

Chapter 11

Sensitivity Analysis of Discrete Markov Chains



11.1 Introduction

As we have seen repeatedly, Markov chains are often used as mathematical models of demographic (as well as other natural) phenomena, with transition probabilities defined in terms of parameters that are of interest in the scientific question at hand. Sensitivity analysis is an important way to quantify the effects of changes in these parameters on the behavior of the chain. This chapter revisits, in a more rigorous way, some of the quantities already explored for absorbing Markov chains (Chaps. 4, 5, and 6). It will also consider ergodic Markov chains (in which no absorbing states exist), and calculate the sensitivity of the stationary distribution and measures of the rate of convergence.

Perturbation (or sensitivity) analysis is a long-standing problem in the theory of Markov chains (Schweitzer 1968; Conlisk 1985; Golub and Meyer 1986; Funderlic and Meyer 1986; Seneta 1988, 1993; Meyer 1994; Cho and Meyer 2000; Mitrophanov 2003, 2005; Mitrophanov et al. 2005; Kirkland et al. 2008). When Markov chains are applied as models of physical, biological, or social systems, they are often defined as functions of parameters that have substantive meaning.

Chapter 11 is modified, under the terms of a Journal Publishing Agreement with Elsevier Publishers, from: Caswell, H. Sensitivity analysis of discrete Markov chains via matrix calculus. *Linear Algebra and its Applications* 438:1727–1745. ©Elsevier.

11.2 Absorbing Chains

The transition matrix for a discrete-time absorbing chain can be written

$$\mathbf{P} = \left(\begin{array}{c|c} \mathbf{U} & \mathbf{0} \\ \hline \mathbf{M} & \mathbf{I} \end{array} \right) \quad (11.1)$$

where \mathbf{U} , of dimension $s \times s$, is the transition matrix among the s transient states, and \mathbf{M} , of dimension $a \times s$, contains probabilities of transition from the transient states to the a absorbing states. Assume that the spectral radius of \mathbf{U} is strictly less than 1. Because we are concerned here with absorption, but not what happens after, we ignore transitions among absorbing states; hence the identity matrix ($a \times a$) in the lower right corner. The matrices $\mathbf{U}[\boldsymbol{\theta}]$ and $\mathbf{M}[\boldsymbol{\theta}]$ are functions of a vector of parameters. We assume that $\boldsymbol{\theta}$ varies over some set in which the column sums of \mathbf{P} are 1 and the spectral radius of \mathbf{U} is strictly less than one.

11.2.1 Occupancy: Visits to Transient States

Let v_{ij} be the number of visits to transient state i , prior to absorption, by an individual starting in transient state j . The expectations of the v_{ij} are entries of the fundamental matrix $\mathbf{N} = \mathbf{N}_1 = (E(\eta_{ij}))$:

$$\mathbf{N} = (\mathbf{I} - \mathbf{U})^{-1} \quad (11.2)$$

(e.g., Kemeny and Snell 1960; Iosifescu 1980). Let $\mathbf{N}_k = (E(\eta_{ij}^k))$ be a matrix containing the k th moments about the origin of the v_{ij} . The first several of these matrices are (Iosifescu 1980, Thm. 3.1)

$$\mathbf{N}_1 = (\mathbf{I} - \mathbf{U})^{-1} \quad (11.3)$$

$$\mathbf{N}_2 = (2\mathbf{N}_{\text{dg}} - \mathbf{I}) \mathbf{N}_1 \quad (11.4)$$

$$\mathbf{N}_3 = (6\mathbf{N}_{\text{dg}}^2 - 6\mathbf{N}_{\text{dg}} + \mathbf{I}) \mathbf{N}_1 \quad (11.5)$$

$$\mathbf{N}_4 = (24\mathbf{N}_{\text{dg}}^3 - 36\mathbf{N}_{\text{dg}}^2 + 14\mathbf{N}_{\text{dg}} - \mathbf{I}) \mathbf{N}_1. \quad (11.6)$$

Theorem 11.2.1 *Let \mathbf{N}_k be the matrix of k th moments of the v_{ij} , as given by (11.3), (11.4), (11.5), and (11.6). The sensitivities of \mathbf{N}_k , for $k = 1, \dots, 4$ are*

$$d\text{vec } \mathbf{N}_1 = (\mathbf{N}_1^\top \otimes \mathbf{N}_1) d\text{vec } \mathbf{U} \quad (11.7)$$

$$d\text{vec } \mathbf{N}_2 = \left[2(\mathbf{I} \otimes \mathbf{N}_{\text{dg}}) - \mathbf{I}_{s^2} \right] d\text{vec } \mathbf{N}_1 + 2(\mathbf{N}^\top \otimes \mathbf{I}) d\text{vec } \mathbf{N}_{\text{dg}} \quad (11.8)$$

$$d\text{vec } \mathbf{N}_3 = \left[\mathbf{I} \otimes (6\mathbf{N}_{\text{dg}}^2 - 6\mathbf{N}_{\text{dg}} + \mathbf{I}) \right] d\text{vec } \mathbf{N}_1 \\ + \left[6(\mathbf{N}^\top \mathbf{N}_{\text{dg}} \otimes \mathbf{I}) + 6(\mathbf{N}^\top \otimes \mathbf{N}_{\text{dg}}) - 6(\mathbf{N}^\top \otimes \mathbf{I}) \right] d\text{vec } \mathbf{N}_{\text{dg}} \quad (11.9)$$

$$d\text{vec } \mathbf{N}_4 = \left[\mathbf{I} \otimes (24\mathbf{N}_{\text{dg}}^3 - 36\mathbf{N}_{\text{dg}}^2 + 14\mathbf{N}_{\text{dg}} - \mathbf{I}) \right] d\text{vec } \mathbf{N}_1 \\ + \left[24(\mathbf{N}^\top \mathbf{N}_{\text{dg}}^2 \otimes \mathbf{I}) + 24(\mathbf{N}^\top \mathbf{N}_{\text{dg}} \otimes \mathbf{N}_{\text{dg}}) + 24(\mathbf{N}^\top \otimes \mathbf{N}_{\text{dg}}^2) \right. \\ \left. - 36(\mathbf{N}^\top \mathbf{N}_{\text{dg}} \otimes \mathbf{I}) - 36(\mathbf{N}^\top \otimes \mathbf{N}_{\text{dg}}) + 14(\mathbf{N}^\top \otimes \mathbf{I}) \right] d\text{vec } \mathbf{N}_{\text{dg}} \quad (11.10)$$

where (see Sect. 2.8)

$$d\mathbf{N}_{\text{dg}} = \mathbf{I} \circ d\mathbf{N}_1 \quad (11.11)$$

$$d\text{vec } \mathbf{N}_{\text{dg}} = \mathcal{D}(\text{vec } \mathbf{I}) d\text{vec } \mathbf{N}_1. \quad (11.12)$$

Proof The result (11.7) is derived in Caswell (2006, Section 3.1). For $k > 1$, and considering \mathbf{N}_k as a function of \mathbf{N}_1 and \mathbf{N}_{dg} , the total differential of \mathbf{N}_k is

$$d\text{vec } \mathbf{N}_k = \frac{\partial \text{vec } \mathbf{N}_k}{\partial \text{vec } \mathbf{N}_1} d\text{vec } \mathbf{N}_1 + \frac{\partial \text{vec } \mathbf{N}_k}{\partial \text{vec } \mathbf{N}_{\text{dg}}} d\text{vec } \mathbf{N}_{\text{dg}}. \quad (11.13)$$

The two terms of (11.13) are the partial differentials of $\text{vec } \mathbf{N}_k$, obtained by taking differentials treating only \mathbf{N}_1 or only \mathbf{N}_{dg} as variables, respectively. Denote these partial differentials as $\partial_{\mathbf{N}_1}$, $\partial_{\mathbf{N}_1}$ and $\partial_{\mathbf{N}_{\text{dg}}}$ and $\partial_{\mathbf{N}_{\text{dg}}}$. Differentiating \mathbf{N}_2 in (11.4), gives

$$\partial_{\mathbf{N}_1} \mathbf{N}_2 = 2\mathbf{N}_{\text{dg}} (d\mathbf{N}_1) - d\mathbf{N}_1 \quad (11.14)$$

$$\partial_{\mathbf{N}_{\text{dg}}} \mathbf{N}_2 = 2(d\mathbf{N}_{\text{dg}}) \mathbf{N}_1. \quad (11.15)$$

Applying the vec operator gives

$$\partial_{\mathbf{N}_1} \text{vec } \mathbf{N}_2 = \left[2(\mathbf{I} \otimes \mathbf{N}_{\text{dg}}) - \mathbf{I}_{s^2} \right] d\text{vec } \mathbf{N}_1 \quad (11.16)$$

$$\partial_{\mathbf{N}_{\text{dg}}} \text{vec } \mathbf{N}_2 = 2(\mathbf{N}_1^\top \otimes \mathbf{I}) d\text{vec } \mathbf{N}_{\text{dg}}, \quad (11.17)$$

and (11.13) becomes

$$d\text{vec } \mathbf{N}_2 = \left[2(\mathbf{I} \otimes \mathbf{N}_{\text{dg}}) - \mathbf{I}_{s^2} \right] d\text{vec } \mathbf{N}_1 + 2(\mathbf{N}_1^\top \otimes \mathbf{I}) d\text{vec } \mathbf{N}_{\text{dg}}, \quad (11.18)$$

which is (11.8). The derivations of $d\text{vec } \mathbf{N}_3$ and $d\text{vec } \mathbf{N}_4$ follow the same sequence of steps. The details are given in Appendix A. \square

The derivatives of \mathbf{N}_2 , \mathbf{N}_3 , and \mathbf{N}_4 can be used to study the variance, standard deviation, coefficient of variation, skewness, and kurtosis of the number of visits to the transient states (Caswell 2006, 2009, 2011).

11.2.2 Time to Absorption

Let η_j be the time to absorption starting in transient state j and let $\boldsymbol{\eta}_k = E(\eta_1^k, \dots, \eta_s^k)^\top$. The first several of these moments are (Iosifescu 1980, Thm. 3.2)

$$\boldsymbol{\eta}_1^\top = \mathbf{1}^\top \mathbf{N}_1 \quad (11.19)$$

$$\boldsymbol{\eta}_2^\top = \boldsymbol{\eta}_1^\top (2\mathbf{N}_1 - \mathbf{I}) \quad (11.20)$$

$$\boldsymbol{\eta}_3^\top = \boldsymbol{\eta}_1^\top (6\mathbf{N}_1^2 - 6\mathbf{N}_1 + \mathbf{I}) \quad (11.21)$$

$$\boldsymbol{\eta}_4^\top = \boldsymbol{\eta}_1^\top (24\mathbf{N}_1^3 - 36\mathbf{N}_1^2 + 14\mathbf{N}_1 - \mathbf{I}). \quad (11.22)$$

Theorem 11.2.2 *Let $\boldsymbol{\eta}_k$ be the vector of the k th moments of the η_i . The sensitivities of these moment vectors are*

$$d\boldsymbol{\eta}_1 = (\mathbf{I} \otimes \mathbf{1}^\top) d\text{vec } \mathbf{N}_1 \quad (11.23)$$

$$d\boldsymbol{\eta}_2 = (2\mathbf{N}_1^\top - \mathbf{I}) d\boldsymbol{\eta}_1 + 2(\mathbf{I} \otimes \boldsymbol{\eta}_1^\top) d\text{vec } \mathbf{N}_1 \quad (11.24)$$

$$\begin{aligned} d\boldsymbol{\eta}_3 = & (6\mathbf{n}^2 - 6\mathbf{N}_1 + \mathbf{I})^\top d\boldsymbol{\eta}_1 \\ & + \left[6(\mathbf{N}_1^\top \otimes \boldsymbol{\eta}_1^\top) + 6(\mathbf{I} \otimes \boldsymbol{\eta}_1^\top \mathbf{N}_1) - 6(\mathbf{I} \otimes \boldsymbol{\eta}_1^\top) \right] d\text{vec } \mathbf{N}_1 \end{aligned} \quad (11.25)$$

$$\begin{aligned} d\boldsymbol{\eta}_4 = & (24\mathbf{N}_1^3 - 36\mathbf{N}_1^2 + 14\mathbf{N}_1 - \mathbf{I})^\top d\boldsymbol{\eta}_1 \\ & + \left\{ 24 \left[(\mathbf{N}_1^\top)^2 \otimes \boldsymbol{\eta}_1^\top \right] + 24(\mathbf{N}_1^\top \otimes \boldsymbol{\eta}_1^\top \mathbf{N}_1) + 24(\mathbf{I} \otimes \boldsymbol{\eta}_1^\top \mathbf{N}_1^2) \right. \\ & \left. - 36(\mathbf{N}_1^\top \otimes \boldsymbol{\eta}_1^\top) - 36(\mathbf{I} \otimes \boldsymbol{\eta}_1^\top \mathbf{N}_1) + 14(\mathbf{I} \otimes \boldsymbol{\eta}_1^\top) \right\} d\text{vec } \mathbf{N}_1 \end{aligned} \quad (11.26)$$

where $d\text{vec } \mathbf{N}_1$ is given by (11.7).

Proof The derivative of η_1 is obtained (Caswell 2006) by differentiating to get $d\eta_1^\top = \mathbf{1}^\top (d\mathbf{N}_1)$ and then applying the vec operator. For the higher moments, consider the η_k to be functions of η_1 and \mathbf{N}_1 , and write the total differential

$$d\eta_k = \frac{\partial \eta_k}{\partial \eta_1^\top} d\eta_1 + \frac{\partial \eta_k}{\partial \text{vec } \mathbf{N}_1} d\text{vec } \mathbf{N}_1. \quad (11.27)$$

The partial differentials of η_2 with respect to η_1 and \mathbf{N}_1 are

$$\partial_{\eta_1} \eta_2^\top = \left(d\eta_1^\top \right) (2\mathbf{N}_1 - \mathbf{I}) \quad (11.28)$$

$$\partial_{\mathbf{N}_1} \eta_2^\top = 2\eta_1^\top (d\mathbf{N}_1). \quad (11.29)$$

Applying the vec operator gives

$$\partial_{\eta_1} \eta_2 = \left(2\mathbf{N}_1^\top - \mathbf{I} \right) d\eta_1 \quad (11.30)$$

$$\partial_{\mathbf{N}_1} \eta_2 = 2 \left(\mathbf{I} \otimes \eta_1^\top \right) d\text{vec } \mathbf{N}_1 \quad (11.31)$$

which combine according to (11.27) to yield (11.24). The derivations of $d\eta_3$ and $d\eta_4$ follow the same sequence of steps; the details are shown in Appendix A. \square

11.2.3 Number of States Visited Before Absorption

Let $\xi_i \geq 1$ be the number of distinct transient states visited before absorption, and let $\xi_1 = E(\xi)$. Then

$$\xi_1^\top = \mathbf{1}^\top \mathbf{N}_{\text{dg}}^{-1} \mathbf{N}_1 \quad (11.32)$$

(Iosifescu 1980, Sect. 3.2.5), where $\mathbf{N}_{\text{dg}}^{-1} = (\mathbf{N}_{\text{dg}})^{-1}$.

Theorem 11.2.3 *Let $\xi_1 = E(\xi)$. The sensitivity of ξ is*

$$d\xi_1 = \left[- \left(\mathbf{N}_1^\top \otimes \mathbf{1}^\top \right) \left(\mathbf{N}_{\text{dg}}^{-1} \otimes \mathbf{N}_{\text{dg}}^{-1} \right) \mathcal{D}(\text{vec } \mathbf{I}) + \left(\mathbf{I} \otimes \mathbf{1}^\top \mathbf{N}_{\text{dg}}^{-1} \right) \right] d\text{vec } \mathbf{N}_1, \quad (11.33)$$

where $d\text{vec } \mathbf{N}_1$ is given by (11.7).

Proof Differentiating (11.32) yields

$$d\xi_1^\top = \mathbf{1}^\top \left(d\mathbf{N}_{\text{dg}}^{-1} \right) \mathbf{N}_1 + \mathbf{1}^\top \mathbf{N}_{\text{dg}}^{-1} d\mathbf{N}_1. \quad (11.34)$$

Applying the vec operator yields

$$d\xi_1 = \left(\mathbf{N}_1^\top \otimes \mathbf{1}^\top \right) d\text{vec } \mathbf{N}_{\text{dg}}^{-1} + \left(\mathbf{I} \otimes \mathbf{1}^\top \mathbf{N}_{\text{dg}}^{-1} \right) d\text{vec } \mathbf{N}_1. \quad (11.35)$$

Applying (2.82) to $d\text{vec } \mathbf{N}_{\text{dg}}^{-1}$ and using (11.12) for $d\text{vec } \mathbf{N}_{\text{dg}}$ gives

$$d\xi_1 = - \left(\mathbf{N}_1^\top \otimes \mathbf{1}^\top \right) \left(\mathbf{N}_{\text{dg}}^{-1} \otimes \mathbf{N}_{\text{dg}}^{-1} \right) \mathcal{D}(\text{vec } \mathbf{I}) d\text{vec } \mathbf{N}_1 + \left(\mathbf{I} \otimes \mathbf{1}^\top \mathbf{N}_{\text{dg}}^{-1} \right) d\text{vec } \mathbf{N}_1 \quad (11.36)$$

which simplifies to (11.33). \square

11.2.4 Multiple Absorbing States and Probabilities of Absorption

When the chain includes $a > 1$ absorbing states, the entry m_{ij} of the $a \times s$ submatrix \mathbf{M} in (11.1) is the probability of transition from transient state j to absorbing state i . The result of the competing risks of absorption is a set of probabilities $b_{ij} = P[\text{absorption in } i \mid \text{starting in } j]$ for $i = 1, \dots, a$ and $j = 1, \dots, s$. The matrix $\mathbf{B} = (b_{ij}) = \mathbf{M}\mathbf{N}_1$ (Iosifescu 1980, Thm. 3.3).

Theorem 11.2.4 *Let $\mathbf{B} = \mathbf{M}\mathbf{N}_1$ be the matrix of absorption probabilities. Then*

$$d\text{vec } \mathbf{B} = \left(\mathbf{N}_1^\top \otimes \mathbf{I} \right) d\text{vec } \mathbf{M} + \left(\mathbf{N}_1^\top \otimes \mathbf{B} \right) d\text{vec } \mathbf{U}. \quad (11.37)$$

Proof Differentiating \mathbf{B} yields

$$d\mathbf{B} = (d\mathbf{M})\mathbf{N}_1 + \mathbf{M}(d\mathbf{N}_1). \quad (11.38)$$

Applying the vec operator gives

$$d\text{vec } \mathbf{B} = \left(\mathbf{N}_1^\top \otimes \mathbf{I} \right) d\text{vec } \mathbf{M} + \left(\mathbf{I} \otimes \mathbf{M} \right) d\text{vec } \mathbf{N}_1. \quad (11.39)$$

Substituting (11.7) for $d\text{vec } \mathbf{N}_1$ and simplifying gives (11.37). \square

Column j of \mathbf{B} is the probability distribution of the eventual absorption state for an individual starting in transient state j . Usually a few of those starting states are of particular interest (e.g., states corresponding to “birth” or to the start of some

process). Let $\mathbf{B}(:, j) = \mathbf{B}\mathbf{e}_j$ denote column j of \mathbf{B} , where \mathbf{e}_j is the j th unit vector of length s . Thus the derivative of $\mathbf{B}(:, j)$ is

$$d\text{vec } \mathbf{B}(:, j) = \left(\mathbf{e}_j^\top \otimes \mathbf{I}_s \right) d\text{vec } \mathbf{B} \tag{11.40}$$

where $d\text{vec } \mathbf{B}$ is given by (11.37). Similarly, row i of \mathbf{B} is $\mathbf{B}(i, :) = \mathbf{e}_i^\top \mathbf{B}$ and

$$d\text{vec } \mathbf{B}(i, :) = \left(\mathbf{I}_s \otimes \mathbf{e}_i^\top \right) d\text{vec } \mathbf{B} \tag{11.41}$$

where \mathbf{e}_i is the i th unit vector of length a .

11.2.5 The Quasistationary Distribution

The quasistationary distribution of an absorbing Markov chain gives the limiting probability distribution, over the set of transient states, of the state of an individual that has yet to be absorbed. Let \mathbf{w} and \mathbf{v} be the right and left eigenvectors associated with the dominant eigenvalue of \mathbf{U} , normalized so that $\|\mathbf{w}\| = \|\mathbf{v}\| = 1$. Darroch and Seneta (1965) defined two quasistationary distributions in terms of \mathbf{w} and \mathbf{v} . The limiting probability distribution of the state of an individual, given that absorption has not yet happened, converges to

$$\mathbf{q}_a = \mathbf{w} \tag{11.42}$$

The limiting probability distribution of the state of an individual, given that absorption has not happened and will not happen for a long time, is

$$\mathbf{q}_b = \frac{\mathbf{w} \circ \mathbf{v}}{\mathbf{w}^\top \mathbf{v}} \tag{11.43}$$

Horvitz and Tuljapurkar (2008) pointed out that the convergence to the quasistationary distribution implies that, in a stage-classified model, mortality eventually becomes independent of age.

Lemma 1 *Let the dominant eigenvalue of \mathbf{U} , guaranteed real and nonnegative by the Perron-Frobenius theorem, satisfy $0 < \lambda < 1$, and let \mathbf{w} and \mathbf{v} be the right and left eigenvectors corresponding to λ , scaled so that $\mathbf{w}^\top \mathbf{v} = 1$. Then*

$$d\mathbf{w} = \left(\lambda \mathbf{I}_s - \mathbf{U} + \mathbf{w} \mathbf{1}^\top \mathbf{U} \right)^{-1} \left[\mathbf{w}^\top \otimes \left(\mathbf{I}_s - \mathbf{w} \mathbf{1}^\top \right) \right] d\text{vec } \mathbf{U} \tag{11.44}$$

$$d\mathbf{v} = \left(\lambda \mathbf{I}_s - \mathbf{U}^\top + \mathbf{v} \mathbf{e}_1^\top \mathbf{U}^\top \right)^{-1} \left[\left(\mathbf{I}_s - \mathbf{v} \mathbf{e}_1^\top \right) \otimes \mathbf{v}^\top \right] d\text{vec } \mathbf{U} \tag{11.45}$$

Proof Equation (11.44) is proven in Caswell (2008, Section 6.1). Equation (11.45) is obtained by treating \mathbf{v} as the right eigenvector of \mathbf{U}^\top . \square

Theorem 11.2.5 *The derivative of the quasistationary distribution \mathbf{q}_a is given by (11.44). The derivative of the quasistationary distribution \mathbf{q}_b is*

$$d\mathbf{q}_b = \frac{1}{\mathbf{v}^\top \mathbf{w}} \left[\left(\mathcal{D}(\mathbf{v}) - \mathbf{q}_b \mathbf{v}^\top \right) d\mathbf{w} + \left(\mathcal{D}(\mathbf{w}) - \mathbf{q}_b \mathbf{w}^\top \right) d\mathbf{v} \right] \quad (11.46)$$

where $d\mathbf{w}$ and $d\mathbf{v}$ are given by (11.44) and (11.45) respectively.

Proof The derivative of \mathbf{q}_a follows from its definition as the scaled right eigenvector of \mathbf{U} . For \mathbf{q}_b , differentiating (11.43) gives

$$d\mathbf{q}_b = \frac{1}{(\mathbf{v}^\top \mathbf{w})^2} \left\{ (\mathbf{v}^\top \mathbf{w}) d(\mathbf{v} \circ \mathbf{w}) - (\mathbf{v} \circ \mathbf{w}) \left[(d\mathbf{v}^\top) \mathbf{w} + \mathbf{v}^\top (d\mathbf{w}) \right] \right\} \quad (11.47)$$

$$= \frac{1}{\mathbf{v}^\top \mathbf{w}} \left[d(\mathbf{v} \circ \mathbf{w}) - \mathbf{q}_b (d\mathbf{v}^\top) \mathbf{w} - \mathbf{q}_b \mathbf{v}^\top (d\mathbf{w}) \right] \quad (11.48)$$

Applying the vec operator gives

$$d\mathbf{q}_b = \frac{1}{\mathbf{v}^\top \mathbf{w}} \left[\mathcal{D}(\mathbf{v}) d\mathbf{w} + \mathcal{D}(\mathbf{w}) d\mathbf{v} - (\mathbf{w}^\top \otimes \mathbf{q}_b) d\mathbf{v} - \mathbf{q}_b \mathbf{v}^\top d\mathbf{w} \right] \quad (11.49)$$

which simplifies to give (11.46). \square

11.3 Life Lost Due to Mortality

The approach here makes it easy to compute the sensitivity of a variety of dependent variables calculated from the Markov chain. As an example of this flexibility, consider a recently developed demographic index, the number of years of life lost due to mortality (Vaupel and Canudas Romo 2003).

The transient states of the chains are age classes, absorption corresponds to death, and absorbing states correspond to age at death. Let μ_i be the mortality rate and $p_i = \exp(-\mu_i)$ the survival probability at age i . The matrix \mathbf{U} has the p_i on the subdiagonal and zeros elsewhere. The matrix \mathbf{M} has $1 - p_i$ on the diagonal and zeros elsewhere. Let $\mathbf{f} = \mathbf{B}(:, 1)$ be the distribution of age at death and $\boldsymbol{\eta}_1$ the vector of expected longevity as a function of age.

A death at age i represents the loss of some number of years of life beyond that age. The expectation of that loss is given by the i th entry of $\boldsymbol{\eta}_1$, and the expected number of years lost over the distribution of age at death is $\eta^\dagger = \boldsymbol{\eta}_1^\top \mathbf{f}$. This quantity also measures the disparity among individuals in longevity (Vaupel and Canudas Romo 2003). If everyone died at the identical age x , \mathbf{f} would be a delta function at x

and further life expectancy at age x would be zero; their product would give $\eta^\dagger = 0$. Declines in discrepancy have accompanied increases in life expectancy observed in developed countries (Edwards and Tuljapurkar 2005; Wilmoth and Horiuchi 1999). Thus it is useful to know how η^\dagger responds to changes in mortality.

Differentiating η^\dagger gives

$$d\eta^\dagger = (d\boldsymbol{\eta}_1^\top) \mathbf{B}\mathbf{e}_1 + \boldsymbol{\eta}_1^\top (d\mathbf{B}) \mathbf{e}_1. \quad (11.50)$$

Applying the vec operator gives

$$d\eta^\dagger = \mathbf{e}_1^\top \mathbf{b}^\top d\boldsymbol{\eta}_1^\top + (\mathbf{e}_1^\top \otimes \boldsymbol{\eta}_1^\top) d\text{vec } \mathbf{B}. \quad (11.51)$$

Substituting (11.23) for $d\boldsymbol{\eta}_1$ and (11.37) for $d\text{vec } \mathbf{B}$ gives

$$\begin{aligned} d\eta^\dagger = & \mathbf{f}^\top (\mathbf{I} \otimes \mathbf{1}^\top) d\text{vec } \mathbf{N}_1 + (\mathbf{e}_1^\top \otimes \boldsymbol{\eta}_1^\top) \\ & \left[(\mathbf{N}_1^\top \otimes \mathbf{I}) d\text{vec } \mathbf{M} + (\mathbf{N}_1^\top \otimes \mathbf{B}) d\text{vec } \mathbf{U} \right] \end{aligned} \quad (11.52)$$

Simplifying and writing derivatives in terms of $\boldsymbol{\mu}$ gives

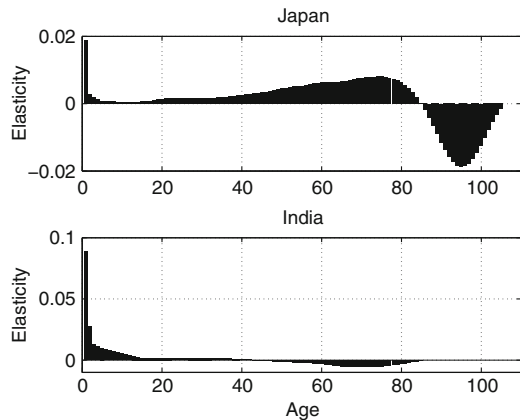
$$\begin{aligned} \frac{d\eta^\dagger}{d\boldsymbol{\mu}^\top} = & \left[\mathbf{f}^\top (\mathbf{N}_1^\top \otimes \boldsymbol{\eta}_1^\top) + (\mathbf{e}_1^\top \mathbf{N}_1^\top \otimes \boldsymbol{\eta}_1^\top \mathbf{B}) \right] \\ & \frac{d\text{vec } \mathbf{U}}{d\boldsymbol{\mu}^\top} + (\mathbf{e}_1^\top \mathbf{N}_1^\top \otimes \boldsymbol{\eta}_1^\top) \frac{d\text{vec } \mathbf{M}}{d\boldsymbol{\mu}^\top} \end{aligned} \quad (11.53)$$

Because mortality rates vary over several orders of magnitude with age, it is useful to present the results as elasticities,

$$\frac{\epsilon \eta^\dagger}{\epsilon \boldsymbol{\mu}^\top} = \frac{1}{\eta^\dagger} \frac{d\eta^\dagger}{d\boldsymbol{\mu}^\top} \mathcal{D}(\boldsymbol{\mu}). \quad (11.54)$$

Figure 11.1 shows these elasticities for two populations chosen to have very different life expectancies: India in 1961, with female life expectancy of 45 years and $\eta^\dagger = 23.9$ years and Japan in 2006, with female life expectancy of 86 years and $\eta^\dagger = 10.1$ years (Human Mortality Database 2016). In both cases, elasticities are positive from birth to some age (≈ 50 for India, ≈ 85 for Japan) and negative thereafter. This implies that reductions in infant and early life mortality would reduce η^\dagger , whereas reductions in old age mortality would increase η^\dagger . Zhang and Vaupel (2009) have shown that the existence of such a critical age is a general property of these models.

Fig. 11.1 The elasticity of mean years of life lost due to mortality, η^\dagger , to changes in age-specific mortality, calculated from the female life tables of India in 1961 and of Japan in 2006. (Data obtained from the Human Mortality Database [2016](#))



11.4 Ergodic Chains

Now let us consider perturbations of an ergodic finite-state Markov chain with an irreducible, primitive, column-stochastic transition matrix \mathbf{P} of dimension $s \times s$. The stationary distribution $\boldsymbol{\pi}$ is given by the right eigenvector, scaled to sum to 1, corresponding to the dominant eigenvalue $\lambda_1 = 1$ of \mathbf{P} . The fundamental matrix of the chain is $\mathbf{Z} = (\mathbf{I} - \mathbf{P} + \boldsymbol{\pi}\mathbf{1}^\top)^{-1}$ (Kemeny and Snell [1960](#)).

We are interested only in perturbations that preserve the column-stochasticity of \mathbf{P} ; i.e., for which \mathbf{P} remains a stochastic matrix. Such perturbations are easily defined when the p_{ij} depend explicitly on a parameter vector $\boldsymbol{\theta}$. However, when the parameters of interest are the p_{ij} themselves, an implicit parameterization must be defined to preserve the stochastic nature of \mathbf{P} under perturbation (Conlisk [1985](#); Caswell [2001](#)). In Sect. [11.4.5](#) we will explore new expressions for two different forms of implicit parameterization.

Previous studies of perturbations of ergodic chains focus almost completely on perturbations of the stationary distribution, and are divided between those focusing on sensitivity as a derivative (e.g., Schweitzer [1968](#); Conlisk [1985](#); Golub and Meyer [1986](#)) and studies focusing on perturbation bounds and condition numbers (Funderlic and Meyer [1986](#); Meyer [1994](#); Seneta [1988](#); Hunter [2005](#); Kirkland [2003](#)); for reviews see Cho and Meyer ([2000](#)) and Kirkland et al. ([2008](#)). The approach here is similar in spirit to that of Schweitzer ([1968](#)), Conlisk ([1985](#)), and Golub and Meyer ([1986](#)), in that we focus on derivatives of Markov chain properties with respect to parameter perturbations, but taking advantage of the matrix calculus approach. We do not consider perturbation bounds here.

11.4.1 The Stationary Distribution

Theorem 11.4.1 *Let $\boldsymbol{\pi}$ be the stationary distribution, satisfying $\mathbf{P}\boldsymbol{\pi} = \boldsymbol{\pi}$ and $\mathbf{1}^\top \boldsymbol{\pi} = 1$. The sensitivity of $\boldsymbol{\pi}$ is*

$$d\boldsymbol{\pi} = \left[\boldsymbol{\pi}^\top \otimes (\mathbf{Z} - \boldsymbol{\pi}\mathbf{1}^\top) \right] d\text{vec } \mathbf{P} \quad (11.55)$$

where \mathbf{Z} is the fundamental matrix of the chain.

Proof The vector $\boldsymbol{\pi}$ is the right eigenvector of \mathbf{P} , scaled to sum to 1. Applying Lemma 1, and noting that $\lambda = 1$ and $\mathbf{1}^\top \mathbf{P} = \mathbf{1}^\top$, gives $d\boldsymbol{\pi} = \mathbf{Z} \left[\boldsymbol{\pi}^\top \otimes (\mathbf{I}_s - \boldsymbol{\pi}\mathbf{1}^\top) \right] d\text{vec } \mathbf{P}$. Noting that $\mathbf{Z}\boldsymbol{\pi} = \boldsymbol{\pi}$ and simplifying the Kronecker products yields (11.55). \square

Based on an analysis of eigenvector sensitivity (Meyer and Stewart 1982), Golub and Meyer (1986) derived an expression for the derivative of $\boldsymbol{\pi}$ to a change in a single element of \mathbf{P} using the group generalized inverse $(\mathbf{I} - \mathbf{P})^\#$ of $\mathbf{I} - \mathbf{P}$. Since $(\mathbf{I} - \mathbf{P})^\# = \mathbf{Z} - \boldsymbol{\pi}\mathbf{1}^\top$ (Golub and Meyer 1986), expression (11.55) is exactly the Golub-Meyer result expressed in matrix calculus notation. Our results here permit sensitivity analysis of functions of $\boldsymbol{\pi}$ using only the chain rule. If $g(\boldsymbol{\pi})$ is a vector- or scalar-valued function of $\boldsymbol{\pi}$, then

$$dg(\boldsymbol{\pi}) = \frac{dg}{d\boldsymbol{\pi}^\top} \frac{d\boldsymbol{\pi}}{d\text{vec } \mathbf{P}} d\text{vec } \mathbf{P} \quad (11.56)$$

Some examples will appear in Sect. 11.5.

11.4.2 The Fundamental Matrix

The fundamental matrix $\mathbf{Z} = (\mathbf{I} - \mathbf{P} + \boldsymbol{\pi}\mathbf{1}^\top)^{-1}$ plays a role in ergodic chains similar to that played by \mathbf{N}_1 in absorbing chains (Kemeny and Snell 1960). It has been extended using generalized inverses (Meyer 1975; Kemeny 1981), but we do not consider those extensions here.

Theorem 11.4.2 *The sensitivity of the fundamental matrix is*

$$d\text{vec } \mathbf{Z} = \left(\mathbf{Z}^\top \otimes \mathbf{Z} \right) \left\{ \mathbf{I}_{s^2} - \left[\mathbf{1}\boldsymbol{\pi}^\top \otimes (\mathbf{Z} - \boldsymbol{\pi}\mathbf{1}^\top) \right] \right\} d\text{vec } \mathbf{P} \quad (11.57)$$

Proof From (2.82),

$$d\text{vec } \mathbf{Z} = - \left(\mathbf{Z}^\top \otimes \mathbf{Z} \right) d\text{vec } \left(\mathbf{I} - \mathbf{P} + \boldsymbol{\pi}\mathbf{1}^\top \right) \quad (11.58)$$

$$= \left(\mathbf{Z}^\top \otimes \mathbf{Z} \right) \left(d\text{vec } \mathbf{P} - (\mathbf{1} \otimes \mathbf{I}_s) d\boldsymbol{\pi} \right) \quad (11.59)$$

Substituting (11.55) for $d\boldsymbol{\pi}$ and simplifying gives (11.57). \square

11.4.3 The First Passage Time Matrix

Let $\mathbf{R} = (r_{ij})$ be the matrix of mean first passage times from j to i , given by Iosifescu (1980, Thm. 4.7).

$$\mathbf{R} = \mathcal{D}(\boldsymbol{\pi})^{-1} (\mathbf{I} - \mathbf{Z} + \mathbf{Z}_{\text{dg}} \mathbf{E}). \quad (11.60)$$

Again, this is the transpose of the expression obtained when \mathbf{P} is row-stochastic.

Theorem 11.4.3 *The sensitivity of \mathbf{R} is*

$$\begin{aligned} d\text{vec } \mathbf{R} = & - \left[\mathbf{R}^{\text{T}} \otimes \mathcal{D}(\boldsymbol{\pi})^{-1} \right] \mathcal{D}(\text{vec } \mathbf{I}_s) (\mathbf{1} \otimes \mathbf{I}_s) d\boldsymbol{\pi} \\ & - \left\{ \left[\mathbf{I}_s \otimes \mathcal{D}(\boldsymbol{\pi})^{-1} \right] - \left[\mathbf{E} \otimes \mathcal{D}(\boldsymbol{\pi})^{-1} \right] \mathcal{D}(\text{vec } \mathbf{I}_s) \right\} d\text{vec } \mathbf{Z} \end{aligned} \quad (11.61)$$

where $d\boldsymbol{\pi}$ is given by (11.55) and $d\text{vec } \mathbf{Z}$ is given by (11.57).

Proof Differentiating (11.60) gives

$$d\mathbf{R} = d \left[\mathcal{D}(\boldsymbol{\pi})^{-1} \right] (\mathbf{I} - \mathbf{Z} + \mathbf{Z}_{\text{dg}} \mathbf{E}) + \mathcal{D}(\boldsymbol{\pi})^{-1} \left[-d\mathbf{Z} + (d\mathbf{Z}_{\text{dg}}) \mathbf{E} \right]. \quad (11.62)$$

Applying the vec operator gives

$$\begin{aligned} d\text{vec } \mathbf{R} = & \left[(\mathbf{I} - \mathbf{Z} + \mathbf{Z}_{\text{dg}} \mathbf{E})^{\text{T}} \otimes \mathbf{I}_s \right] d\text{vec} \left[\mathcal{D}(\boldsymbol{\pi})^{-1} \right] \\ & - \left[\mathbf{I}_s \otimes \mathcal{D}(\boldsymbol{\pi})^{-1} \right] d\text{vec } \mathbf{Z} + \left[\mathbf{E} \otimes \mathcal{D}(\boldsymbol{\pi})^{-1} \right] d\text{vec } \mathbf{Z}_{\text{dg}}. \end{aligned} \quad (11.63)$$

Using (2.82) for $d\text{vec} \left[\mathcal{D}(\boldsymbol{\pi})^{-1} \right]$, (2.69) for $d\text{vec } \mathcal{D}(\boldsymbol{\pi})$, and (11.12) for $d\text{vec } \mathbf{Z}_{\text{dg}}$ yields

$$\begin{aligned} d\text{vec } \mathbf{R} = & - \left[\mathbf{R}^{\text{T}} \mathcal{D}(\boldsymbol{\pi}) \otimes \mathbf{I}_s \right] \left[\mathcal{D}(\boldsymbol{\pi})^{-1} \otimes \mathcal{D}(\boldsymbol{\pi})^{-1} \right] \mathcal{D}(\text{vec } \mathbf{I}) (\mathbf{1} \otimes \mathbf{I}) d\boldsymbol{\pi} \\ & - \left[\mathbf{I} \otimes \mathcal{D}(\boldsymbol{\pi})^{-1} \right] d\text{vec } \mathbf{Z} + \left[\mathbf{E} \otimes \mathcal{D}(\boldsymbol{\pi})^{-1} \right] \mathcal{D}(\text{vec } \mathbf{I}) d\text{vec } \mathbf{Z} \end{aligned} \quad (11.64)$$

which simplifies to give (11.61). \square

11.4.4 Mixing Time and the Kemeny Constant

The mixing time K of a chain is the mean time required to get from a specified state to a state chosen at random from the stationary distribution $\boldsymbol{\pi}$. Remarkably, K is independent of the starting state (Grinstead and Snell 2003; Hunter 2006) and is sometimes called Kemeny's constant; it is a measure of the rate of convergence to stationarity, and is $K = \text{trace}(\mathbf{Z})$ (Hunter 2006). In addition to being a quantity of interest in itself, the rate of convergence also plays a role in the sensitivity of the stationary distribution of ergodic chains (Hunter 2005; Mitrophanov 2005).

Theorem 11.4.4 *The sensitivity of K is*

$$dK = (\text{vec } \mathbf{I}_s)^\top d\text{vec } \mathbf{Z}. \quad (11.65)$$

Proof Differentiating $K = \text{trace}(\mathbf{Z})$ gives

$$dK = \mathbf{1}^\top (\mathbf{I} \circ d\mathbf{Z}) \mathbf{1}. \quad (11.66)$$

Applying the vec operator gives

$$dK = (\mathbf{1}^\top \otimes \mathbf{1}^\top) \mathcal{D}(\text{vec } \mathbf{I}) d\text{vec } \mathbf{Z} \quad (11.67)$$

which simplifies to (11.65). \square

11.4.5 Implicit Parameters and Compensation

Theorems 11.4.1, 11.4.2, 11.4.3, and 11.4.4 are written in terms of $d\text{vec } \mathbf{P}$. However, perturbation of any element, say p_{kj} , to $p_{kj} + \theta_{kj}$, must be compensated for by adjustments of the other elements in column j so that the column sum remains equal to 1 (Conlisk 1985). Two kinds of compensation are likely to be of use in applications: additive and proportional. Additive compensation adjusts all the elements of the column by an equal amount, distributing the perturbation θ_{kj} additively over column j . Proportional compensation distributes θ_{kj} in proportion to the values of the p_{ij} , for $i \neq k$. Proportional compensation is attractive because it preserves the pattern of zero and non-zero elements within \mathbf{P} .

To develop the compensation formulae, let us start by considering a probability vector \mathbf{p} , of dimension $s \times 1$, with $p_i \geq 0$ and $\sum_i p_i = 1$. Let θ_i be the perturbation of p_i , and write

$$\mathbf{p}(\boldsymbol{\theta}) = \mathbf{p}(0) + \mathbf{A}\boldsymbol{\theta} \quad (11.68)$$

for some matrix \mathbf{A} to be determined. If y is a function of \mathbf{p} , then

$$dy = \frac{dy}{d\mathbf{p}^\top} \frac{d\mathbf{p}}{d\boldsymbol{\theta}^\top} d\boldsymbol{\theta} \quad (11.69)$$

evaluated at $\boldsymbol{\theta} = 0$.

Additive compensation For the case of additive compensation, we write

$$\begin{aligned} p_1(\boldsymbol{\theta}) &= p_1(0) + \theta_1 - \frac{\theta_2}{s-1} - \dots - \frac{\theta_s}{s-1} \\ p_2(\boldsymbol{\theta}) &= p_2(0) - \frac{\theta_1}{s-1} + \theta_2 - \dots - \frac{\theta_s}{s-1} \\ &\vdots \\ p_s(\boldsymbol{\theta}) &= p_s(0) - \frac{\theta_1}{s-1} - \frac{\theta_2}{s-1} - \dots + \theta_s \end{aligned} \quad (11.70)$$

The perturbation θ_1 is added to p_1 and compensated for by subtracting $\theta_1/(s-1)$ from all other entries of \mathbf{p} ; clearly $\sum_i p_i(\boldsymbol{\theta}) = 1$ for any perturbation vector $\boldsymbol{\theta}$.

The system of Eqs. (11.70) can be written

$$\mathbf{p}(\boldsymbol{\theta}) = \mathbf{p}(0) + \left(\mathbf{I} - \frac{1}{s-1} \mathbf{C} \right) \boldsymbol{\theta}. \quad (11.71)$$

Defining \mathbf{E} to be a matrix of ones, then the matrix \mathbf{C} can be written (as a so-called Toeplitz matrix) as $\mathbf{C} = \mathbf{E} - \mathbf{I}$, with zeros on the diagonal and ones elsewhere. Thus the matrix \mathbf{A} in (11.68) is

$$\mathbf{A} = \mathbf{I} - \frac{1}{s-1} \mathbf{C} \quad (11.72)$$

Proportional compensation For proportional compensation, assume that $p_i < 1$ for all i . The vector $\mathbf{p}(\boldsymbol{\theta})$ is

$$\begin{aligned} p_1(\boldsymbol{\theta}) &= p_1(0) + \theta_1 - \frac{p_1\theta_2}{1-p_2} - \dots - \frac{p_1\theta_s}{1-p_s} \\ p_2(\boldsymbol{\theta}) &= p_2(0) - \frac{p_2\theta_1}{1-p_1} + \theta_2 - \dots - \frac{p_2\theta_s}{1-p_s} \\ &\vdots \\ p_s(\boldsymbol{\theta}) &= p_s(0) - \frac{p_s\theta_1}{1-p_1} - \frac{p_s\theta_2}{1-p_2} - \dots + \theta_s \end{aligned} \quad (11.73)$$

The perturbation θ_1 is added to p_1 and compensated for by subtracting $\theta_1 p_i / (1 - p_1)$ from the i th entry of \mathbf{p} . Again, $\sum_i p_i(\boldsymbol{\theta}) = 1$ for any perturbation vector $\boldsymbol{\theta}$.

Equation (11.73) can be written

$$\mathbf{p}(\boldsymbol{\theta}) = \mathbf{p}(0) + \left[\mathbf{I} - \mathcal{D}(\mathbf{p}) \mathbf{C} \mathcal{D}(\mathbf{1} - \mathbf{p})^{-1} \right] \boldsymbol{\theta} \tag{11.74}$$

so that the matrix \mathbf{A} in (11.68) is

$$\mathbf{A} = \mathbf{I} - \mathcal{D}(\mathbf{p}) \mathbf{C} \mathcal{D}(\mathbf{1} - \mathbf{p})^{-1} \tag{11.75}$$

The transition matrix We have derived compensation formulae for a single probability vector \mathbf{p} . Now consider perturbation of a probability matrix \mathbf{P} , each column of which is a probability vector. Define a perturbation matrix $\boldsymbol{\Theta}$ where θ_{ij} is the perturbation of p_{ij} . Perturbations of column j are to be compensated by a matrix \mathbf{A}_j , so that

$$\mathbf{P}(\boldsymbol{\Theta}) = \mathbf{P}(0) + \left[\mathbf{A}_1 \boldsymbol{\Theta}(:, 1) \cdots \mathbf{A}_s \boldsymbol{\Theta}(:, s) \right] \tag{11.76}$$

where \mathbf{A}_i compensates for the changes in column i of \mathbf{P} . Applying the vec operator to (11.76) gives

$$\text{vec } \mathbf{P}(\boldsymbol{\Theta}) = \text{vec } \mathbf{P}(0) + \begin{pmatrix} \mathbf{A}_1 & & \\ & \ddots & \\ & & \mathbf{A}_s \end{pmatrix} \text{vec } \boldsymbol{\Theta} \tag{11.77}$$

$$= \text{vec } \mathbf{P}(0) + \sum_{i=1}^s (\mathbf{E}_{ii} \otimes \mathbf{A}_i) \text{vec } \boldsymbol{\Theta}. \tag{11.78}$$

The terms in the summation in (11.78) are recognizable as the vec of the product $\mathbf{A}_i \boldsymbol{\Theta} \mathbf{E}_{ii}$; thus

$$\mathbf{P}(\boldsymbol{\Theta}) = \mathbf{P}(0) + \sum_{i=1}^s \mathbf{A}_i \boldsymbol{\Theta} \mathbf{E}_{ii} \tag{11.79}$$

where \mathbf{E}_{ii} is a matrix with a 1 in the (i, i) entry and zeros elsewhere.

Theorem 11.4.5 *Let \mathbf{P} be a column-stochastic $s \times s$ transition matrix. Let $\boldsymbol{\Theta}$ be a matrix of perturbations, where θ_{ij} is applied to p_{ij} , and the other entries of $\boldsymbol{\Theta}$ compensate for the perturbation. Let $\mathbf{C} = \mathbf{E} - \mathbf{I}$. If compensation is additive, then*

$$\mathbf{P}(\boldsymbol{\Theta}) = \mathbf{P}(0) + \left(\mathbf{I} - \frac{1}{s-1} \mathbf{C} \right) \boldsymbol{\Theta} \tag{11.80}$$

$$\frac{d\text{vec } \mathbf{P}}{d\text{vec } \mathbf{T} \Theta} = \left[\mathbf{I}_{s^2} - \frac{1}{s-1} (\mathbf{I}_s \otimes \mathbf{C}) \right]. \quad (11.81)$$

If compensation is proportional, then

$$\mathbf{P}(\Theta) = \mathbf{P}(0) + \sum_{i=1}^s \left\{ \mathbf{I} - \mathcal{D} [\mathbf{P}(:, i)] \mathbf{C} \mathcal{D} [\mathbf{I} - \mathbf{P}(:, i)]^{-1} \right\} \Theta \mathbf{E}_{ii} \quad (11.82)$$

$$\frac{d\text{vec } \mathbf{P}}{d\text{vec } \mathbf{T} \Theta} = \mathbf{I}_{s^2} - \sum_{i=1}^s \left\{ \mathbf{E}_{ii} \otimes \mathcal{D} [\mathbf{P}(:, i)] \mathbf{C} \mathcal{D} [\mathbf{I} - \mathbf{P}(:, i)]^{-1} \right\}. \quad (11.83)$$

Proof $\mathbf{P}(\Theta)$ is given by (11.79). If compensation is additive, \mathbf{A}_i is given by (11.72) for all i . Substituting into (11.79) gives (11.80). Differentiating (11.80) and applying the vec operator gives (11.81).

If compensation is proportional, substituting (11.75) for \mathbf{A}_i in (11.79) gives (11.82). Differentiating yields

$$d\mathbf{P} = (d\theta) \sum_{i=1}^s \mathbf{E}_{ii} - \sum_{i=1}^s \mathcal{D} [\mathbf{P}(:, i)] \mathbf{C} \mathcal{D} [\mathbf{I} - \mathbf{P}(:, i)]^{-1} (d\Theta) \mathbf{E}_{ii}. \quad (11.84)$$

Using the vec operator gives (11.83). □

Perturbations of \mathbf{P} subject to compensation are given by perturbations of Θ . Thus for any function $y(\mathbf{P})$ we can write

$$\left. \frac{dy}{d\text{vec } \mathbf{T} \mathbf{P}} \right|_{\text{comp}} = \frac{dy}{d\text{vec } \mathbf{T} \mathbf{P}} \frac{d\text{vec } \mathbf{P}}{d\text{vec } \mathbf{T} \Theta} \quad (11.85)$$

where $d\text{vec } \mathbf{P}/d\text{vec } \mathbf{T} \Theta$ is given (for additive and proportional compensation) by Theorem 11.4.5. The slight notational complexity is worthwhile for clarifying how to use Theorem 11.4.5 in practice.

11.5 Species Succession in a Marine Community

Markov chains are used by ecologists as models of species replacement (succession) in ecological communities; (e.g., Horn 1975; Hill et al. 2004; Nelis and Wootton 2010). In these models, the state of a point on a landscape is given by the species occupying that point. The entry p_{ij} of \mathbf{P} is the probability that species j is replaced by species i between t and $t + 1$. If a community consists of a large number of points independently subject to the transition probabilities in \mathbf{P} , the stationary distribution $\boldsymbol{\pi}$ will give the relative frequencies of species in the community at equilibrium.

Hill et al. (2004) used a Markov chain to describe a community of encrusting organisms occupying rock surfaces at 30–35 m depth in the Gulf of Maine. The Markov chain contained 14 species plus an additional state (“bare rock”) for unoccupied substrate. The matrix \mathbf{P} was estimated from longitudinal data (Hill et al. 2002, 2004) and is given, along with a list of species names, in Appendix B. We will use the results of this chapter to analyze the sensitivity of species diversity and the Kemeny constant to the processes of colonization and replacement that determining \mathbf{P} .

11.5.1 Biotic Diversity

The stationary distribution π , with the species numbered in order of decreasing abundance and bare rock placed at the end as state 15, is shown in Fig. 11.2. The two dominant species are an encrusting sponge (called *Hymedesmia*) and a bryozoan (*Crisia*).

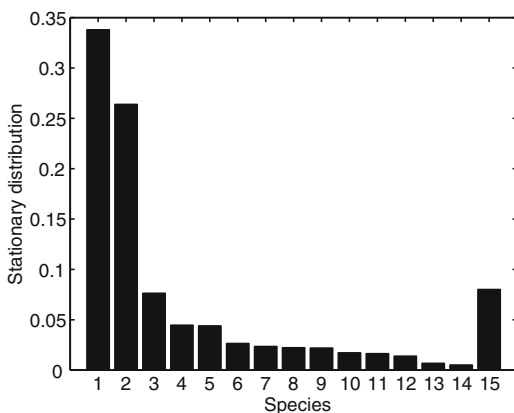
The entropy of this stationary distribution, $H(\pi) = -\pi^T(\log \pi)$, where the logarithm is applied elementwise, is used as an index of biodiversity; it is maximal when all species are equally abundant and goes to 0 in a community dominated by a single species. The sensitivity of H is

$$dH = -\left(\log \pi^T + \mathbf{1}^T\right) d\pi \tag{11.86}$$

Most ecologists, however, would not include bare substrate in a measure of biodiversity, so we define instead a “biotic diversity” $H_b(\pi) = H(\pi_b)$ where

$$\pi_b = \frac{\mathbf{G}\pi}{\|\mathbf{G}\pi\|}. \tag{11.87}$$

Fig. 11.2 The stationary distribution for the subtidal benthic community succession model of Hill et al. (2004). States 1–14 correspond to species, numbered in decreasing order of abundance in the stationary distribution. State 15 is bare rock, unoccupied by any species. For the identity of species and the transition matrix, see Appendix B



The matrix \mathbf{G} , of dimension 14×15 , is a 0–1 matrix that selects rows 1–14 of $\boldsymbol{\pi}$. Because $\boldsymbol{\pi}$ is positive, $\|\mathbf{G}\boldsymbol{\pi}\| = \mathbf{1}^\top \mathbf{G}\boldsymbol{\pi}$. Differentiating $\boldsymbol{\pi}_b$ gives

$$d\boldsymbol{\pi}_b = \left(\frac{\mathbf{G}}{\mathbf{1}^\top \mathbf{G}\boldsymbol{\pi}} - \frac{\mathbf{G}\boldsymbol{\pi}\mathbf{1}^\top \mathbf{G}}{(\mathbf{1}^\top \mathbf{G}\boldsymbol{\pi})^2} \right) d\boldsymbol{\pi} \quad (11.88)$$

which simplifies to

$$d\boldsymbol{\pi}_b = \left(\frac{\mathbf{G} - \boldsymbol{\pi}_b \mathbf{1}^\top \mathbf{G}}{\mathbf{1}^\top \mathbf{G}\boldsymbol{\pi}} \right) d\boldsymbol{\pi} \quad (11.89)$$

This model contains no explicit parameters; perturbations of the transition probabilities themselves are of interest and a compensation pattern is needed. Because the relative magnitudes of the entries in a column of \mathbf{P} reflect the relative abilities of species to capture or to hold space, proportional compensation is appropriate in this case because it preserves these relative abilities.

The sensitivity and elasticity of the biotic diversity H_b to changes in the matrix \mathbf{P} , subject to proportional compensation, are

$$\left. \frac{dH_b}{d\text{vec } \mathbf{P}} \right|_{\text{comp}} = \underbrace{\frac{dH_b}{d\boldsymbol{\pi}_b^\top}}_1 \underbrace{\frac{d\boldsymbol{\pi}_b}{d\boldsymbol{\pi}^\top}}_2 \underbrace{\frac{d\boldsymbol{\pi}}{d\text{vec } \mathbf{P}}}_3 \underbrace{\frac{d\text{vec } \mathbf{P}}{d\text{vec } \mathbf{P}}}_4 \quad (11.90)$$

$$\left. \frac{\epsilon H_b}{\epsilon \text{vec } \mathbf{P}} \right|_{\text{comp}} = \frac{1}{H_b} \frac{dH_b}{d\text{vec } \mathbf{P}} \mathcal{D}(\text{vec } \mathbf{P}) \quad (11.91)$$

Term 1 on the right hand side of (11.90) is the derivative of H_b with respect to $\boldsymbol{\pi}_b$, and is given by (11.86). Term 2 is the derivative of the biotic diversity vector $\boldsymbol{\pi}_b$ with respect to the full diversity vector $\boldsymbol{\pi}$, given by (11.89). Term 3 is the derivative of the diversity vector $\boldsymbol{\pi}$ with respect to the transition matrix \mathbf{P} , given by, (11.55). Finally, Term 4 is the derivative of the matrix \mathbf{P} taking into account the compensation structure in (11.83).

The sensitivity and elasticity vectors (11.90) and (11.91) are of dimension $1 \times s^2 = 1 \times 255$. To reduce the number of independent perturbations, we consider subsets of the p_{ij} : disturbance (in which a species is replaced by bare rock), colonization of unoccupied space, replacement of one species by another, and persistence of a species in its location, where

$$P[\text{disturbance of sp. } i] = p_{si}$$

$$P[\text{colonization by sp. } i] = p_{is}$$

$$P[\text{persistence of sp. } i] = p_{ii}$$

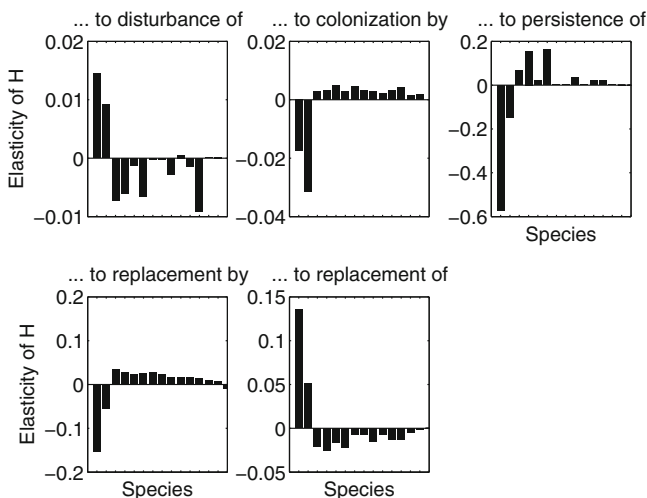


Fig. 11.3 The elasticity of the biotic diversity $H_b(\pi)$ calculated over the biotic states of the stationary distribution of the subtidal benthic community succession model of Hill et al. (2004). States 1–14 correspond to species, numbered in decreasing order of abundance in the stationary distribution. State 15 is bare rock, unoccupied by any species. For the identity of species and the transition matrix, see Appendix B

$$P[\text{replacement of sp. } i] = \sum_{k \neq i, s} p_{ki}$$

$$P[\text{replacement by sp. } i] = \sum_{j \neq i, s} p_{ij}$$

Extracting the corresponding elements of $\frac{\epsilon H_b}{\epsilon \text{vec}^T \mathbf{P}}$ gives the elasticities to these classes of probabilities. Figure 11.3 shows that the dominant species (1 and 2) have impacts that are larger than, and opposite in sign to, those of the remaining species. Biodiversity would be enhanced by increasing the disturbance of, or the replacement of, species 1 and 2, and reduced by increasing the rates of colonization by, persistence of, or replacement by species 1 and 2.

11.5.2 The Kemeny Constant and Ecological Mixing

Ecologists have used several measures of the rate of convergence of communities modelled by Markov chains, including the damping ratio and Dobrushin’s coefficient of ergodicity (Hill et al. 2004). The Kemeny constant K is an interesting addition to this list; it gives the expected time to get from any initial state to

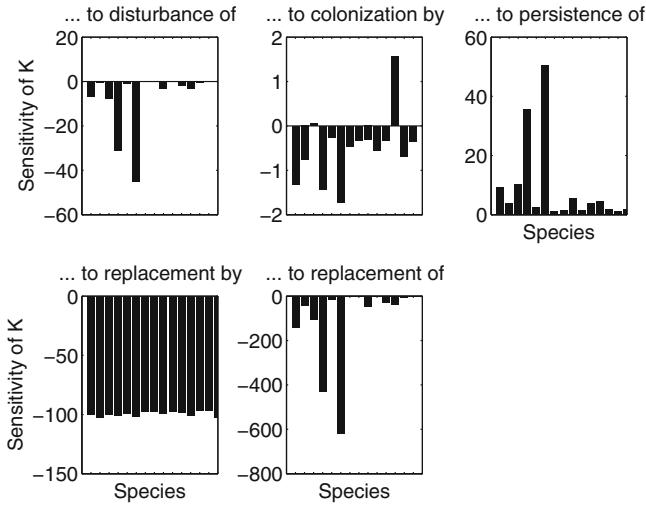


Fig. 11.4 The sensitivity of the Kemeny constant K of the subtidal benthic community succession model of Hill et al. (2004). States 1–14 correspond to species, numbered in decreasing order of abundance in the stationary distribution. State 15 is bare rock, unoccupied by any species. For the identity of species and the transition matrix, see Appendix B

a state selected at random from the stationary distribution (Hunter 2006). Once reaching that state, the behavior of the chain and the stationary process are indistinguishable.

The sensitivity of K , subject to compensation, is

$$\left. \frac{dK}{d\text{vec}^T \mathbf{P}} \right|_{\text{comp}} = \frac{dK}{d\text{vec}^T \mathbf{Z}} \frac{d\text{vec} \mathbf{Z}}{d\text{vec}^T \mathbf{P}} \frac{d\text{vec} \mathbf{P}}{d\text{vec}^T \Theta} \tag{11.92}$$

where the three terms on the right hand side are given by (11.65), (11.57), and (11.83), respectively.

Figure 11.4 shows the sensitivities $dK/d\text{vec}^T \mathbf{P}$, subject to proportional compensation, and aggregated as in Fig. 11.3. Unlike the case with H_b , the two dominant species do not stand out from the others. Increases in the rates of replacement will speed up convergence, and increases in persistence will slow convergence. The disturbance of, colonization by, persistence of, and replacement of species 6 (it is a sea anemone, *Urticina crassicornis*) have particularly large impacts on K . Examination of row 6 and column 6 of \mathbf{P} (Appendix B) shows that *U. crassicornis* has the highest probability of persistence ($p_{66} = 0.86$), and one of the lowest rates of disturbance, in the community. While it is far from dominant (Fig. 11.2), it has a major impact on the rate of mixing.

11.6 Discussion

Given that many properties of finite state Markov chains can be expressed as simple matrix expressions, matrix calculus is an attractive approach to finding the sensitivity and elasticity to parameter perturbations. Most of the literature on perturbation analysis of Markov chains has focused on the stationary distribution of ergodic chains, but the approach here is equally applicable to absorbing chains, and to dependent variables other than the stationary distribution. The perturbation of ergodic chains is often studied using generalized inverses, since the influential studies of Meyer (Meyer 1975, 1994; Golub and Meyer 1986; Funderlic and Meyer 1986). Matrix calculus provides a complementary approach; the sensitivity of the stationary distribution π obtained here agrees with the result obtained by Golub and Meyer (1986) using the group generalized inverse.

The examples shown here are typical of cases where absorbing or ergodic Markov chains are used in population biology and ecology. In each example, the dependent variables of interest are functions several steps removed from the chain itself. The ease with which one can differentiate such functions is a particularly attractive property of the matrix calculus approach.

A Appendix A: Proofs

Theorems 11.2.1 and 11.2.2 give the sensitivities of the moments of the number of visits to transient states and of the time to absorption, respectively. These results are obtained by applying matrix calculus to the expressions for the moments. Proofs are given in the text for the first two moments; the proofs for the others follow the same steps but introduce no new concepts, and so are presented here.

A.1 Derivatives of the Moments of Occupancy Times

To continue the proof of Theorem 11.2.1, take partial differentials of \mathbf{N}_3 in (11.5) with respect to \mathbf{N}_1 and \mathbf{N}_{dg} , to obtain

$$\partial_{\mathbf{N}_1} \mathbf{N}_3 = \left(6\mathbf{N}_{dg}^2 - 6\mathbf{N}_{dg} + \mathbf{I} \right) d\mathbf{N}_1 \quad (11.93)$$

$$\partial_{\mathbf{N}_{dg}} \mathbf{N}_3 = 6 \left(d\mathbf{N}_{dg} \right) \mathbf{N}_{dg} \mathbf{N}_1 + 6\mathbf{N}_{dg} \left(d\mathbf{N}_{dg} \right) \mathbf{N}_1 - 6 \left(d\mathbf{N}_{dg} \right) \mathbf{N}_1 \quad (11.94)$$

Applying the vec operator to each term and using Roth's theorem gives

$$\partial_{\mathbf{N}_1} \text{vec } \mathbf{N}_3 = \left[\mathbf{I} \otimes \left(6\mathbf{N}_{dg}^2 - 6\mathbf{N}_{dg} + \mathbf{I} \right) \right] d\text{vec } \mathbf{N}_1 \quad (11.95)$$

$$\begin{aligned} \partial_{\mathbf{N}_{\text{dg}}} \text{vec } \mathbf{N}_3 = & \left[6 \left(\mathbf{N}_1^{\text{T}} \mathbf{N}_{\text{dg}} \otimes \mathbf{I} \right) + 6 \left(\mathbf{N}^{\text{T}} \otimes \mathbf{N}_{\text{dg}} \right) - 6 \left(\mathbf{N}_1^{\text{T}} \otimes \mathbf{I} \right) \right] \\ & d \text{vec } \mathbf{N}_{\text{dg}}. \end{aligned} \quad (11.96)$$

Substituting (11.95) and (11.96) into (11.13) gives (11.9).

Taking partial differentials of \mathbf{N}_4 in (11.6) gives

$$\partial_{\mathbf{N}_1} \mathbf{N}_4 = \left(24 \mathbf{N}_{\text{dg}}^3 - 36 \mathbf{N}_{\text{dg}}^2 + 14 \mathbf{N}_{\text{dg}} - \mathbf{I} \right) d \mathbf{N}_1 \quad (11.97)$$

$$\begin{aligned} \partial_{\mathbf{N}_{\text{dg}}} \mathbf{N}_4 = & 24 \left(d \mathbf{N}_{\text{dg}} \right) \mathbf{N}_{\text{dg}}^2 \mathbf{N}_1 + 24 \mathbf{N}_{\text{dg}} \left(d \mathbf{N}_{\text{dg}} \right) \mathbf{N}_{\text{dg}} \mathbf{N}_1 + 24 \mathbf{N}_{\text{dg}}^2 \left(d \mathbf{N}_{\text{dg}} \right) \mathbf{N}_1 \\ & - 36 \left(d \mathbf{N}_{\text{dg}} \right) \mathbf{N}_{\text{dg}} \mathbf{N}_1 - 36 \mathbf{N}_{\text{dg}} \left(d \mathbf{N}_{\text{dg}} \right) \mathbf{N}_1 + 14 \left(d \mathbf{N}_{\text{dg}} \right) \mathbf{N}_1. \end{aligned} \quad (11.98)$$

Applying the vec operator yields

$$\partial_{\mathbf{N}_1} \text{vec } \mathbf{N}_4 = \left[\mathbf{I} \otimes \left(24 \mathbf{N}_{\text{dg}}^3 - 36 \mathbf{N}_{\text{dg}}^2 + 14 \mathbf{N}_{\text{dg}} - \mathbf{I} \right) \right] d \text{vec } \mathbf{N}_1 \quad (11.99)$$

$$\begin{aligned} \partial_{\mathbf{N}_{\text{dg}}} \text{vec } \mathbf{N}_4 = & \left[24 \left(\mathbf{N}_1^{\text{T}} \mathbf{N}_{\text{dg}}^2 \otimes \mathbf{I} \right) + 24 \left(\mathbf{N}_1^{\text{T}} \mathbf{N}_{\text{dg}} \otimes \mathbf{N}_{\text{dg}} \right) + 24 \left(\mathbf{N}_1^{\text{T}} \otimes \mathbf{N}_{\text{dg}}^2 \right) \right. \\ & - 36 \left(\mathbf{N}_1^{\text{T}} \mathbf{N}_{\text{dg}} \otimes \mathbf{I} \right) - 36 \left(\mathbf{N}_1^{\text{T}} \otimes \mathbf{N}_{\text{dg}} \right) \\ & \left. + 14 \left(\mathbf{N}_1^{\text{T}} \otimes \mathbf{I} \right) \right] d \text{vec } \mathbf{N}_{\text{dg}}. \end{aligned} \quad (11.100)$$

Substituting (11.99) and (11.100) into (11.13) gives (11.10).

A.2 Derivatives of the Moments of Time to Absorption

To continue the proof of Theorem 11.2.2, take partial differentials of η_3 , in (11.21) with respect to η_1 and \mathbf{N}_1 , to obtain

$$\partial_{\eta_1} \eta_3^{\text{T}} = \left(d \eta_1^{\text{T}} \right) \left(6 \mathbf{N}_1^2 - 6 \mathbf{N}_1 + \mathbf{I} \right) \quad (11.101)$$

$$\partial_{\mathbf{N}_1} \eta_3^{\text{T}} = 6 \eta_1^{\text{T}} \left(d \mathbf{N}_1 \right) + 6 \eta_1^{\text{T}} \mathbf{N}_1 \left(d \mathbf{N}_1 \right) - 6 \eta_1^{\text{T}} \left(d \mathbf{N}_1 \right). \quad (11.102)$$

Applying the vec operator yields

$$\partial_{\eta_1} \eta_3 = \left(6 \mathbf{N}_1^2 - 6 \mathbf{N}_1 + \mathbf{I} \right)^{\text{T}} d \eta_1 \quad (11.103)$$

$$\partial_{\mathbf{N}_1} \eta_3 = \left[6 \left(\mathbf{N}_1^{\text{T}} \otimes \eta_1^{\text{T}} \right) + 6 \left(\mathbf{I} \otimes \eta_1^{\text{T}} \mathbf{N}_1 \right) - 6 \left(\mathbf{I} \otimes \eta_1^{\text{T}} \right) \right] d \text{vec } \mathbf{N}_1 \quad (11.104)$$

which combine to yield (11.25).

The partial differentials of η_4 in (11.22) with respect to η_1 and \mathbf{N}_1 are

$$\partial_{\eta_1} \eta_4^\top = d\eta_1^\top \left(24\mathbf{N}_1^3 - 36\mathbf{N}_1^2 + 14\mathbf{N}_1 - \mathbf{I} \right) \quad (11.105)$$

$$\begin{aligned} \partial_{\mathbf{N}_1} \eta_4^\top = \eta_1^\top \left[24 (d\mathbf{N}_1) \mathbf{N}_1^2 + 24\mathbf{N}_1 (d\mathbf{N}_1) \mathbf{N}_1 + 24\mathbf{N}_1^2 (d\mathbf{N}_1) \right. \\ \left. - 36 (d\mathbf{N}_1) \mathbf{N}_1 - 36\mathbf{N}_1 (d\mathbf{N}_1) + 14d\mathbf{N}_1 \right]. \end{aligned} \quad (11.106)$$

Applying the vec operator to each equation gives

$$\partial_{\eta_1} \eta_4 = \left(24\mathbf{N}_1^3 - 36\mathbf{N}_1^2 + 14\mathbf{N}_1 - \mathbf{I} \right)^\top d\eta_1 \quad (11.107)$$

$$\begin{aligned} \partial_{\mathbf{N}_1} \eta_4 = \left\{ 24 \left[\left(\mathbf{N}_1^2 \right)^\top \otimes \eta_1^\top \right] + 24 \left(\mathbf{N}_1^\top \otimes \eta_1^\top \mathbf{N}_1 \right) + 24 \left(\mathbf{I} \otimes \eta_1^\top \mathbf{N}_1^\top \right) \right. \\ \left. - 36 \left(\mathbf{N}_1^\top \otimes \eta_1^\top \right) - 36 \left(\mathbf{I} \otimes \eta_1^\top \mathbf{N}_1 \right) + 14 \left(\mathbf{I} \otimes \eta_1^\top \right) \right\} d\text{vec } \mathbf{N}_1 \end{aligned} \quad (11.108)$$

which combine to give (11.26).

B Appendix B: Marine Community Matrix

	Model states	Species type	State ID	Number
1	<i>Hymedesmia</i> 1 sp.	Sponge	HY1	14875
2	<i>Crisia eburnea</i>	Bryozoan	CRI	9915
3	<i>Myxilla fimbriata</i>	Sponge	MYX	4525
4	<i>Mycale lingua</i>	Sponge	MYC	3001
5	<i>Filograna implexa</i>	Polychaete	FIL	2219
6	<i>Urticina crassicornis</i>	Sea anemone	URT	992
7	<i>Ascidia callosa</i>	Ascidian	ASC	1052
8	<i>Aplidium pallidum</i>	Ascidian	APL	1166
9	<i>Hymedesmia</i> 2 sp.	Sponge	HY2	1226
10	<i>Idmidronea atlantica</i>	Bryozoan	IDM	730
11	<i>Coralline Algae</i>	Encrusting algae	COR	875
12	<i>Metridium senile</i>	Sea anemone	MET	1298
13	<i>Parasmittina jeffreysi</i>	Bryozoan	PAR	402
14	<i>Spirorbis spirorbis</i>	Polychaete	SPI	225
15	Bare Rock		BR	4266

The transition matrix for the marine benthic community (Hill et al. 2004) is

$\mathbf{P} =$

$$\begin{pmatrix} 0.771 & 0.145 & 0.052 & 0.017 & 0.117 & 0.009 & 0.241 & 0.199 & 0.056 & 0.309 & 0.056 & 0.025 & 0.321 & 0.158 & 0.101 \\ 0.102 & 0.609 & 0.061 & 0.054 & 0.218 & 0.024 & 0.223 & 0.235 & 0.147 & 0.228 & 0.222 & 0.068 & 0.179 & 0.448 & 0.320 \\ 0.017 & 0.031 & 0.710 & 0.006 & 0.035 & 0.012 & 0.051 & 0.038 & 0.026 & 0.031 & 0.028 & 0.018 & 0.023 & 0.018 & 0.025 \\ 0.004 & 0.011 & 0.004 & 0.839 & 0.004 & 0.000 & 0.016 & 0.018 & 0.011 & 0.010 & 0.008 & 0.030 & 0.000 & 0.018 & 0.009 \\ 0.015 & 0.028 & 0.020 & 0.005 & 0.404 & 0.016 & 0.080 & 0.089 & 0.020 & 0.027 & 0.036 & 0.016 & 0.063 & 0.085 & 0.062 \\ 0.001 & 0.005 & 0.004 & 0.000 & 0.008 & 0.863 & 0.024 & 0.007 & 0.006 & 0.006 & 0.000 & 0.000 & 0.000 & 0.006 & 0.005 \\ 0.018 & 0.022 & 0.008 & 0.004 & 0.033 & 0.001 & 0.105 & 0.044 & 0.011 & 0.042 & 0.025 & 0.010 & 0.030 & 0.030 & 0.048 \\ 0.012 & 0.025 & 0.008 & 0.006 & 0.032 & 0.007 & 0.041 & 0.154 & 0.026 & 0.031 & 0.020 & 0.016 & 0.020 & 0.018 & 0.034 \\ 0.002 & 0.011 & 0.025 & 0.008 & 0.013 & 0.016 & 0.014 & 0.015 & 0.586 & 0.010 & 0.007 & 0.004 & 0.003 & 0.018 & 0.013 \\ 0.014 & 0.015 & 0.003 & 0.004 & 0.007 & 0.003 & 0.033 & 0.027 & 0.021 & 0.165 & 0.007 & 0.003 & 0.020 & 0.030 & 0.031 \\ 0.003 & 0.012 & 0.005 & 0.006 & 0.006 & 0.004 & 0.025 & 0.016 & 0.006 & 0.013 & 0.507 & 0.001 & 0.017 & 0.006 & 0.017 \\ 0.002 & 0.008 & 0.007 & 0.011 & 0.005 & 0.007 & 0.005 & 0.020 & 0.005 & 0.008 & 0.002 & 0.537 & 0.000 & 0.006 & 0.017 \\ 0.005 & 0.005 & 0.002 & 0.000 & 0.006 & 0.000 & 0.014 & 0.009 & 0.001 & 0.012 & 0.005 & 0.003 & 0.248 & 0.000 & 0.011 \\ 0.003 & 0.004 & 0.008 & 0.003 & 0.005 & 0.000 & 0.012 & 0.009 & 0.005 & 0.006 & 0.003 & 0.003 & 0.000 & 0.030 & 0.013 \\ 0.029 & 0.069 & 0.084 & 0.036 & 0.108 & 0.036 & 0.115 & 0.122 & 0.074 & 0.104 & 0.076 & 0.266 & 0.076 & 0.127 & 0.294 \end{pmatrix}$$

(11.109)

References

- Caswell, H. 2001. *Matrix Population Models: Construction, Analysis, and Interpretation*. 2nd edition. Sinauer Associates, Sunderland, MA.
- Caswell, H., 2006. Applications of Markov chains in demography. Pages 319–334 in *MAM2006: Markov Anniversary Meeting*. Boson Books, Raleigh, North Carolina.
- Caswell, H. 2008. Perturbation analysis of nonlinear matrix population models. *Demographic Research* **18**:59–116.
- Caswell, H. 2009. Stage, age and individual stochasticity in demography. *Oikos* **118**:1763–1782.
- Caswell, H. 2011. Perturbation analysis of continuous-time absorbing Markov chains. *Numerical Linear Algebra with Applications* **18**:901–917.
- Cho, G. E., and C. D. Meyer. 2000. Comparison of perturbation bounds for the stationary distribution of a Markov chain. *Linear Algebra and its Applications* **335**:137–150.
- Conlisk, J. 1985. Comparative statics for Markov chains. *Journal of Economic Dynamics and Control* **9**:139–151.
- Darroch, J. N., and E. Seneta. 1965. On quasi-stationary distributions in absorbing discrete-time finite Markov chains. *Journal of Applied Probability* **2**:88–100.
- Edwards, R. D., and S. Tuljapurkar. 2005. Inequality in life spans and a new perspective on mortality convergence across industrialized countries. *Population and Development Review* **31**:645–674.
- Funderlic, R. E., and C. D. Meyer, Jr. 1986. Sensitivity of the stationary distribution vector for an ergodic Markov chain. *Linear Algebra and its Applications* **76**:1–17.
- Golub, G. H., and C. D. Meyer, Jr. 1986. Using the QR factorization and group inversion to compute, differentiate, and estimate the sensitivity of stationary probabilities for Markov chains. *SIAM Journal on Algebraic and Discrete Methods* **7**:273–281.

- Grinstead, C. M., and J. L. Snell. 2003. Introduction to probability. Second edition. American Mathematical Society.
- Hill, M. F., J. D. Witman, and H. Caswell. 2002. Spatio-temporal variation in Markov chain models of subtidal community succession. *Ecology Letters* **5**:665–675.
- Hill, M. F., J. D. Witman, and H. Caswell. 2004. Markov chain analysis of succession in a rocky subtidal community. *The American Naturalist* **164**:E46–E61.
- Horn, H. S., 1975. Markovian properties of forest succession. Pages 196–211 in M. L. Cody and J. M. Diamond, editors. *Ecology and evolution of communities*. Harvard University Press, Cambridge, MA.
- Horvitz, C. C., and S. Tuljapurkar. 2008. Stage dynamics, period survival, and mortality plateaus. *American Naturalist* **172**:203–215.
- Human Mortality Database. 2016. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). www.mortality.org URL www.mortality.org.
- Hunter, J. J. 2005. Stationary distributions and mean first passage times of perturbed Markov chains. *Linear Algebra and its Applications* **410**:217–243.
- Hunter, J. J. 2006. Mixing times with applications to perturbed Markov chains. *Linear Algebra and its Applications* **417**:108–123.
- Iosifescu, M. 1980. *Finite Markov Processes and Their Applications*. Wiley, New York, New York.
- Kemeny, J. G. 1981. Generalization of a fundamental matrix. *Linear Algebra and its Applications* **38**:193–206.
- Kemeny, J. G., and J. L. Snell. 1960. *Finite Markov Chains*. Van Nostrand, Princeton, New Jersey.
- Kirkland, S. 2003. Conditioning properties of the stationary distribution for a Markov chain. *Electronic Journal of Linear Algebra* **10**:1–15.
- Kirkland, S. J., M. M. Neumann, and N.-S. Sze. 2008. On optimal condition numbers for Markov chains. *Numerische Mathematik* **110**:521–537.
- Meyer, C. D. 1975. The role of the group generalized inverse in the theory of finite Markov chains. *SIAM Review* **17**:443–464.
- Meyer, C. D. 1994. Sensitivity of the stationary distribution of a Markov chain. *SIAM Journal of Matrix Analysis and Applications* **15**:715–728.
- Meyer, C. D., and G. W. Stewart. 1982. Derivatives and perturbations of eigenvectors. *SIAM Journal of Numerical Analysis* **25**:679–691.
- Mitrophanov, A. Y. 2003. Stability and exponential convergence of continuous-time Markov chains. *Journal of Applied Probability* **40**:970–979.
- Mitrophanov, A. Y. 2005. Sensitivity and convergence of uniformly ergodic Markov chains. *Journal of Applied Probability* **42**:1003–1014.
- Mitrophanov, A. Y., A. Lomsadze, and M. Borodovsky. 2005. Sensitivity of hidden Markov models. *Journal of Applied Probability* **42**:632–642.
- Nelis, L. C., and J. T. Wootton. 2010. Treatment-based Markov chain models clarify mechanisms of invasion in an invaded grassland community. *Proceedings B, The Royal Society of London* **277**:539.
- Schweitzer, P. J. 1968. Perturbation theory and finite Markov chains. *Journal of Applied Probability* **5**:401–413.
- Seneta, E. 1988. Perturbation of the stationary distribution measured by ergodicity coefficients. *Advances in Applied Probability* **20**:228–230.
- Seneta, E. 1993. Sensitivity of finite Markov chains under perturbation. *Statistics and Probability Letters* **17**:163–168.
- Vaupel, J. W., and V. Canudas Romo. 2003. Decomposing change in life expectancy: a bouquet of formulas in honor of Nathan Keyfitz's 90th birthday. *Demography* **40**:201–216.

- Wilmoth, J. R., and S. Horiuchi. 1999. Rectangularization revisited: variability of age at death within human populations. *Demography* **36**:475–495.
- Zhang, Z., and J. W. Vaupel. 2009. The age separating early deaths from late deaths. *Demographic Research* **20**:721–730.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

