# Chapter 12
# Sensitivity Analysis of Continuous Markov Chains

## 12.1 Introduction

When Markov chains are used as mathematical models of natural or social phenomena, the transition intensities or probabilities are usually defined in terms of parameters that are relevant to the scientific question at hand. Sensitivity analysis of such models is important because it quantifies the dependence of the model behavior on the parameters. This chapter presents sensitivity results for finite-state, continuous-time absorbing Markov chains, paralleling the approach for discrete-time chains in Chap. 11. In absorbing chains, interest focuses on behavior prior to absorption (time spent in transient states and time to absorption) and on the probabilities of absorption in each absorbing state. Here we will derive formulae for the sensitivity and the elasticity (i.e., proportional sensitivity) of the moments of the time to absorption, the time spent in each transient state, and the number of visits to each transient state.

The most basic difference between discrete-time and continuous-time Markov chains is that the former are defined by transition probabilities, while the latter are defined by transition rates. This leads to differences in the structure of the matrices, but there is a nice parallelism in the results.

Perturbation analysis of Markov chains has a long history (Schweitzer 1968; Meyer 1975). Most of the literature, however, is devoted to discrete-time chains, and most of that focuses on ergodic chains and the perturbation analysis of the stationary distribution; e.g. Funderlic and Meyer (1986), Golub and Meyer (1986), Hunter (2005), Cho and Meyer (2000), and Seneta (1993). Much less attention has been paid to continuous-time chains. Perturbation expansions have been developed

---

for the stationary distribution of ergodic continuous-time chains, with application to queueing models (Altman et al. 2004), and sensitivity results and perturbation bounds presented for transient solutions (Ramesh and Trivedi 1993; Mitrophanov 2004). The operations research literature contains many studies of the sensitivity of performance measures calculated over realizations of a continuous-time ergodic Markov chain; e.g., Cao (1989), Glasserman (1992), and Cao et al. (1996). The results to be presented here complement and extend the existing literature on perturbation analysis of Markov chains, by focusing on the statistical properties of the solutions of absorbing continuous-time chains, by introducing the use of matrix calculus, and (as a consequence of that technique) extending the range of parameters whose effects can be evaluated.

### 12.1.1  Absorbing Markov Chains

I consider a finite state, homogeneous, continuous-time Markov chain with intensity matrix $\mathbf{Q}$, where $q_{ij}$ is the rate of transition from stage $j$ to stage $i$. The intensity matrix satisfies $q_{ij} \geq 0$ for $i \neq j$ and $q_{jj} = -\sum_{i \neq j} q_{ij}$. Note that $\mathbf{Q}$ is written in column-to-row orientation, and operates on column vectors. An absorbing chain contains at least one absorbing class of states. Numbering the states so that the transient states appear before the absorbing states leads to the intensity matrix

$$\mathbf{Q} = \left( \begin{array}{c|c} \mathbf{U} & \mathbf{0} \\ \hline \mathbf{M} & \mathbf{0} \end{array} \right). \tag{12.1}$$

The matrix $\mathbf{U}$ contains rates of transitions among the transient states, and $\mathbf{M}$ contains the rates of transition from transient to absorbing states.

I assume that $\mathbf{U}$ and $\mathbf{M}$ are differentiable functions of a vector $\boldsymbol{\theta}$ of parameters, and that $\mathbf{Q}[\boldsymbol{\theta}]$ remains an intensity matrix for sufficiently small perturbations of $\boldsymbol{\theta}$. This includes as a special case the situation where the elements of $\boldsymbol{\theta}$ are simply some or all of the $q_{ij}$, $i \neq j$. The goal of the perturbation analysis is to obtain the derivatives of properties of the chain with respect to $\boldsymbol{\theta}$.

## 12.2   Occupancy Time in Transient States

Let $s$ be the number of transient states, and $\nu_{ij}$ be the time spent in transient state $i$ by an individual starting in transient state $j$. Define $\mathbf{N}_k = E\left(\nu_{ij}^k\right)$ as the matrix whose entries are the $k$th moments, and $\mathbf{N}_{\mathrm{dg}} = (\mathbf{N}_1)_{\mathrm{dg}}$. The matrix $\mathbf{N}_1$ of expectations is the fundamental matrix of the chain. The first several moments of occupancy times are given by the entries of the matrices

$$\mathbf{N}_1 = -\mathbf{U}^{-1} \tag{12.2}$$

$$\mathbf{N}_2 = 2\mathbf{N}_{dg}\mathbf{N}_1 \tag{12.3}$$

$$\mathbf{N}_3 = 6\mathbf{N}_{dg}^2\mathbf{N}_1 \tag{12.4}$$

$$\mathbf{N}_4 = 24\mathbf{N}_{dg}^3\mathbf{N}_1 \tag{12.5}$$

and, in general, by

$$\mathbf{N}_k = k\mathbf{N}_{dg}\mathbf{N}_{k-1} \qquad k \geq 2 \tag{12.6}$$

(Iosifescu 1980, Thm. 8.7).

The differentials of the moments (12.2), (12.3), (12.4), and (12.5) are

$$d\text{vec}\,\mathbf{N}_1 = \left(\mathbf{N}_1^\mathsf{T} \otimes \mathbf{N}_1\right) d\text{vec}\,\mathbf{U} \tag{12.7}$$

$$d\text{vec}\,\mathbf{N}_2 = 2\left\{\left(\mathbf{N}_1^\mathsf{T} \otimes \mathbf{I}\right)\mathcal{D}\,(\text{vec}\,\mathbf{I}) + \left(\mathbf{I} \otimes \mathbf{N}_{dg}\right)\right\}\left(\mathbf{N}_1^\mathsf{T} \otimes \mathbf{N}_1\right) d\text{vec}\,\mathbf{U} \tag{12.8}$$

$$d\text{vec}\,\mathbf{N}_3 = 6\left\{2\left(\mathbf{N}_1^\mathsf{T} \otimes \mathbf{N}_{dg}\right)\mathcal{D}\,(\text{vec}\,\mathbf{I}) + \left(\mathbf{I} \otimes \mathbf{N}_{dg}^2\right)\right\}\left(\mathbf{N}_1^\mathsf{T} \otimes \mathbf{N}_1\right) d\text{vec}\,\mathbf{U} \tag{12.9}$$

$$d\text{vec}\,\mathbf{N}_4 = 24\left\{3\left(\mathbf{N}_1^\mathsf{T} \otimes \mathbf{N}_{dg}^2\right)\mathcal{D}\,(\text{vec}\,\mathbf{I}) + \left(\mathbf{I} \otimes \mathbf{N}_{dg}^3\right)\right\}\left(\mathbf{N}_1^\mathsf{T} \otimes \mathbf{N}_1\right) d\text{vec}\,\mathbf{U} \tag{12.10}$$

where $\mathbf{I} = \mathbf{I}_s$ throughout. A recursive relation for all the moments is

$$d\text{vec}\,\mathbf{N}_k = k\left(\mathbf{N}_{k-1}^\mathsf{T} \otimes \mathbf{I}\right)\mathcal{D}\,(\text{vec}\,\mathbf{I})d\text{vec}\,\mathbf{N} + k\left(\mathbf{I} \otimes \mathbf{N}_{dg}\right) d\text{vec}\,\mathbf{N}_{k-1} \qquad k \geq 2. \tag{12.11}$$

The variance, standard deviation, and coefficient of variation of the $v_{ij}$ are important in applications; they are

$$V\left(v_{ij}\right) = \mathbf{N}_2 - \mathbf{N}_1 \circ \mathbf{N}_1 \tag{12.12}$$

$$SD\left(v_{ij}\right) = \sqrt{V\left(v_{ij}\right)} \tag{12.13}$$

$$CV\left(v_{ij}\right) = \mathcal{D}\,(\text{vec}\,\mathbf{N}_1)^{-1}\,\text{vec}\,SD\left(v_{ij}\right) \tag{12.14}$$

where the square root is taken elementwise. Their derivatives are

$$d\text{vec}\,V = 2\left[\left(\mathbf{N}^\mathsf{T} \otimes \mathbf{I}\right)\mathcal{D}\,(\text{vec}\,\mathbf{I}) + \left(\mathbf{I} \otimes \mathbf{N}_{dg}\right) - \mathcal{D}\,(\text{vec}\,\mathbf{N})\right]d\text{vec}\,\mathbf{N}_1 \tag{12.15}$$

$$d\text{vec}\,SD = \frac{1}{2}\mathcal{D}\left[\text{vec}\,SD\left(v_{ij}\right)\right]^{-1}d\text{vec}\,V \tag{12.16}$$

$$d\text{vec}\, CV = \mathcal{D}\,(\text{vec}\,\mathbf{N}_1)^{-1}\, d\text{vec}\, SD$$
$$- \left[ (\text{vec}\, SD)^\mathsf{T}\, \mathcal{D}\,(\text{vec}\,\mathbf{N}_1)^{-1} \otimes \mathcal{D}\,(\text{vec}\,\mathbf{N}_1)^{-1} \right]$$
$$\times \mathcal{D}\,\left(\text{vec}\,\mathbf{I}_{s^2}\right)\left(\mathbf{1}_{s^2} \otimes \mathbf{I}_{s^2}\right) d\text{vec}\,\mathbf{N}_1 \qquad (12.17)$$

(suppressing the arguments of $V$, $SD$ and $CV$). Because $\mathbf{N}_1$ usually contains zeros, $\mathcal{D}\,(\text{vec}\,\mathbf{N}_1)^{-1}$ must be restricted to the non-zero entries; the coefficient of variation is undefined if the mean is zero.

**Derivation** The fundamental matrix $\mathbf{N}_1 = -\mathbf{U}^{-1}$. Applying (2.82) yields (12.7). The derivatives of the higher moments are obtained by differentiating $\mathbf{N}_2 - \mathbf{N}_4$ in (12.3), (12.4), and (12.5). For example, the differential of $\mathbf{N}_4$ is

$$d\mathbf{N}_4 = 24 \left\{ 3\mathbf{N}_{\mathrm{dg}}^2 \left( d\mathbf{N}_{\mathrm{dg}} \right) \mathbf{N}_1 + \mathbf{N}_{\mathrm{dg}}^3 \left( d\mathbf{N}_1 \right) \right\}, \qquad (12.18)$$

using the fact that $\mathbf{N}_{\mathrm{dg}}$ commutes with itself and $d\mathbf{N}_{\mathrm{dg}}$. Applying the vec operator gives

$$d\text{vec}\,\mathbf{N}_4 = 24 \left\{ 3 \left( \mathbf{N}_2^\mathsf{T} \otimes \mathbf{N}_{\mathrm{dg}}^2 \right) d\text{vec}\,\mathbf{N}_{\mathrm{dg}} + \left( \mathbf{I}_s \otimes \mathbf{N}_{\mathrm{dg}}^3 \right) d\text{vec}\,\mathbf{N}_1 \right\}. \qquad (12.19)$$

Substituting (11.12) for $d\text{vec}\,\mathbf{N}_{\mathrm{dg}}$ and (12.7) for $d\text{vec}\,\mathbf{N}_1$ gives (12.10). Results (12.8) and (12.9) are obtained in similar fashion.

Differentiating the recurrence relationship (12.6) gives

$$d\mathbf{N}_k = k \left( d\mathbf{N}_{\mathrm{dg}} \right) \mathbf{N}_{k-1} + s\mathbf{N}_{\mathrm{dg}} \left( d\mathbf{N}_{k-1} \right). \qquad (12.20)$$

Apply the vec operator,

$$d\text{vec}\,\mathbf{N}_k = k \left( \mathbf{N}_{k-1}^\mathsf{T} \otimes \mathbf{I}_s \right) d\text{vec}\,\mathbf{N}_{\mathrm{dg}} + k \left( \mathbf{I}_s \otimes \mathbf{N}_{\mathrm{dg}} \right) d\text{vec}\,\mathbf{N}_{k-1}, \qquad (12.21)$$

and substitute (11.12) for $d\text{vec}\,\mathbf{N}_{\mathrm{dg}}$ to obtain (12.11).

The derivative of $V$ in (12.15) comes from differentiating (12.12),

$$dV = d\mathbf{N}_2 - 2\mathbf{N}_1 \circ d\mathbf{N}_1, \qquad (12.22)$$

applying the vec operator,

$$D\text{vec}\, V = d\text{vec}\,\mathbf{N}_2 - 2\mathcal{D}\,(\text{vec}\,\mathbf{N}_1)\, d\text{vec}\,\mathbf{N}_1, \qquad (12.23)$$

and then using (12.7) and (12.8). The derivative of $SD\,(v_{ij})$ in (12.16) follows from (2.83). The derivative of $CV\,(v_{ij})$ in (12.17) is obtained using (2.84), with $\mathbf{x} = \text{vec}\, SD$ and $\mathbf{y} = \text{vec}\,\mathbf{N}_1$.

## 12.3   Longevity: Time to Absorption

Let $\eta_j$ be the time to absorption for an individual currently in transient state $j$. The vectors of the $k$th moments of the time to absorption, $\eta_k$, satisfy

$$\eta_1^\mathsf{T} = \mathbf{1}^\mathsf{T}\mathbf{N}_1 \tag{12.24}$$

$$\eta_2^\mathsf{T} = (2)\mathbf{1}^\mathsf{T}\mathbf{N}_1^2 \tag{12.25}$$

$$\eta_3^\mathsf{T} = (6)\mathbf{1}^\mathsf{T}\mathbf{N}_1^3 \tag{12.26}$$

$$\eta_4^\mathsf{T} = (24)\mathbf{1}^\mathsf{T}\mathbf{N}_1^4 \tag{12.27}$$

and in general

$$\eta_k^\mathsf{T} = k\eta_{k-1}^\mathsf{T}\mathbf{N}_1 \qquad k \geq 2 \tag{12.28}$$

(Iosifescu 1980, Thm. 8.6)

   The variance, standard deviation, and coefficient of variation of the time to absorption are

$$V(\eta) = \eta_2 - \eta_1 \circ \eta_1 \tag{12.29}$$

$$SD(\eta) = \sqrt{V(\eta)} \tag{12.30}$$

$$CV(\eta) = \mathcal{D}\left(SD(\eta)\right)^{-1}\eta_1 \tag{12.31}$$

with the square root taken elementwise.

   The derivatives of the moments in (12.24), (12.25), (12.26), and (12.27) are given by

$$d\eta_1 = \left(\mathbf{N}_1^\mathsf{T} \otimes \eta_1^\mathsf{T}\right) d\text{vec}\,\mathbf{U} \tag{12.32}$$

$$d\eta_2 = \left\{2\left[\left(\mathbf{N}_1^\mathsf{T}\right)^2 \otimes \eta_1^\mathsf{T}\right] + 2\left(\mathbf{N}_1^\mathsf{T} \otimes \eta_1^\mathsf{T}\mathbf{N}_1\right)\right\} d\text{vec}\,\mathbf{U} \tag{12.33}$$

$$d\eta_3 = \left\{6\left[\left(\mathbf{N}_1^\mathsf{T}\right)^3 \otimes \eta_1^\mathsf{T}\right] + 6\left[\left(\mathbf{N}_1^\mathsf{T}\right)^2 \otimes \eta_1^\mathsf{T}\mathbf{N}_1\right]\right.$$
$$\left. + 3\left(\mathbf{N}_1^\mathsf{T} \otimes \eta_2^\mathsf{T}\mathbf{N}_1\right)\right\} d\text{vec}\,\mathbf{U} \tag{12.34}$$

$$d\eta_4 = \left\{24\left[\left(\mathbf{N}_1^\mathsf{T}\right)^4 \otimes \eta_1^\mathsf{T}\right] + 24\left[\left(\mathbf{N}_1^\mathsf{T}\right)^3 \otimes \eta_1^\mathsf{T}\mathbf{N}_1\right]\right.$$
$$\left. +12\left[\left(\mathbf{N}_1^\mathsf{T}\right)^2 \otimes \eta_2^\mathsf{T}\mathbf{N}_1\right] + 4\left(\mathbf{N}_1^\mathsf{T} \otimes \eta_3^\mathsf{T}\mathbf{N}_1\right)\right\} d\text{vec}\,\mathbf{U} \tag{12.35}$$

and, recursively,

$$d\boldsymbol{\eta}_k = k\mathbf{N}_1^\mathsf{T} d\boldsymbol{\eta}_{k-1} + k\left(\mathbf{I}_s \otimes \boldsymbol{\eta}_{k-1}^\mathsf{T}\right) d\text{vec}\,\mathbf{N}_1. \tag{12.36}$$

The derivatives of the variance, standard deviation, and coefficient of variation of the time to absorption are (suppressing the arguments)

$$dV = 2\left\{\left[\left(\mathbf{N}_1^\mathsf{T}\right)^2 \otimes \boldsymbol{\eta}_1^\mathsf{T}\right] + \left(\mathbf{N}_1^\mathsf{T} \otimes \boldsymbol{\eta}_1^\mathsf{T}\mathbf{N}_1\right) - \mathcal{D}\left(\boldsymbol{\eta}_1\right)\left(\mathbf{N}_1^\mathsf{T} \otimes \boldsymbol{\eta}_1^\mathsf{T}\right)\right\} d\text{vec}\,\mathbf{U}$$
$$\tag{12.37}$$

$$dSD = \frac{1}{2}\mathcal{D}\,(SD)^{-1}\,dV \tag{12.38}$$

$$dCV = \mathcal{D}\left(\boldsymbol{\eta}_1\right)^{-1} dSD - \left[SD^\mathsf{T}\mathcal{D}\left(\boldsymbol{\eta}_1\right)^{-1} \otimes \mathcal{D}\left(\boldsymbol{\eta}_1\right)^{-1}\right]$$
$$\times \mathcal{D}\,(\text{vec}\,\mathbf{I}_s)\,(\mathbf{1}_s \otimes \mathbf{I}_s)\,d\boldsymbol{\eta}_1. \tag{12.39}$$

**Derivation** Differentiating (12.24) for the expected time to absorption gives

$$d\boldsymbol{\eta}_1^\mathsf{T} = \mathbf{1}_s^\mathsf{T} d\mathbf{N}_1, \tag{12.40}$$

Applying the vec operator, substituting (12.7) for $d\text{vec}\,\mathbf{N}_1$, and simplifying gives (12.32). The derivatives of the higher moments are obtained in the same way; e.g., for $\boldsymbol{\eta}_4$,

$$d\boldsymbol{\eta}_4^\mathsf{T} = (24)\mathbf{1}_s^\mathsf{T}\left[(d\mathbf{N}_1)\,\mathbf{N}_1^3 + \mathbf{N}_1\,(d\mathbf{N}_1)\,\mathbf{N}_1^2 + \mathbf{N}_1^2\,(d\mathbf{N}_1)\,\mathbf{N}_1 + \mathbf{N}_1^3\,(d\mathbf{N}_1)\right]. \tag{12.41}$$

Applying the vec operator yields

$$d\boldsymbol{\eta}_4 = 24\left\{\left[\left(\mathbf{N}_1^\mathsf{T}\right)^3 \otimes \mathbf{1}_s^\mathsf{T}\right] + \left[\left(\mathbf{N}_1^\mathsf{T}\right)^2 \otimes \mathbf{1}_s^\mathsf{T}\mathbf{N}_1\right] + \left[\mathbf{N}_1^\mathsf{T} \otimes \mathbf{1}_s^\mathsf{T}\mathbf{N}_1^2\right]\right.$$
$$\left. + \left[\mathbf{I}_s \otimes \mathbf{1}_s^\mathsf{T}\mathbf{N}_1^3\right]\right\} d\text{vec}\,\mathbf{N}_1. \tag{12.42}$$

Substituting (12.7) for $d\text{vec}\,\mathbf{N}_1$ and simplifying using Eqs. (12.24), (12.25), and (12.26) gives (12.35). The derivatives of the second and third moments, (12.33) and (12.34), are obtained in similar fashion.

   The recursive formula (12.36) is obtained by differentiating (12.28)

$$d\boldsymbol{\eta}_k^\mathsf{T} = k\left(d\boldsymbol{\eta}_{k-1}^\mathsf{T}\right)\mathbf{N}_1 + k\boldsymbol{\eta}_{k-1}^\mathsf{T} d\mathbf{N}_1. \tag{12.43}$$

Apply the vec operator,

$$d\boldsymbol{\eta}_k = k\mathbf{N}_1^\mathsf{T} d\boldsymbol{\eta}_{k-1} + k\left(\mathbf{I}_s \otimes \boldsymbol{\eta}_{k-1}^\mathsf{T}\right) d\text{vec}\,\mathbf{N}_1, \tag{12.44}$$

substitute (12.7) for $d\text{vec}\,\mathbf{N}_1$, and simplify, to obtain (12.36).

Differentiating (12.29) for the variance yields

$$dV = d\boldsymbol{\eta}_2 - 2\boldsymbol{\eta}_1 \circ d\boldsymbol{\eta}_1. \tag{12.45}$$

Applying the vec operator gives

$$dV = d\boldsymbol{\eta}_2 - 2\mathcal{D}\left(\boldsymbol{\eta}_1\right) d\boldsymbol{\eta}_1. \tag{12.46}$$

Substituting (12.32) for $d\boldsymbol{\eta}_1$ and (12.33) for $d\boldsymbol{\eta}_2$ gives the result (12.37). The derivatives of the standard deviation, in (12.38), and the coefficient of variation, in (12.39), are obtained by differentiating (12.30) and (12.31) and applying (2.83) and (2.84).

## 12.4   Multiple Absorbing States and Probabilities of Absorption

Consider a chain that includes $a > 1$ absorbing states. The entry $m_{ij}$ of the $a \times s$ submatrix $\mathbf{M}$ in (12.1) is the rate of transition from transient state $j$ to absorbing state $i$. The probabilities of absorption are defined as

$$b_{ij} = P\left[\text{absorption in } i \,|\text{starting in } j\,\right]. \tag{12.47}$$

The $a \times s$ matrix $\mathbf{B} = \left(b_{ij}\right)$ is

$$\mathbf{B} = \mathbf{M}\mathbf{N}_1 \tag{12.48}$$

(Iosifescu 1980, Section 8.5.6). 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 starting states are of particular interest (e.g., states corresponding to "birth"). 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 $s$. Then

$$d\mathbf{B}(:, j) = \left(\mathbf{e}_j^{\mathsf{T}} \otimes \mathbf{I}_s\right) d\text{vec}\,\mathbf{B}. \tag{12.49}$$

Similarly, row $i$ of $\mathbf{B}$ is $\mathbf{B}(i, :) = \mathbf{e}_i^{\mathsf{T}}\mathbf{B}$ and

$$d\text{vec}\,\mathbf{B}(i, :) = \left(\mathbf{I}_s \otimes \mathbf{e}_i^{\mathsf{T}}\right) d\text{vec}\,\mathbf{B} \tag{12.50}$$

where $\mathbf{e}_i$ is the $i$th unit vector of length $a$. The derivative of $\mathbf{B}$ in (12.49) and (12.50) is

$$d\text{vec}\,\mathbf{B} = \left(\mathbf{N}_1^{\mathsf{T}} \otimes \mathbf{I}\right) d\text{vec}\,\mathbf{M} + \left(\mathbf{N}_1^{\mathsf{T}} \otimes \mathbf{B}\right) d\text{vec}\,\mathbf{U}. \tag{12.51}$$

**Derivations** Differentiating (12.48) yields

$$d\mathbf{B} = (d\mathbf{M})\,\mathbf{N}_1 + \mathbf{M}\,(d\mathbf{N}_1)\,.\tag{12.52}$$

Applying the vec operator and simplifying gives

$$d\text{vec}\,\mathbf{B} = \left(\mathbf{N}_1^\mathsf{T} \otimes \mathbf{I}\right) d\text{vec}\,\mathbf{M} + (\mathbf{I} \otimes \mathbf{M})\,d\text{vec}\,\mathbf{N}_1\tag{12.53}$$

Substituting (12.7) for $d\text{vec}\,\mathbf{N}_1$ and simplifying gives (12.51).

## 12.5 The Embedded Chain: Discrete Transitions Within a Continuous Process

If a continuous-time chain is observed only at the moments when it changes state, the result is a discrete-time process called the embedded Markov chain, or the jump chain, associated with $\mathbf{Q}$ (Iosifescu 1980, Section 8.3.2). The transition matrix of this embedded chain can be written

$$\widehat{\mathbf{P}} = \left(\begin{array}{c|c}\widehat{\mathbf{U}} & \mathbf{0} \\ \hline \widehat{\mathbf{M}} & \mathbf{I}_a\end{array}\right)\tag{12.54}$$

where

$$\widehat{\mathbf{U}} = \mathbf{I}_s - \mathbf{U}\mathbf{U}_{\text{dg}}^{-1}\tag{12.55}$$

$$\widehat{\mathbf{M}} = -\mathbf{M}\mathbf{U}_{\text{dg}}^{-1}\,.\tag{12.56}$$

The embedded chain provides information on the number of visits to each transient state, rather than the time spent in each transient state. The expected numbers of such visits are given by the fundamental matrix

$$\widehat{\mathbf{N}}_1 = \left(\mathbf{I} - \widehat{\mathbf{U}}\right)^{-1}\,.\tag{12.57}$$

The sensitivity analysis of the embedded chain follows directly from the discrete-time results in previous chapters (Chaps. 4 and 5).

In particular, the differential of $\widehat{\mathbf{N}}_1$ is Caswell (2006)

$$d\text{vec}\,\widehat{\mathbf{N}}_1 = \left(\widehat{\mathbf{N}}_1^\mathsf{T} \otimes \widehat{\mathbf{N}}_1\right) d\text{vec}\,\widehat{\mathbf{U}}\,.\tag{12.58}$$

However, this derivative is unlikely to be the sensitivity we are looking for. The continuous-time chain is likely to be parameterized in terms of the rate matrices $\mathbf{U}$ and $\mathbf{M}$, rather than the probability matrices $\widehat{\mathbf{U}}$ and $\widehat{\mathbf{M}}$. To express the perturbation analysis of $\widehat{\mathbf{P}}$ in terms of the parameters of $\mathbf{Q}$ requires the derivatives of the embedded chain with respect to the continuous chain; i.e.,

$$\frac{d\text{vec}\,\widehat{\mathbf{U}}}{d\text{vec}^\mathsf{T}\mathbf{U}} \quad \text{and} \quad \frac{d\text{vec}\,\widehat{\mathbf{M}}}{d\text{vec}^\mathsf{T}\mathbf{M}}.$$

These derivatives are

$$d\text{vec}\,\widehat{\mathbf{U}} = \left[-\left(\mathbf{U}_{\text{dg}}^{-1} \otimes \mathbf{I}_s\right) + \left(\mathbf{U}_{\text{dg}}^{-1} \otimes \mathbf{U}\mathbf{U}_{\text{dg}}^{-1}\right)\mathcal{D}\left(\text{vec}\,\mathbf{I}_s\right)\right]d\text{vec}\,\mathbf{U} \quad (12.59)$$

$$d\text{vec}\,\widehat{\mathbf{M}} = -\left(\mathbf{U}_{\text{dg}}^{-1} \otimes \mathbf{I}_a\right)d\text{vec}\,\mathbf{M} \quad (12.60)$$

$$+ (\mathbf{I}_s \otimes \mathbf{M})\left(\mathbf{U}_{\text{dg}}^{-1} \otimes \mathbf{U}_{\text{dg}}^{-1}\right) \times \mathcal{D}\left(\text{vec}\,\mathbf{I}_s\right)d\text{vec}\,\mathbf{U}.$$

Using (12.59) and (12.61), one can write

$$\frac{d\text{vec}\,\widehat{\mathbf{N}}_1}{d\boldsymbol{\theta}^\mathsf{T}} = \left(\widehat{\mathbf{N}}_1^\mathsf{T} \otimes \widehat{\mathbf{N}}_1\right)\frac{d\text{vec}\,\widehat{\mathbf{U}}}{d\text{vec}^\mathsf{T}\mathbf{U}}\frac{d\text{vec}\,\mathbf{U}}{d\boldsymbol{\theta}^\mathsf{T}}. \quad (12.61)$$

**Derivation** Differentiate $\widehat{\mathbf{U}}$ in (12.55),

$$d\widehat{\mathbf{U}} = -(d\mathbf{U})\,\mathbf{U}_{\text{dg}}^{-1} - \mathbf{U}\left(d\mathbf{U}_{\text{dg}}^{-1}\right), \quad (12.62)$$

apply the vec operator, and use (2.82) and (11.12) for $d\text{vec}\,\mathbf{U}_{\text{dg}}^{-1}$. The result is

$$d\text{vec}\,\widehat{\mathbf{U}} = -\left[\left(\mathbf{U}_{\text{dg}}^{-1}\right)^\mathsf{T} \otimes \mathbf{I}_s\right]d\text{vec}\,\mathbf{U} - (\mathbf{I}_s \otimes \mathbf{U})\,d\text{vec}\,\mathbf{U}_{\text{dg}}^{-1}$$

$$= -\left(\mathbf{U}_{\text{dg}}^{-1} \otimes \mathbf{I}_s\right)d\text{vec}\,\mathbf{U} + (\mathbf{I}_s \otimes \mathbf{U})\left(\mathbf{U}_{\text{dg}}^{-1} \otimes \mathbf{U}_{\text{dg}}^{-1}\right)\mathcal{D}\left(\text{vec}\,\mathbf{I}_s\right)d\text{vec}\,\mathbf{U}$$

which simplifies to give (12.59). Similarly, differentiating $\widehat{\mathbf{M}}$ in (12.56) and applying the vec operator gives

$$d\text{vec}\,\widehat{\mathbf{M}} = -\left(\mathbf{U}_{\text{dg}}^{-1} \otimes \mathbf{I}_a\right)d\text{vec}\,\mathbf{M} - (\mathbf{I}_s \otimes \mathbf{M})\,d\text{vec}\,\mathbf{U}_{\text{dg}}^{-1}. \quad (12.63)$$

Using (2.82) and (11.12) for $d\text{vec}\,\mathbf{U}_{\text{dg}}^{-1}$ and simplifying gives (12.61).

## 12.6    An Example: A Model of Disease Progression

An important area of application of continuous-time Markov chains is the modelling of transitions among disease states. In this context, the time to absorption is longevity, and the time spent in various transient states has implications for the quality of life during the disease. Fix and Neyman (1951) introduced the idea and proposed a 4-state model for cancer, with two transient states (under treatment or not) and two absorbing states (death from cancer or from other causes). Kay (1986) proposed a model with $k$ disease states and an absorbing state representing death. There is now a large literature on such models and their estimation. Recently, studies have proliferated that use Markov chain models of disease transmission to explore the cost-effectiveness of screening and treatment procedures (e.g., Kuo et al. 1999; Chen et al. 1999; Wu et al. 2006; Sonnenberg and Beck 1993).

Sensitivity analysis reveals how these demographic properties respond to changes in parameters. As an example, I consider a model for the progression of colorectal cancer (CRC) that was developed to study the cost-effectiveness of a new CRC screening technique based on DNA testing of stool samples (Wu et al. 2006). The model includes 7 transient states (normal, small and large adenoma, early and late preclinical CRC, and early and late clinical CRC) and 2 absorbing states (death from CRC and death from other causes); see Fig. 12.1. Parameters were estimated from the literature and from clinical studies in Taiwan.

This model, which describes the so-called natural history of the disease, was embedded in a larger decision model to compare the cost-effectiveness of screening strategies. The intensity matrix (12.1) corresponding to Fig. 12.1 is
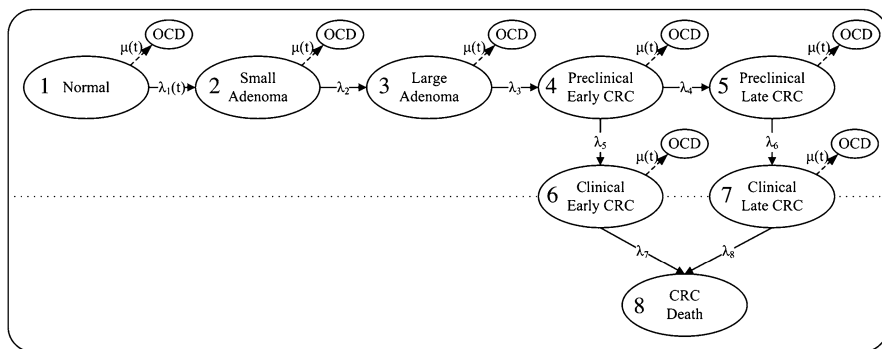


**Fig. 12.1** State transition diagram for an absorbing Markov chain model of colorectal cancer (CRC) progression. The model includes 7 transient states based on the stage of development of adenoma (polyps) or cancer, and two absorbing states corresponding to death from CRC and death from other causes (OCD). Transition rates are given by $\lambda_i$, and mortality rate from other causes by $\mu$. (Modified, under the terms of a Creative Commons Attribution License, from Figure 1 of Wu et al. 2006)

$$
\mathbf{Q} =
\left(
\begin{array}{ccccccc|cc}
-\lambda_1 - \mu & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\lambda_1 & -\lambda_2 - \mu & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & \lambda_2 & -\lambda_3 - \mu & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & \lambda_3 & -\lambda_4 - \lambda_5 - \mu & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & \lambda_4 & -\lambda_6 - \mu & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & \lambda_5 & 0 & -\lambda_7 - \mu & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \lambda_6 & 0 & -\lambda_8 - \mu & 0 & 0 \\
\hline
0 & 0 & 0 & 0 & 0 & \lambda_7 & \lambda_8 & 0 & 0 \\
\mu & \mu & \mu & \mu & \mu & \mu & \mu & 0 & 0
\end{array}
\right).
$$

$$(12.64)$$

The $\lambda_i$ are transition rates; $\mu$ is the mortality rate from other causes of death. The incidence rate of small adenoma ($\lambda_1$) and the mortality rate due to other causes of death ($\mu$) are age-dependent. Here I have analyzed values for age 70; based on figures in Wu et al. (2006). This leads to a parameter vector (all rates are per year):

$$
\boldsymbol{\theta} =
\begin{pmatrix}
\lambda_1 \\
\vdots \\
\lambda_8 \\
\mu
\end{pmatrix}
=
\begin{pmatrix}
1.52 \times 10^{-2} \\
3.46 \times 10^{-2} \\
2.15 \times 10^{-2} \\
3.70 \times 10^{-1} \\
2.38 \times 10^{-1} \\
4.85 \times 10^{-1} \\
3.02 \times 10^{-2} \\
2.10 \times 10^{-1} \\
2.20 \times 10^{-2}
\end{pmatrix}.
$$

$$(12.65)$$

### 12.6.1  Sensitivity Results

The fundamental matrix (12.2) is

$$
\mathbf{N}_1 =
\begin{pmatrix}
26.9 & 0 & 0 & 0 & 0 & 0 & 0 \\
7.2 & 17.7 & 0 & 0 & 0 & 0 & 0 \\
5.7 & 14.0 & 23.0 & 0 & 0 & 0 & 0 \\
0.2 & 0.5 & 0.8 & 1.6 & 0 & 0 & 0 \\
0.1 & 0.4 & 0.6 & 1.2 & 2.0 & 0 & 0 \\
0.9 & 2.2 & 3.6 & 7.2 & 0 & 19.2 & 0 \\
0.3 & 0.7 & 1.2 & 2.4 & 4.1 & 0.00 & 4.3
\end{pmatrix}.
$$

$$(12.66)$$

Thus, given these rates, a 70-year old normal condition individual would expect to spend 27 years in stage 1, and only 0.9 and 0.3 years in stages 6 and 7 (early

and late clinical CRC).[1] Individuals in more advanced stages can expect to spend progressively longer periods in stages 6 and 7 (compare across rows 6 and 7 of $\mathbf{N}_1$).

The standard deviations (12.13) of the times spent in the transient states are

$$
SD\left(v_{ij}\right) =
\begin{pmatrix}
26.9 & 0 & 0 & 0 & 0 & 0 & 0 \\
14.2 & 17.7 & 0 & 0 & 0 & 0 & 0 \\
15.2 & 21.2 & 23.0 & 0 & 0 & 0 & 0 \\
0.8 & 1.1 & 1.4 & 1.6 & 0 & 0 & 0 \\
0.7 & 1.1 & 1.4 & 1.8 & 2.0 & 0 & 0 \\
5.8 & 8.9 & 11.2 & 15.0 & 0 & 19.2 & 0 \\
1.6 & 2.4 & 3.0 & 3.9 & 4.3 & 0 & 4.3
\end{pmatrix}. \tag{12.67}
$$

Clearly, considerable variation can be expected in the times spent in the various states; the standard deviation equals or exceeds the mean in every case.

Considering the sensitivity analysis of the time spent in transient states, focus on the fate of a normal (state 1) individual. The expected times spent in each state by such an individual are give by $\mathbf{N}_1(:, 1)$. From (12.7) and (2.55) the sensitivity and elasticity of $\mathbf{N}(:, 1)$ are

$$
\frac{d\mathbf{N}_1(:, 1)}{d\boldsymbol{\theta}^\mathsf{T}} =
\begin{pmatrix}
-722.6 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -722.6 \\
280.9 & -127.5 & 0 & 0 & 0 & 0 & 0 & 0 & -321.6 \\
223.4 & 64.5 & -132.0 & 0 & 0 & 0 & 0 & 0 & -387.8 \\
7.6 & 2.2 & 4.6 & -0.3 & -0.3 & 0 & 0 & 0 & -13.5 \\
5.6 & 1.6 & 3.4 & 0.2 & -0.2 & -0.3 & 0 & 0 & -10.2 \\
34.8 & 10.0 & 21.0 & -1.4 & 2.3 & 0 & -17.1 & 0 & -79.0 \\
11.6 & 3.4 & 7.0 & 0.3 & -0.5 & 0 & 0 & -1.3 & -22.5
\end{pmatrix}
$$

$$
\frac{\epsilon\mathbf{N}_1(:, 1)}{\epsilon\boldsymbol{\theta}^\mathsf{T}} =
\begin{pmatrix}
-0.4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -0.6 \\
0.6 & -0.6 & 0 & 0 & 0 & 0 & 0 & 0 & -1.0 \\
0.6 & 0.4 & -0.5 & 0 & 0 & 0 & 0 & 0 & -1.5 \\
0.6 & 0.4 & 0.5 & -0.6 & -0.4 & 0 & 0 & 0 & -1.5 \\
0.6 & 0.4 & 0.5 & 0.4 & -0.4 & -1.0 & 0 & 0 & -1.5 \\
0.6 & 0.4 & 0.5 & -0.6 & 0.6 & 0 & -0.6 & 0 & -1.9 \\
0.6 & 0.4 & 0.5 & 0.4 & -0.4 & 0.0 & 0 & -0.9 & -1.7
\end{pmatrix}. \tag{12.68}
$$

These elasticities imply that a 1% increase in $\lambda_1$ will (to first order) cause about a 0.4% decrease in the mean time spent in the normal state and a 0.6% increase in the mean time spent in each other state. A 1% increase in $\lambda_4$ (the rate of transition between early and late preclinical CRC) creates a 0.6% decrease in the time spent

---

[1]This calculation holds the mortality rate fixed at its values at age 70; in reality it increases with age. Wu et al. (2006) included age variation by providing values of $\lambda_1$ (the rate of progression from normal to small adenoma) specific to 5-year intervals from 50 to 70 years of age; all other parameters were age-invariant.

in stages 4 and 6 (the early CRC stages) and a 0.4% increase in the time spent in stages 5 and 7 (the late CRC stages). An increase in the mortality rate $\mu$ due to other causes of death reduces the time spent in any of the transient states.

The elasticity of the variance in the time spent in the transient states by an individual in state 1 is

$$
\frac{\epsilon V(v_{i1})}{\epsilon \boldsymbol{\theta}^{\mathsf{T}}} =
\begin{pmatrix}
-0.8 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1.2 \\
0.4 & -1.2 & 0 & 0 & 0 & 0 & 0 & 0 & -1.2 \\
0.5 & 0.3 & -1.0 & 0 & 0 & 0 & 0 & 0 & -1.8 \\
0.5 & 0.4 & 0.5 & -1.2 & -0.8 & 0 & 0 & 0 & -1.5 \\
0.6 & 0.4 & 0.5 & 0.4 & -0.4 & -1.9 & 0 & 0 & -1.6 \\
0.6 & 0.4 & 0.5 & -0.6 & 0.6 & 0 & -1.2 & 0 & -2.3 \\
0.6 & 0.4 & 0.5 & 0.4 & -0.4 & 0.0 & 0 & -1.8 & -1.7
\end{pmatrix}.
\tag{12.69}
$$

The sign pattern is the same as that of the elasticities of the mean times in (12.68), so we conclude that any parameter change that increases the mean time spent in a transient state will also increase the variance in that time. The elasticities of the variance are comparable to those of the mean (cf. (12.68) and (12.69)), showing that the means and the variance respond with roughly equal proportional changes.

Longevity is measured by the time to absorption, and is a primary concern in analyses of screening or treatment protocols. The vectors of the mean, standard deviation, and coefficient of variation of longevity are

$$
\boldsymbol{\eta}_1 =
\begin{pmatrix}
41.4 \\
35.5 \\
29.1 \\
12.4 \\
6.1 \\
19.2 \\
4.3
\end{pmatrix}
\qquad
SD(\boldsymbol{\eta}) =
\begin{pmatrix}
37.4 \\
30.3 \\
25.8 \\
14.1 \\
4.7 \\
19.2 \\
4.3
\end{pmatrix}
\qquad
CV(\boldsymbol{\eta}) =
\begin{pmatrix}
0.9 \\
0.9 \\
0.9 \\
1.1 \\
0.8 \\
1.0 \\
1.0
\end{pmatrix}.
\tag{12.70}
$$

The sensitivity and elasticity of expected longevity (life expectancy) with respect to $\boldsymbol{\theta}$ are

$$
\frac{d\boldsymbol{\eta}_1}{d\boldsymbol{\theta}^{\mathsf{T}}} =
\begin{pmatrix}
-158.7 & -45.8 & -96.0 & -1.2 & 1.3 & -0.2 & -17.1 & -1.3 & -1557.2 \\
0 & -112.2 & -234.9 & -3.0 & 3.2 & -0.6 & -41.9 & -3.2 & -1089.1 \\
0 & 0 & -384.2 & -5.0 & 5.3 & -1.0 & -68.6 & -5.2 & -756.5 \\
0 & 0 & 0 & -10.0 & 10.7 & -2.1 & -138.8 & -10.4 & -176.0 \\
0 & 0 & 0 & 0 & 0 & -3.5 & 0 & -17.8 & -29.8 \\
0 & 0 & 0 & 0 & 0 & 0 & -367.0 & 0 & -367.0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & -18.6 & -18.6
\end{pmatrix}
$$

$$\frac{\epsilon \boldsymbol{\eta}_1}{\epsilon \boldsymbol{\theta}^\mathsf{T}} = \begin{pmatrix} -0.06 & -0.04 & -0.05 & -0.01 & 0.01 & -0.00 & -0.01 & -0.01 & -0.83 \\ 0 & -0.11 & -0.14 & -0.03 & 0.02 & -0.01 & -0.04 & -0.02 & -0.68 \\ 0 & 0 & -0.28 & -0.06 & 0.04 & -0.02 & -0.07 & -0.04 & -0.57 \\ 0 & 0 & 0 & -0.30 & 0.21 & -0.08 & -0.34 & -0.18 & -0.31 \\ 0 & 0 & 0 & 0 & 0 & -0.28 & 0 & -0.61 & -0.11 \\ 0 & 0 & 0 & 0 & 0 & 0 & -0.58 & 0 & -0.42 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -0.91 & -0.09 \end{pmatrix} . \qquad (12.71)$$

Almost all the nonzero elements are negative, because increasing any of the rates leading towards clinical CRC reduces life expectancy, as does increasing the mortality rate due to other causes of death. The exceptions are the sensitivities and elasticities of $\boldsymbol{\eta}_1$ to $\lambda_5$ (in column 5 of these matrices), which are positive because $\lambda_5$ delays the onset of clinical CRC (cf. Fig. 12.1).

The elasticities of $E(\eta_1)$, the life expectancy of a normal individual, to a change in $\boldsymbol{\theta}$, appear in the first row of (12.71). The largest of these (except for the last column, representing mortality from other causes of death) are to changes in $\lambda_1$, $\lambda_2$, and $\lambda_3$, the rates of transition from normal to small adenoma, small to large adenoma, and large adenoma to preclinical CRC. The rates $\lambda_2$ and $\lambda_3$ have large effects on $E(\eta_2)$, and $\lambda_3$ has a large effect on $E(\eta_3)$. These transitions are targets of screening and early treatment; this analysis quantifies the effect that such interventions could have.

The sensitivity and elasticity of the standard deviation of longevity are

$$\frac{dSD(\boldsymbol{\eta})}{d\boldsymbol{\theta}^\mathsf{T}} = \begin{pmatrix} -0.27 & -0.07 & -0.16 & -0.00 & 0.00 & -0.00 & -0.03 & -0.00 & -1.19 \\ 0 & -0.13 & -0.31 & -0.00 & 0.00 & -0.00 & -0.06 & -0.00 & -0.76 \\ 0 & 0 & -0.43 & -0.00 & 0.00 & -0.00 & -0.09 & -0.00 & -0.61 \\ 0 & 0 & 0 & -0.01 & 0.01 & 0.00 & -0.27 & 0.00 & -0.27 \\ 0 & 0 & 0 & 0 & 0 & -0 & 0.00 & -0.02 & -0.02 \\ 0 & 0 & 0 & 0 & 0 & 0 & -0.37 & 0 & -0.37 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -0.02 & -0.02 \end{pmatrix} \times 10^3$$

$$\qquad (12.72)$$

and

$$\frac{\epsilon SD(\boldsymbol{\eta})}{\epsilon \boldsymbol{\theta}^\mathsf{T}} = \begin{pmatrix} -0.11 & -0.06 & -0.09 & -0.02 & 0.01 & -0.00 & -0.02 & -0.01 & -0.70 \\ 0 & -0.15 & -0.22 & -0.04 & 0.03 & -0.00 & -0.06 & -0.01 & -0.55 \\ 0 & 0 & -0.36 & -0.05 & 0.05 & -0.00 & -0.11 & -0.01 & -0.52 \\ 0 & 0 & 0 & -0.23 & 0.23 & 0.01 & -0.58 & 0.00 & -0.43 \\ 0 & 0 & 0 & 0 & 0 & -0.16 & 0.00 & -0.75 & -0.09 \\ 0 & 0 & 0 & 0 & 0 & 0 & -0.58 & 0 & -0.42 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -0.91 & -0.09 \end{pmatrix} . \qquad (12.73)$$

These have the same sign pattern as the sensitivity of $\boldsymbol{\eta}_1$, indicating that any increase in life expectancy will be accompanied by an increase in the variance of longevity. The coefficient of variation takes this joint change into account; from (12.39),

$$\frac{\epsilon CV(\eta)}{\epsilon \theta^{\mathsf{T}}} = \begin{pmatrix} 0.04 & 0.02 & 0.03 & 0.00 & -0.00 & -0.00 & 0.01 & -0.00 & -0.31 \\ 0 & -0.00 & 0.02 & -0.01 & 0.00 & -0.01 & 0.01 & -0.01 & -0.38 \\ 0 & 0 & -0.01 & -0.03 & 0.01 & -0.02 & 0.01 & -0.04 & -0.21 \\ 0.00 & 0.00 & 0.00 & -0.00 & -0.07 & -0.08 & 0.32 & -0.14 & 0.19 \\ 0 & 0 & 0 & 0.00 & 0.00 & -0.30 & 0.00 & -0.27 & -0.09 \\ 0 & 0 & 0 & 0 & 0 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.00 & 0.00 \end{pmatrix}. \quad (12.74)$$

Most of these elasticities are small, suggesting that the mean and standard deviation respond roughly proportionally, so that the $CV$ does not change much.

The matrix $\mathbf{B}$ in (12.48), giving the ultimate probability of death from CRC (row 1) or other causes of death (row 2) is

$$\mathbf{B} = \begin{pmatrix} 0.1\ 0.2\ 0.4\ 0.7\ 0.9\ 0.6\ 0.9 \\ 0.9\ 0.8\ 0.6\ 0.3\ 0.1\ 0.4\ 0.1 \end{pmatrix}. \quad (12.75)$$

Focusing on the probability of death due to CRC, the sensitivity and elasticity, from (12.50), are

$$\frac{d\text{vec}\,\mathbf{B}(1,:)}{d\theta^{\mathsf{T}}} = \begin{pmatrix} 3.5 & 1.0 & 2.1 & 0.0 & -0.0 & 0.0 & 0.4 & 0.0 & -7.1 \\ 0 & 2.5 & 5.2 & 0.1 & -0.1 & 0.0 & 0.9 & 0.1 & -11.5 \\ 0 & 0 & 8.4 & 0.1 & -0.1 & 0.0 & 1.5 & 0.1 & -12.5 \\ 0 & 0 & 0 & 0.2 & -0.2 & 0.1 & 3.0 & 0.2 & -8.5 \\ 0 & 0 & 0 & 0 & 0 & 0.1 & 0.00 & 0.4 & -5.4 \\ 0 & 0 & 0 & 0 & 0 & 0 & 8.1 & 0 & -11.1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.4 & -3.9 \end{pmatrix}$$

$$\frac{\epsilon \text{vec}\,\mathbf{B}(1,:)}{\epsilon \theta^{\mathsf{T}}} = \begin{pmatrix} 0.6 & 0.4 & 0.5 & 0.1 & -0.1 & 0.0 & 0.1 & 0.1 & -1.7 \\ 0 & 0.4 & 0.5 & 0.1 & -0.1 & 0.0 & 0.1 & 0.1 & -1.2 \\ 0 & 0 & 0.5 & 0.1 & -0.1 & 0.0 & 0.1 & 0.1 & -0.8 \\ 0 & 0 & 0 & 0.1 & -0.1 & 0.0 & 0.1 & 0.0 & -0.3 \\ 0 & 0 & 0 & 0 & 0 & 0.0 & 0 & 0.1 & -0.1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.4 & 0 & -0.4 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.1 & -0.1 \end{pmatrix}.$$

The probability of death from CRC could be reduced by increasing the mortality rate due to other causes (last column), although this is not an attractive treatment option. A more useful interpretation of the last column is as an indication of the increase in death from CRC that would result from reducing other causes of death.

For normal individuals, the risk of death from CRC is most elastic to changes in $\lambda_2$, $\lambda_3$, and $\lambda_4$ (row 1). The row sums of the elasticity matrix, corresponding to the effects of a proportional change in all rates, sum to zero because a change of time scale does not affect the probability of absorption.

### 12.6.2   Sensitivity of the Embedded Chain

The transition matrix $\widehat{\mathbf{P}}$ in (12.54) for the embedded chain is

$$
\widehat{\mathbf{P}} = \left(\begin{array}{ccccccc|cc}
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0.41 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0.61 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0.49 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0.59 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0.38 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0.96 & 0 & 0 & 0 & 0 \\
\hline
0 & 0 & 0 & 0 & 0 & 0.58 & 0.91 & 0 & 0 \\
0.59 & 0.39 & 0.51 & 0.03 & 0.04 & 0.42 & 0.09 & 0 & 0
\end{array}\right). \tag{12.76}
$$

The fundamental matrix $\widehat{\mathbf{N}}_1$ from (12.57) is

$$
\widehat{\mathbf{N}}_1 = \begin{pmatrix}
1.0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0.4 & 1.0 & 0 & 0 & 0 & 0 & 0 \\
0.2 & 0.6 & 1.0 & 0 & 0 & 0 & 0 \\
0.1 & 0.3 & 0.5 & 1.0 & 0 & 0 & 0 \\
0.1 & 0.2 & 0.3 & 0.6 & 1.0 & 0 & 0 \\
0.1 & 0.1 & 0.2 & 0.4 & 0 & 1.0 & 0 \\
0.1 & 0.2 & 0.3 & 0.6 & 1.0 & 0 & 1.0
\end{pmatrix}. \tag{12.77}
$$

In this continuous-time chain, states cannot be re-entered (cf. Fig. 12.1). Because a state can be visited at most once, the mean number of visits is also the probability of ever entering the state. Thus the probabilities that a normal individual will ever suffer early or late clinical CRC are $\widehat{\mathbf{N}}_1(6, 1) = 0.1$, and $\widehat{\mathbf{N}}_1(7, 1) = 0.07$, respectively. These probabilities increase for individuals in successively later stages; for an individual with large adenoma the probabilities are $\widehat{\mathbf{N}}_1(6.3) = 0.2$ and $\widehat{\mathbf{N}}_1(7, 3) = 0.3$, respectively.

Focusing sensitivity analysis on individuals in the normal state (state 1), the sensitivities and elasticities of the number of visits are

$$
\frac{d\widehat{\mathbf{N}}_1(:, 1)}{d\boldsymbol{\theta}^{\mathsf{T}}} = \begin{pmatrix}
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
15.9 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -11.0 \\
9.7 & 2.8 & 0 & 0 & 0 & 0 & 0 & 0 & -11.1 \\
4.8 & 1.4 & 2.9 & 0 & 0 & 0 & 0 & 0 & -8.3 \\
2.8 & 0.8 & 1.7 & 0.1 & -0.1 & 0 & 0 & 0 & -5.0 \\
1.8 & 0.5 & 1.1 & -0.1 & 0.1 & 0 & 0 & 0 & -3.2 \\
2.7 & 0.8 & 1.6 & 0.1 & -0.1 & 0.0 & 0 & 0 & -4.9
\end{pmatrix} \tag{12.78}
$$

and

$$\frac{\epsilon \widehat{\mathbf{N}}_1(:, 1)}{\epsilon \boldsymbol{\theta}^{\mathsf{T}}} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0\ 0\ 0 & 0 \\ 0.6 & 0 & 0 & 0 & 0 & 0\ 0\ 0 & -0.6 \\ 0.6 & 0.4 & 0 & 0 & 0 & 0\ 0\ 0 & -1.0 \\ 0.6 & 0.4 & 0.5 & 0 & 0 & 0\ 0\ 0 & -1.5 \\ 0.6 & 0.4 & 0.5 & 0.4 & -0.4 & 0\ 0\ 0 & -1.5 \\ 0.6 & 0.4 & 0.5 & -0.6 & 0.6 & 0\ 0\ 0 & -1.5 \\ 0.6 & 0.4 & 0.5 & 0.41 & -0.4 & 0.04\ 0\ 0 & -1.5 \end{pmatrix}. \qquad (12.79)$$

The sensitivities and elasticities of the probability of contracting clinical CRC are given by the last two rows. These probabilities are highly elastic to $\lambda_1$, $\lambda_2$ and $\lambda_3$. The elasticities to $\mu$ indicate that every 1% reduction in mortality due to other causes will cause about a 1.5% increase in the probability of experiencing clinical CRC.

## 12.7   Discussion

The results of this chapter have been presented in terms of differentials of, or derivatives with respect to, a general vector $\boldsymbol{\theta}$ of parameters. The nature of these parameters and their relation to $\mathbf{Q}$, $\mathbf{U}$, or $\mathbf{M}$ can be very general. At its simplest, $\boldsymbol{\theta}$ could consist of some subset of the elements of $\mathbf{Q}$. This is the case in the CRC example (Sect. 12.6), in which the parameters are transition rates $\lambda_i$ and mortality rates $\mu_i$. More generally, the transition rates might themselves be written as functions of other variables. For example, in Van Den Hout and Matthews (2009a,b) the rates are written as $q_{ij} = \exp\left(\boldsymbol{\beta}_{ij}^{\mathsf{T}} \mathbf{z}\right)$, $i \neq j$, where $\mathbf{z}$ is a vector of covariates (e.g., age, medical care) and $\boldsymbol{\beta}_{ij}$ is a vector of coefficients to be estimated. The results presented here can be applied directly to such cases, and indeed to even more complicated functional dependencies, using the chain rule. Thus, focusing on parametric dependence is not only scientifically valuable (these are, after all, the relationships of interest in applications of Markov chains) but also extremely general.

Epidemic models are often written as continuous-time Markov chains, specified in terms of rates of movement among infection states. Gómez-Corral and López-García (2018) extended the methods of this chapter to a model in which individuals are classified by two state variables (a level-dependent quasi-birth-death process). The model may be considered a continuous-time analog of the age×stage models of Chap. 6 (Caswell 2012; Caswell and Salguero-Gómez 2013; Caswell et al. 2018). Their approach takes advantage of the block structure of the intensity matrix for such processes. They have also applied the approach to receptor-ligand complexes within cells (López-García et al. 2018). As far removed from demography as molecules may seem, the concepts of i-state transitions, of inferring population behavior from individual trajectories, and of sensitivity analysis still apply. That's a good thing.

# References

Altman, E., K. Avrachenkov, and R. Núñez-queija. 2004. Perturbation analysis for denumerable Markov chains with application to queueing models. Advances in Applied Probability **36**:839–853.

Cao, X. 1989. Estimates of performance sensitivity of a stochastic system. IEEE Transactions on Information Theory **35**:1058–1068.

Cao, X., X. Yuan, and L. Qiu. 1996. A single sample path-based performance sensitivity formula for Markov chains. IEEE Transactions on Automatic Control **41**:1814–1817.

Caswell, H., 2006. Applications of Markov chains in demography. Pages 319–334 *in* MAM2006: Markov Anniversary Meeting. Boson Books, Raleigh, North Carolina.

Caswell, H. 2012. Matrix models and sensitivity analysis of populations classified by age and stage: a vec-permutation matrix approach. Theoretical Ecology **5**:403–417.

Caswell, H., C. de Vries, N. Hartemink, G. Roth, and S. F. van Daalen. 2018. Age×stage-classified demographic analysis: a comprehensive approach. Ecological Monographs **88**:560–584.

Caswell, H., and R. Salguero-Gómez. 2013. Age, stage and senescence in plants. Journal of Ecology **101**:585–595.

Chen, T.-H., M.-F. Yen, S.-S. Lai, K. S-L, W. C-Y, W. J-M, T. C. Prevost, and D. S. W. 1999. Evaluation of a selective screening for colorectal carcinoma: the Taiwan Multicenter Cancer Screening (TAMCAS) Project. Cancer **86**:1116–1128.

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.

Fix, E., and J. A. Neyman. 1951. A simple stochastic model of recovery, relapse, death and loss of patients. Human Biology **23**:205–241.

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.

Glasserman, P. 1992. Derivative estimates from simulation of continuous-time Markov chains. Operations Research **40**:292–308.

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.

Gómez-Corral, A., and M. López-García. 2018. Perturbation analysis in finite LD-QBD processes and applications to epidemic models. Numerical Linear Algebra with Applications page e2160.

Hunter, J. J. 2005. Stationary distributions and mean first passage times of perturbed Markov chains. Linear Algebra and its Applications **410**:217–243.

Iosifescu, M. 1980. Finite Markov Processes and Their Applications. Wiley, New York, New York.

Kay, R. A. 1986. Markov model for analysing cancer markers and disease states in survival studies. Biometrics **42**:855–865.

Kuo, H. S., H. J. Chang, P. Chou, L. Teng, and T. H. H. Chan. 1999. A Markov chain model to assess the efficacy of screening for non-insulin dependent diabetes mellitus (NIDDM). International Journal of Epidemiology **28**:233–240.

López-García, M., M. Nowicka, C. Bendtsen, G. Lythe, S. Ponnambalam, and C. Molina-París. 2018. Quantifying the phosphorylation timescales of receptor–ligand complexes: a Markovian matrix-analytic approach. Open Biology **8**:180126.

Meyer, C. D. 1975. The role of the group generalized inverse in the theory of finite Markov chains. SIAM Review **17**:443–464.

Mitrophanov, A. Y. 2004. The spectral gap and perturbation bounds for reversible continuous-time Markov chains. Journal of Applied Probability **41**:1219–1222.

Ramesh, A. V., and K. Trivedi, 1993. On the sensitivity of transient solutions of Markov models. Pages 122–134 *in* Proceedings of the 1993 ACM SIGMETRICS Conference on measurement and modeling of computer systems.

Schweitzer, P. J. 1968. Perturbation theory and finite Markov chains. Journal of Applied Probability **5**:401–413.

Seneta, E. 1993. Sensitivity of finite Markov chains under perturbation. Statistics and Probability Letters **17**:163–168.

Sonnenberg, F. A., and R. Beck. 1993. Markov models in medical decision making: a practical guide. Medical Decision Making **13**:322–338.

Van Den Hout, A., and F. E. Matthews. 2009a. Estimating dementia-free life expectancy for Parkinson's patients using Bayesian inference and microsimulation. Biostatistics **10**:729–743.

Van Den Hout, A., and F. E. Matthews. 2009b. A piecewise-constant Markov model and the effects of study design on the estimation of life expectancies in health and ill health. Statistical Methods in Medical Research **18**:145–162.

Wu, G.-M., Y.-M. Wang, M.-F. Yen, J.-M. Wong, H.-C. Lai, J. Warwick, and C. TH-H. 2006. Cost-effectiveness analysis of colorectal cancer screening with stool DNA testing in intermediate-incidence countries. BMC Cancer **6**:136.