution of the aggregated chain defined by $\mathbf{A}$, then the stationary distribution for the unaggregated chain defined by $\mathbf{P}$ is

$$\boldsymbol{\pi}^T = (\alpha_1\mathbf{s}_1^T \mid \alpha_2\mathbf{s}_2^T \mid \cdots \alpha_k\mathbf{s}_k^T).$$

In other words, the censored distributions $\mathbf{s}_i^T$ can be pasted together to form the global distribution $\boldsymbol{\pi}^T$, and the α_i's provide the glue to do the job.

It's clear that disaggregation as stated above can serve as an algorithm for computing the stationary probabilities of any irreducible chain. But while the aggregation/disaggregation results are beautiful theoretical theorems, their straightforward implementation usually doesn't result in a computational advantage over more standard methods. Computing the stochastic complements $\mathbf{S}_i$ in order to determine the censored distributions $\mathbf{s}_i^T$ is generally a computationally intensive task, so, as far as computation is concerned, the goal is to somehow exploit special structure exhibited by the chain to judiciously implement the aggregation/disaggregation procedure. The following situation is an example of a structure that is common in large-scale Markov chain applications. (The section on *nearly uncoupled chains* will be added at a later date.)