Sensitivity Analysis: Matrix Methods in Demography and Ecology

Hal Caswell

Sensitivity Analysis: Matrix Methods in Demography and Ecology

Springer Open

Demographic Research Monographs

A Series of the Max Planck Institute
for Demographic Research

Editor-in-chief
Mikko Myrskylä
Max Planck Institute for Demographic Research
Rostock, Germany

More information about this series at http://www.springer.com/series/5521

Hal Caswell

Sensitivity Analysis: Matrix Methods in Demography and Ecology

Springer Open

Hal Caswell
Biodiversity & Ecosystem Dynamics
University of Amsterdam
Amsterdam, The Netherlands

For Moira

Preface

Sensitivity analysis addresses one of the most persistent of all questions: what would happen *if*? Within the field of demography, sensitivity analysis might be said to have originated with the groundbreaking, yet very different, papers of Hamilton (1966) and Keyfitz (1971). Hamilton calculated the sensitivity of the intrinsic rate of increase, r, to changes in age-specific mortality. He interpreted r as a measure of individual fitness, capturing the effects of the phenotype on mortality and fertility. The resulting sensitivities are measures of the strength of natural selection on aging and senescence. Keyfitz calculated sensitivities of population growth rate, life expectancy, and other quantities. Taking a demographic perspective, he interpreted the results as showing the linkage between age-specific rates at the individual level and the "intrinsic" rates expressed at the population level. Both these perspectives on sensitivity analysis continue to play major roles in demography and population biology. Connecting traits to individual rates, and those rates to measures of fitness, is the foundation of evolutionary demography. Understanding linkages between individual rates and population outcomes informs population projections, policy and spending, conservation, health demography, ecotoxicology, and so on.

Fast forward to today. The diversity of demographic models, of the outcomes that can be calculated, and the power of the mathematical tools available to analyze them far exceed those of 50 years ago. Much of this progress is due to the formulation of demographic models in terms of matrices. P. H. Leslie formulated matrix models in the 1940s (Leslie 1945), but they were mostly ignored for two decades until revitalized by a series of studies in the 1960s (Keyfitz 1964; Lefkovitch 1965; Rogers 1968). In the very first issue of the first volume of the new journal *Demography*, Nathan Keyfitz described population projection as a matrix operator (Keyfitz 1964). This book relies on matrix formulations generalized beyond projections to age-structured and stage-structured populations, to linear and nonlinear dynamics, to time-invariant and time-varying vital rates, and to multistate models that combine age and stage information.

The matrix formulation provides easily computable outcomes at the level of the individual (e.g., risks of mortality, longevity, lifetime reproduction), the cohort (e.g., distributions of age or stage at death), and the population (e.g., population growth rate). The mathematical connection of matrix models and the theory of finite-state Markov chains make it possible to go beyond expected outcomes to calculate variances and higher moments and to take full advantage of the stochasticity of demographic events at the individual level (individual stochasticity).

The sensitivity analysis of these diverse outcomes is made possible by the even more recently developed mathematical tool of matrix calculus (Magnus and Neudecker 1988). Matrix calculus permits easy differentiation of scalar-, vector-, and matrix-valued functions of scalar-, vector-, and matrix-valued arguments. This entire book is an application of these methods to demographic problems.

Organization The book is (imperfectly) divided into five parts. Part I contains an introduction and a summary of the matrix calculus methods that are used throughout the book.

Part II analyzes linear models for population growth, longevity, and reproduction. In linear models, the per-capita vital rates are independent of population size and structure. When the rates are also time-invariant, these models lead to a stable age or stage structure and exponential growth. The rate of growth is one of the most fundamental outcomes of stable population theory. Chapter 3 analyzes the sensitivity of population growth rate from three directions: differentiation of the characteristic equation, eigenvalue perturbation theory, and matrix calculus, providing the first application of the methods that form the basis of the subsequent chapters. Chapter 4 focuses on longevity, presenting the sensitivity analysis of life expectancy, variance in longevity, and life disparity. Chapter 5 introduces the important concept of individual stochasticity (stochastic outcomes of probabilistic transitions in the life cycle) and explores its effects on longevity, net reproductive rate, birth intervals, and age at reproduction. Some aspects of time variation are introduced, including the first appearance in the book of the powerful vec-permutation matrix method to describe temporally varying environments.

A critical first step in the construction of any demographic model is the choice of the individual state (i-state) variables that capture the relevant information about individuals. Age, developmental stage, body size, and a variety of other properties have been used as i-states. However, it is often the case that a combination of age and some other characteristic is necessary to describe individuals. Chapter 6 presents the sensitivity analysis of such models, using the vec-permutation method to construct multistate models and matrix calculus to differentiate the results.

Part III relaxes the assumption of time invariance. Chapter 7 presents the sensitivity analysis of transient dynamics, i.e., dynamics that happen in the short term, before asymptotic behavior appears. Short-term population growth and structure may differ in important ways from the growth and structure implied by stable population theory. Chapter 7 explores these differences, for cases where the vital rates may be fixed, varying, or even nonlinear. Chapter 8 analyzes periodic models. Such models appear in a variety of guises: as matrix products

describing periodic (e.g., seasonal) environmental variation and as matrix products describing distinct processes embedded within an apparently single projection matrix and in the construction of multistate matrix models. In each case, the goal is to describe the sensitivity of some overall outcome, calculated from the entire periodic matrix product, to changes in parameters affecting each component of the matrix. Chapter 9 analyzes population growth in stochastic environments and the problem of decomposing differences in stochastic growth rates into components due to the environment and to the vital rates. This requires a combination of the first-order approximate decomposition known as life table response experiment (LTRE) analysis with the more specialized Kitagawa-Keyfitz decomposition and has potential implications far beyond the stochastic environment case.

Part IV analyzes nonlinear models, including density-dependent models, frequency-dependent models (e.g., models for the interaction of the sexes), nonlinear models for subsidized populations, and a nonlinear approach to the sensitivity of the stable structure and the reproductive value of linear models.

Finally, Part V returns to the analysis of the Markov chain models that form the basis of many of the demographic calculations throughout the book. These chapters take a more mathematical approach to the sensitivity analysis of Markov chains, including some aspects that have yet to find wide demographic application (but the potential is there). Chapter 11 analyzes discrete-time chains, both the absorbing chains familiar in demography (death is an absorbing state in most models) and ergodic chains that include no absorbing states. Chapter 12 presents the sensitivity analysis of continuous-time absorbing Markov chains, using as an example of a model for the stages of colorectal cancer.

Most of the chapters here are based on, or extended from, papers that have appeared in a variety of journals in ecology, population biology, human demography, and applied mathematics. There is overlap among the chapters. This is a feature, not a bug, because it means that similar calculations are revisited with different perspectives, different derivations, and different examples. When choices arose, I tried to choose the presentation that would make things easier for the reader.

The material here certainly does not exhaust the applications of matrix calculus in the sensitivity analysis of demographic models. I have tried to point out directions for further development.

Bibliography

Hamilton, W. D. 1966. The moulding of senescence by natural selection. Journal of Theoretical Biology **12**:12–45.

Keyfitz, N. 1964. The population projection as a matrix operator. Demography **1**:56–73.

Keyfitz, N. 1971. Linkages of intrinsic to age-specific rates. Journal of the American Statistical Association **66**:275–281.

Lefkovitch, L. P. 1965. The study of population growth in organisms grouped by stages. Biometrics **21**:1–18.

Leslie, P. H. 1945. On the use of matrices in certain population mathematics. Biometrika **33**:183–212.

Magnus, J. R., and H. Neudecker. 1988. Matrix differential calculus with applications in statistics and econometrics. John Wiley and Sons, New York, New York.

Rogers, A. 1968. Matrix Analysis of Interregional Population Growth and Distribution. University of California Press, Berkeley, California.

Amsterdam, The Netherlands   Hal Caswell

Acknowledgements

---

[1] Yes, I know, humans are animals. But it is just unbearably clumsy to write "human and non-human animals" every time.

Heterogeneity in Plant and Animal Demography. This grant and the team that it permitted me to assemble were essential to this research.

People There is a long list of people who deserve thanks (but no blame) for this book. Special thanks to my research group at the University of Amsterdam: Silke van Daalen, Charlotte de Vries, Gregory Roth, Nienke Hartemink, Nora Sanchez Gassen, and Christina Bohk-Ewald. Mike Neubert and Stephanie Jenouvrier at WHOI have been particularly valuable collaborators. My student, Esther Shyu, helped push sensitivity analysis into new directions. Joel Cohen inspired more of this analysis than may be apparent. I thank Nathan Keyfitz for his example. Shripad Tuljapurkar, Carol Horvitz, and Ulrich Steiner have also explored this territory, and discussions with them have been especially valuable. I have presented courses and workshops on sensitivity analysis at the MPIDR and at meetings of the Ecological Society of America, and participants in those workshops have provided valuable feedback.

I have had the good fortune to collaborate with many researchers on sensitivity analysis, including Azmy Ackleh, Annette Baudisch, Christina Bohk-Ewald, Solange Brault, Silke van Daalen, Michal Engelman, Masami Fujiwara, Nienke Hartemink, Carol Horvitz, Christine Hunter, Stephanie Jenouvrier, Petra Klepac, Tiffany Knight, Eleanor Pardini, Alyson van Raalte, Bonnie Ripley, Gregory Roth, Roberto Salguero-Gomez, Nora Sanchez Gassen, Esther Shyu, Carly Strasser, Yngvild Vindenes, Charlotte de Vries, Martin Wensink, Virginia Zarulli, and Ariane Verdy.

The most tired cliché in book-writing is the one where the author thanks a partner whose support has been essential to completion of the work. Clichés, however, are sometimes true, and, in this case, I owe a huge thanks to my wife, Moira Powers, for her unfailing support.

Contents

Part I
Introductory and Methodological

Chapter 1
Introduction: Sensitivity Analysis – What and Why?

1.1 Introduction

Demography is a science that connects individual processes and events to the development of cohorts and then to the dynamics of populations. It does so with mathematical models that distinguish among individuals based on their characteristics.[1] The most familiar such model is the life table, which records mortality and fertility of the individual as a function of age, and is used to calculate properties of cohorts (e.g., the distribution of age at death) and populations (e.g., the intrinsic rate of increase).

The life table is the most familiar, but demography has proceeded far beyond that in both models and analyses. In any case, though, a model is defined first by its structure (the states of individuals and the transitions possible among them), then by the rates at which individuals develop, survive, and reproduce throughout the life cycle, then by the functional dependence of those rates (time-invariant or time-varying, density-independent or density-dependent, deterministic or stochastic), and finally by the values of the parameters that define the rates. A set of parameters operating within a given model generates the demographic outcomes calculated from the model (population growth rate, population structure, equilibria, cycles, measures of longevity, state occupancy times, transient behavior and projections, and so on). The *sensitivity problem* is to understand how the outcome[s] change in response to changes in the parameters.

---

[1]Technically, these characteristics are known as individual state variables, or i-states (Metz and Diekmann 1986; Caswell 2001). Their task is to capture all the information about the individual's history that is relevant to determining its future fate, and a major task of demography is to discover those aspects of the individual necessary for a successful i-state (e.g., de Vries and Caswell 2017). In the models considered here, the population state (p-state) is a distribution function over the set of i-states. Thus, for example, age as an i-state leads to a population described by its age distribution.

Why should we care about the effects of change?

- We may be concerned with particular changes that we have reason to believe will occur (due to, e.g., changes in society, changes in the environment, changes in policy) and want to know how they will affect the outcome of interest.
- We may want to evaluate the effect of changes that we hope to cause as matters of policy, or to compare alternative policies for their effects.
- We may be interested in evolutionary demographic questions. Natural selection is a process that explores the consequences for fitness (which is itself a demographic outcome) of changing the phenotypic traits that influence demographic parameters.
- We may want to identify the parameters have the biggest effect on the outcome, in order to allocate sampling or measurement efforts where they are most needed.
- At a very basic level, we may simply want to know how the system works, how the outcomes are determined. Just as an empirical study might include experiments to manipulate factors and see how outcomes change, sensitivity analysis of a mathematical model reveals how outcomes respond to parameter changes.

It is not an overstatement to say that no model is every fully understood if it does not include a sensitivity analysis.

1.2   Sensitivity, Calculus, and Matrix Calculus

The change in an outcome in response to a change in a parameter can be treated as a problem in differential calculus. Let ξ denote some dependent variable and θ some parameter. The sensitivity problem can be approached via the derivative

$$\frac{d\xi}{d\theta} \tag{1.1}$$

or the elasticity, or proportional sensitivity[2]

$$\frac{\epsilon \xi}{\epsilon \theta} = \frac{\theta}{\xi} \frac{d\xi}{d\theta} = \frac{d \log \xi}{d \log \theta} \tag{1.2}$$

Note that I will use "sensitivity analysis" to refer generically to both sensitivity and elasticity.

The sensitivity problem is a challenging task, rather than an exercise in undergraduate calculus, because the dependence of ξ on θ may be complicated, and

---

[2]There seems to be no standard notation for elasticities; the one I am using here is based on a suggestion by Samuelson (1947).

because ξ may be a scalar (e.g. life expectancy at birth, or population growth rate) or a vector (e.g., a stable stage distribution or a projected population structure) or a matrix (e.g., the matrix of mean occupancy times). Similarly, θ may be a scalar (e.g., the Gompertz rate of aging) or a vector (e.g., the age schedule of mortality rates), or a matrix (e.g., the transition matrix among life cycle stages). In addition, the chains of causation in even simple demographic models are complicated. Tracing the causal chains from a set of parameters (of which there may be many) to a set of outcomes (again, many) with complicated interactions is hard.

This book is an in depth exploration of sensitivity analyses based on matrix formulations of demographic calculations. Matrix formulations are designed precisely to map transformations from one multidimensional space to another. Thus they simplify computations, clarify notation, and increase analytical power.[3]

The premise of this book is that demography as a discipline is neither defined by, nor limited to, a taxon. You will find here examples and analyses of humans, of non-human animals, and plants. Human demography and population biology have mutually informed each other from the beginning, and I see no reason for them to stop now.

It is important to remember that the diversity of complex life histories among the species that occupy our world poses a challenge to demographic analysis that is identical to the challenge posed by the complicated lives of humans. The dynamics of health status, family structure, or socio-economic status introduce complications to the life course exactly comparable to the dynamics of size growth in plants, metamorphosis in insects, or breeding status in birds.

A bit of history The earliest focus of demographic sensitivity analysis was population growth rate λ (or the intrinsic rate of increase $r = \log \lambda$) in linear demographic models. Hamilton (1966) was the first to solve this, in the context of the evolution of senescence. Demetrius (1969) derived a corresponding matrix expression, apparently unaware of Hamilton's results. Goodman (1971) was the first to notice the connection to reproductive value (see Chap. 3). Keyfitz (1971) derived the sensitivity of r, but also of life expectancy, mean age at death, and other outcomes.

All these analyses were based on age-classified demographic models. These results were generalized to stage-classified models by applying eigenvalue perturbation theory (Caswell 1978), followed by elasticity calculations (de Kroon et al. 1986), sensitivities of eigenvectors (Caswell 1982), lower-level parameters (Caswell 1989b), second derivatives of eigenvalues (Caswell 1996), the population spreading rate (Neubert and Caswell 2000), transient dynamics (Caswell 2007) and other things. Following the important early work of Tuljapurkar (1990), the sensitivity

---

[3]That does not mean that calculations made by other means are wrong. I am a methodological pluralist, and I do not believe that it is necessary to attack other methods in order to justify the use of matrix methods.

analysis of stochastic models developed in parallel with that of deterministic models (e.g. Tuljapurkar et al. 2003; Haridas and Tuljapurkar 2005; Horvitz et al. 2005; Steinsaltz et al. 2011).

Matrix calculus, permitting differentiation of scalar-, vector-, or matrix-valued functions of scalar, vector, or matrix arguments, began to be developed in the 1960s (see Nel (1980) for some history and comparison of different methods). The approach we will use here was introduced by Neudecker (1969) and expanded by Magnus and Neudecker (1985). A comprehensive, but mathematically difficult, treatment is given in Magnus and Neudecker (1988). Chapter 2 gives a brief presentation of the matrix calculus methods we will utilize in this book.

1.3   Some Issues

Sensitivity analysis is more than an algebraic exercise; it is a tool for making inferences and drawing conclusions about substantive demographic issues. It is useful to bring to the discussion a perspective on some questions.

1.3.1   Prospective and Retrospective Analyses: Sensitivity and Decomposition

If some variable ξ is a function of a set of parameters $\theta_1, \ldots, \theta_p$, then $\frac{\partial \boldsymbol{\xi}}{\partial \theta_i}$ gives the rate of change of ξ in response to a change in the ith parameter, holding the rest constant. Contrary to what is sometimes assumed, this calculation requires no assumption that it is actually possible to change the parameters. If the flight velocity of pigs is one of the parameters in the model, the analysis will happily answer the question of what would happen if pigs could fly.

Nor is there any assumption that changes in θ_i have ever happened in the past. The sensitivity analysis looks forward, asking what would happen *if* this or that parameter were to change. It is thus referred to as *prospective* analysis (Caswell 2000).

On the other hand, suppose you find yourself considering two values of ξ, that have resulted from two different situations (times, places, conditions), each with its own set of parameters:

$$\theta_1^{(1)}, \theta_2^{(1)}, \ldots \longrightarrow \xi^{(1)}$$
$$\theta_1^{(2)}, \theta_2^{(2)}, \ldots \longrightarrow \xi^{(2)}$$

You ask, what caused the difference between $\xi^{(2)}$ and $\xi^{(1)}$. Knowing the derivatives $\frac{\partial \xi}{\partial \theta_i}$ cannot tell you, because you are not asking the counterfactual question of what

would happen *if*, but the very factual question of what actually happened between the two situations. This is a *retrospective* analysis, familiar to human demographers as a decomposition problem (e.g., Kitagawa 1955; Canudas Romo 2003).

One widely used approach to understanding the causes of observed differences is life table response experiment (LTRE) analysis,[4] which uses a first-order approximation to decompose the differences,

$$\Delta\xi = \xi^{(2)} - \xi^{(1)} \approx \sum_i \frac{\partial\xi}{\partial\theta_i}\left(\theta_i^{(2)} - \theta_i^{(1)}\right). \tag{1.3}$$

The ith term in the summation is the contribution of the difference in the parameter θ_i to the difference in the outcome, $\Delta\xi$. These contributions reflect *both* the sensitivity of ξ to the parameters and the differences between conditions in each of the parameters. Parameters to which ξ is not very sensitive can make large contributions if the difference $\Delta\theta_i$ is big enough. Contributions to which ξ is very sensitive can make small contributions if θ_i does not change much. The matrix calculus version of this decomposition is given in Sect. 2.9, applied to differences in life disparity in Chap. 4, to periodic environments in Chap. 8, and explored in the challenging context of stochastic models in Chap. 9.

The distinction between prospective and retrospective analysis is obvious once the questions they address are specified, but it has challenged a number of authors (e.g., Wisdom and Mills 1997; Manlik et al. 2017). A particularly insightful discussion of these ideas, in somewhat different terminology, appears in Nathan Keyfitz's essay, *How do we know the facts of demography?*, which now appears as Chapter 20 of Keyfitz and Caswell (2005).

1.3.2   Uncertainty Propagation

Suppose that ξ is a function of θ, but θ is known only imperfectly. Then ξ is also known only imperfectly; the uncertainty in θ is propagated from θ to ξ. The sensitivity $d\xi/d\theta$ alone says nothing about uncertainty, and the uncertainty in ξ says nothing about the sensitivity.

Uncertainty propagation can be calculated by simulation if a probability distribution is known that can describe the uncertainty in θ. Sampling from this distribution and calculating ξ for each sampled parameter gives the distribution of ξ resulting from the uncertainty in θ (e.g. Caswell et al. 1998; Salomon et al.

---

[4]This awkward but well entrenched nomenclature was created when I was trying to understand the interpretation of experiments in ecotoxicology in which laboratory cohorts would be exposed to some noxious substance, and a life table (mortality and fertility schedule) measured as a response variable (Caswell 1989*a*). It soon became apparent that the method could be applied to any comparison of different conditions, and that the response could be any demographic variable. See (Caswell 2001, Chapter 10) for details.

2001). If the distribution of θ comes from an empirical set of measurements, this approach converges to the bootstrap (Efron and Tibshirani 1993). If θ has a parametric distribution (e.g., the multivariate normal distribution returned by maximum likelihood estimation) the technique is sometimes known as a parametric bootstrap (e.g., Regehr et al. 2010).

Sensitivity analysis can contribute to uncertainty propagation analysis through the first order, small variance approximation to the variance in ξ,

$$V(\xi) \approx \sum_{i,j} \left(\frac{\partial \xi}{\partial \theta_i} \right) \left(\frac{\partial \xi}{\partial \theta_j} \right) \mathrm{Cov}(\theta_i, \theta_j) \tag{1.4}$$

Notice again that sensitivity does not, by itself, say anything about uncertainty, but it does show how the (co)variance in parameters will propagate to the variance in the outcome ξ.

1.3.3 Why Not Just Simulate?

If you work on these problems, or if you apply these methods in particular studies, eventually you will be asked (often by a reviewer), why not just do it all by simulation?Just evaluate ξ at the value θ, and at $\theta + \Delta\theta$, and then approximate the derivative as

$$\frac{\Delta\xi}{\Delta\theta} = \frac{\xi(\theta + \Delta\theta) - \xi(\theta)}{\Delta\theta} \tag{1.5}$$

for some very small value of $\Delta\theta$.

Three answers come to mind. First, if θ and the model are of sufficiently high dimension, there can be a lot of these perturbations to be calculated. For example, population projections of the type analyzed by Caswell and Sanchez Gassen (2015), with 102 ages, 2 sexes, 3 vital rates, and projections on the order of 50 years, have over 30,000 parameters. A numerical perturbation of each of these would be painful.

Second, the computation of derivatives by numerical perturbations is a notoriously ill-behaved problem. A standard reference on computations in applied mathematics says that this approximation is "almost guaranteed to produce inaccurate results" (Press et al. 1992, p. 185). It is subject to truncation error (caused by making the perturbation too large) and roundoff error (caused by making the perturbation too small). In some applications these errors will be unimportant, but in others they can be crucial (e.g., Hunter and Caswell 2009, for an example in mark-recapture analysis)s.

Third, and more basic and telling: an exact answer is always an improvement over an approximation. When an exact answer is available, in an easily computable

form, there must be strong arguments to support the idea that a less efficient and less accurate approximation is just as good. And having both exact and approximate methods is even better.

These arguments apply to numerical calculation of derivatives. But simulation has an important place in analyzing *scenarios*; i.e., the results of specified collections of parameters, usually with multiple and large differences among them. When population projections are reported with "high," "medium," and "low" fertility scenarios, the point is to compare a range of multivariate alternatives. Other examples include comparisons of screening procedures for colorectal cancer (Wu et al. 2006), or projections based on IPCC global climate models (e.g., Jenouvrier et al. 2012). In principle, sensitivity analysis could support these calculations by suggesting interesting scenarios, highlighting the parameters with the biggest impact on the outcome.

1.3.4 Sensitivity and Identifying Targets for Intervention

To intervene is to change something. Population biologists concerned with endangered species would like to intervene to increase the population growth rate. Those concerned with invasive pests would like to do the same, but in the opposite direction. Human demographers focused on aging societies wonder about how policies would change age distributions or dependency ratios. In all these cases, the interventions operate through changes in demographic parameters, and thus sensitivity analysis can reveal something about their effects.

This logic has led to the use of prospective perturbation analyses in conservation biology, using the sensitivity or elasticity of population growth rate to identify promising targets for intervention. The first such use involved the loggerhead sea turtle (Crouse et al. 1987). Standard practice at the time was to focus on protecting eggs and hatchlings on nesting beaches. But a sensitivity analysis showed that population growth rate was not very sensitive to these stages, and much more sensitive to changes in survival of adults at sea. This led to a recommendation, and then a policy, to install "turtle excluder devices" on the nets used in coastal shrimp fisheries in the United States, to reduce mortality due to adult turtles being captured in those nets.

This basic idea has become a part of the toolkit for conservation biology, but has also fallen victim to a kind of magical thinking that first makes unrealistic expectations of the sensitivity analysis and then blames the analysis for failing to meet those expectations. For a recent example see Manlik et al. (2017); for a thorough description of the issues and some of their solutions, see Caswell (2001, Chapter 18).

The fact remains that knowing the sensitivity of some outcome ξ to some parameter θ gives the rate of change of ξ in response to an intervention that changes θ. That is valuable information to have in considering the various interventions that might bring about a desired change.

1.3.5   The Dream of Easy Interpretation

This book is full of long and complicated formulas. Occasionally, these formulas yield easy, readily apparent, qualitative interpretations.[5] But not often. There is a reason for this. The formulas are complicated because the processes are complicated, and because the results are given at a high level of generality. Chapter 10, for example, analyzes the sensitivity of nonlinear, density-dependent models. It derives a complicated formula for the sensitivity of *any* function of the equilibrium population, to changes in *any* parameter affecting *any* of the vital rates, in *any* age- or stage-specific way, for *any* choice of stage classification and *any* survival, fertility, and transition rates, with *any* pattern of density dependence, for *any* species with *any* kind of life history. Accounting for that web of dependencies, in such generality, makes finding an easily interpretable formula an unlikely dream.

Not an impossible dream, but in general, insights of that kind arise from simplifying general methods to address particular situations. Specifying a particular demographic structure, choosing an outcome variable of interest, and carefully specifying the functional dependencies, if done skillfully, can lead to qualitative results.

1.4   The Importance of Change

Questions of *change* lurk in almost every demographic (every scientific?) study. We ask how things have changed in the past, how they differ among populations in the present, and how they will, or may, change in the future. Even apparently simple descriptive statements (the results of a census in a particular time and place, for example) are almost immediately examined in comparison with other times and/or places.

Sensitivity analysis is a powerful tool for analyzing change, in the special case of demographic outcomes that are calculated as functions of some set of parameters. As the chapters to come will make clear, this covers a wide landscape of interesting demographic questions. And the list is not yet complete.

Bibliography

Canudas Romo, V. 2003. Decomposition methods in demography. Population Studies, Rozenberg Publishers, Amsterdam, Netherlands.
Caswell, H. 1978. A general formula for the sensitivity of population growth rate to changes in life history parameters. Theoretical Population Biology **14**:215–230.

---

[5]For example the interpretation of the sensitivity of population growth rate to matrix elements in terms of stable structure and reproductive value in Chap. 3, or the sensitivity of life expectancy to mortality in terms of occupancy time and transition probabilities in Chap. 4.

Caswell, H. 1982. Optimal life histories and the maximization of reproductive value: a general theorem for complex life cycles. Ecology **63**:1218–1222.

Caswell, H. 1989*a*. Analysis of life table response experiments I. Decomposition of effects on population growth rate. Ecological Modelling **46**:221–237.

Caswell, H. 1989*b*. Matrix Population Models: Construction, Analysis, and Interpretation. 1st edition. Sinauer Associates, Sunderland, MA, USA.

Caswell, H. 1996. Second derivatives of population growth rate: calculation and applications. Ecology **77**:870–879.

Caswell, H. 2000. Prospective and retrospective perturbation analyses and their use in conservation biology. Ecology **81**:619–627.

Caswell, H. 2001. Matrix Population Models: Construction, Analysis, and Interpretation. 2nd edition. Sinauer Associates, Sunderland, MA.

Caswell, H. 2007. Sensitivity analysis of transient population dynamics. Ecology Letters **10**:1–15.

Caswell, H., S. Brault, A. J. Read, and T. D. Smith. 1998. Harbor porpoise and fisheries: an uncertainty analysis of incidental mortality. Ecological Applications **8**:1226–1238.

Caswell, H., and N. Sanchez Gassen. 2015. The sensitivity analysis of population projections. Demographic Research **33**:801–840.

Crouse, D. T., L. B. Crowder, and H. Caswell. 1987. A stage-based population model for loggerhead sea turtles and implications for conservation. Ecology **68**:1412–1423.

de Kroon, H., A. Plaisier, J. van Groenendael, and H. Caswell. 1986. Elasticity: the relative contribution of demographic parameters to population growth rate. Ecology **67**:1427–1431.

de Vries, C., and H. Caswell. 2017. Demography when history matters: construction and analysis of second-order matrix population models. Theoretical Ecology https://doi.org/10.1007/s12080-017-0353-0:1–12.

Demetrius, L. 1969. The sensitivity of population growth rate to perturbations in the life cycle components. Mathematical Biosciences **4**:129–136.

Efron, B., and R. J. Tibshirani. 1993. An introduction to the bootstrap. Chapman and Hall, London, England.

Goodman, L. A. 1971. On the sensitivity of the intrinsic growth rate to changes in the age-specific birth and death rates. Theoretical Population Biology **2**:339–354.

Hamilton, W. D. 1966. The moulding of senescence by natural selection. Journal of Theoretical Biology **12**:12–45.

Haridas, C. V., and S. Tuljapurkar. 2005. Elasticities in variable environments: properties and implications. The American Naturalist **166**:481–495.

Horvitz, C. C., S. Tuljapurkar, and J. B. Pascarella. 2005. Plant-animal interactions in random environments: habitat-stage elasticity, seed predators, and hurricanes. Ecology **86**:3312–3322.

Hunter, C. M., and H. Caswell, 2009. Rank and redundancy of multistate mark-recapture models for seabird populations with unobservable states. Pages 797–825 *in* D. L. Thompson, E. G. Cooch, and M. J. Conroy, editors. Modeling demographic processes in marked populations, volume 3 of *Environmental and Ecological Statistics*. Springer, New York.

Jenouvrier, S., M. Holland, J. Stroeve, C. Barbraud, H. Weimerskirch, M. Serreze, and H. Caswell. 2012. Effects of climate change on an emperor penguin population: analysis of coupled demographic and climate models. Global Change Biology **18**:2756–2770.

Keyfitz, N. 1971. Linkages of intrinsic to age-specific rates. Journal of the American Statistical Association **66**:275–281.

Keyfitz, N., and H. Caswell. 2005. Applied mathematical demography. 3rd edition. Springer, New York, New York.

Kitagawa, E. M. 1955. Components of a difference between two rates. Journal of the American Statistical Association **50**:1168–1194.

Magnus, J. R., and H. Neudecker. 1985. Matrix differential calculus with applications to simple, Hadamard, and Kronecker products. Journal of Mathematical Psychology **29**:474–492.

Magnus, J. R., and H. Neudecker. 1988. Matrix differential calculus with applications in statistics and econometrics. John Wiley and Sons, New York, New York.

Manlik, O., R. C. Lacy, and W. B. Sherwin. 2017. Applicability and limitations of sensitivity analyses for wildlife management. Journal of Applied Ecology.

Metz, J. A. J., and O. Diekmann. 1986. The dynamics of physiologically structured populations. Springer-Verlag, Berlin, Germany.

Nel, D. G. 1980. On matrix differentiation in statistics. South African Statistical Journal **14**:137–193.

Neubert, M. G., and H. Caswell. 2000. Demography and dispersal: calculation and sensitivity analysis of invasion speed for structured populations. Ecology **81**:1613–1628.

Neudecker, H. 1969. Some theorems on matrix differentiation with special reference to Kronecker matrix products. Journal of the American Statistical Association **64**:953–963.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. (1992). *Numerical recipes in C: The art of scientific programming* (Second ed.). Cambridge, UK: Cambridge University Press.

Regehr, E. V., C. M. Hunter, H. Caswell, S. C. Amstrup, and I. Stirling. 2010. Survival and breeding of polar bears in the southern Beaufort Sea in relation to sea ice. Journal of Animal Ecology **79**:117–127.

Salomon, J. A., C. D. Mathers, C. J. L. Murray, and B. Ferguson, 2001. Methods for life expectancy and healthy life expectancy uncertainty analysis. Global Programme on Evidence for Health Policy Working Paper 10, World Health Organization, Geneva.

Samuelson, P. A. 1947. Foundations of economic analysis. Harvard University Press, Cambridge, Massachusetts, USA.

Steinsaltz, D., S. Tuljapurkar, and C. Horvitz. 2011. Derivatives of the stochastic growth rate. Theoretical Population Biology **80**:1–15.

Tuljapurkar, S. 1990. Population Dynamics in Variable Enviroinments. Springer-Verlag, New York, New York.

Tuljapurkar, S., C. Horvitz, and J. B. Pascarella. 2003. The many growth rates and elasticities of populations in random environments. American Naturalist **162**:489–502.

Wisdom, M. J., and L. S. Mills. 1997. Sensitivity analysis to guide population recovery: prairie-chickens as an example. The Journal of Wildlife Management pages 302–312.

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.

Chapter 2
Matrix Calculus and Notation

2.1 Introduction: Can It Possibly Be That Simple?

In October of 2005, I scribbled in a notebook, "can it possibly be *that* simple?" I was referring to the sensitivity of transient dynamics (the eventual results appear in Chap. 7), and had just begun to use matrix calculus as a tool. The answer to my question was yes. It can be that simple.

This book relies on this set of mathematical techniques. This chapter introduces the basics, which will be used throughout the text. For more information, I recommend four sources in particular. The most complete treatment, but not the easiest starting point, is the book by Magnus and Neudecker (1988). More accessible introductions can be found in the paper by Magnus and Neudecker (1985) and especially the text by Abadir and Magnus (2005). A review paper by Nel (1980) is helpful in placing the Magnus-Neudecker formulation in the context of other attempts at a calculus of matrices.

Sensitivity analysis asks how much change in an outcome variable y is caused by a change in some parameter x. At its most basic level, and with some reasonable assumptions about the continuity and differentiability of the functional relationships involved, the solution is given by differential calculus. If y is a function of x, then the derivative

$$\frac{dy}{dx}$$

tells how y responds to a change in x, i.e., the sensitivity of y to a change in x.

However, the outcomes of a demographic calculation may be scalar-valued (e.g., the population growth rate λ), vector-valued (e.g., the stable stage distribution), or matrix-valued (e.g., the fundamental matrix). Any of these outcomes may be

functions of scalar-valued parameters (e.g., the Gompertz aging rate), vector-valued parameters (e.g., the mortality schedule), or matrix-valued parameters (e.g., the transition matrix) parameters. Thus, sensitivity analysis in demography requires more than the simple derivative in (2.1). We want a consistent and flexible approach to differentiating

$$\left\{ \begin{array}{l} \text{scalar-valued} \\ \text{vector-valued} \\ \text{matrix-valued} \end{array} \right\} \text{ functions of } \left\{ \begin{array}{l} \text{scalar} \\ \text{vector} \\ \text{matrix} \end{array} \right\} \text{ arguments}$$

2.2   Notation and Matrix Operations

2.2.1   Notation

Matrices are denoted by upper case bold symbols (e.g., $\mathbf{A}$), vectors (usually) by lower case bold symbols ($\mathbf{n}$). The (i, j) entry of the matrix $\mathbf{A}$ is a_{ij}, and the ith entry of the vector $\mathbf{n}$ is n_i. Sometimes we will use MATLAB notation, and write

$$\mathbf{X}(i, :) = \text{row } i \text{ of } \mathbf{X} \tag{2.1}$$

$$\mathbf{X}(:, j) = \text{column } j \text{ of } \mathbf{X} \tag{2.2}$$

The notation

$$\big(x(ij) \big)$$

denotes a matrix whose (i, j) entry is x. For example,

$$\left(\frac{dy_i}{dx_j} \right)$$

is the matrix whose (i, j) entry is the derivative of y_i with respect to x_j.

The transpose of $\mathbf{X}$ is $\mathbf{X}^{\mathsf{T}}$. Logarithms are natural logarithms. The vector norm $\|\mathbf{x}\|$ is, unless noted otherwise, the 1-norm. The symbol $\mathcal{D}(\mathbf{x})$ denotes the square matrix with $\mathbf{x}$ on the diagonal and zeros elsewhere. The symbol $\mathbf{1}$ denotes a vector of ones. The vector $\mathbf{e}_i$ is a unit vector with 1 in the ith entry and zeros elsewhere. The identity matrix is $\mathbf{I}$. Where necessary for clarity, the dimension of matrices or vectors will be indicated by a subscript. Thus $\mathbf{I}_s$ is a $s \times s$ identity matrix, $\mathbf{1}_s$ is an $s \times 1$ vector of ones, and $\mathbf{X}_{m \times n}$ is a $m \times n$ matrix.

In some places (Chaps. 6 and 10) block-structured matrices appear; these are denoted by either $\mathbb{A}$ or $\tilde{\mathbf{A}}$, depending on the context and the role of the matrix.

2.2.2   Operations

In addition to the familiar matrix product $\mathbf{AB}$, we will also use the Hadamard, or elementwise product

$$\mathbf{A} \circ \mathbf{B} = \left(a_{ij} b_{ij} \right) \tag{2.3}$$

and the Kronecker product

$$\mathbf{A} \otimes \mathbf{B} = \left(a_{ij} \mathbf{B} \right) \tag{2.4}$$

The Hadamard product requires that $\mathbf{A}$ and $\mathbf{B}$ be the same size. The Kronecker product is defined for any sizes of $\mathbf{A}$ and $\mathbf{B}$. Some useful properties of the Kronecker product include

$$(\mathbf{A} \otimes \mathbf{B})^{-1} = \left(\mathbf{A}^{-1} \otimes \mathbf{B}^{-1} \right) \tag{2.5}$$

$$(\mathbf{A} \otimes \mathbf{B})^{\mathsf{T}} = \left(\mathbf{A}^{\mathsf{T}} \otimes \mathbf{B}^{\mathsf{T}} \right) \tag{2.6}$$

$$\mathbf{A} \otimes (\mathbf{B} + \mathbf{C}) = (\mathbf{A} \otimes \mathbf{B}) + (\mathbf{A} \otimes \mathbf{C}) \tag{2.7}$$

and, provided that the matrices are of the right size for the products to be defined,

$$(\mathbf{A}_1 \otimes \mathbf{B}_1)(\mathbf{A}_2 \otimes \mathbf{B}_2) = (\mathbf{A}_1 \mathbf{A}_2 \otimes \mathbf{B}_1 \mathbf{B}_2). \tag{2.8}$$

2.2.3   The Vec Operator and Vec-Permutation Matrix

The vec operator transforms a $m \times n$ matrix $\mathbf{A}$ into a $mn \times 1$ vector by stacking the columns one above the next,

$$\text{vec}\,\mathbf{A} = \begin{pmatrix} \mathbf{A}(:, 1) \\ \vdots \\ \mathbf{A}(:, n) \end{pmatrix} \tag{2.9}$$

For example,

$$\text{vec}\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a \\ c \\ b \\ d \end{pmatrix}. \tag{2.10}$$

The vec of $\mathbf{A}$ and the vec of $\mathbf{A}^\mathsf{T}$ are rearrangements of the same entries; they are related by

$$\text{vec}\,\mathbf{A}^\mathsf{T} = \mathbf{K}_{m,n}\,\text{vec}\,\mathbf{A} \qquad (2.11)$$

where $\mathbf{A}$ is $m \times n$ and $\mathbf{K}_{m,n}$ is the *vec-permutation matrix* (Henderson and Searle 1981) or *commutation matrix* (Magnus and Neudecker 1979). The vec-permutation matrix can be calculated as

$$\mathbf{K}_{m,n} = \sum_{i=1}^{m} \sum_{j=1}^{n} \left(\mathbf{E}_{ij} \otimes \mathbf{E}_{ij}^\mathsf{T} \right) \qquad (2.12)$$

where $\mathbf{E}_{ij}$ is a matrix, of dimension $m \times n$, with a 1 in the (i, j) entry and zeros elsewhere. Like any permutation matrix, $\mathbf{K}^{-1} = \mathbf{K}^\mathsf{T}$.

The vec operator and the vec-permutation matrix are particularly important in multistate models (e.g., age$\times$stage-classified models), where they are used in both the formulation and analysis of the models (e.g., Caswell 2012, 2014; Caswell and Salguero-Gómez 2013; Caswell et al. 2018); see also Chap. 6. Extensions to an arbitrary number of dimensions, so-called hyperstate models, have been presented by Roth and Caswell (2016).

2.2.4   Roth's Theorem

The vec operator and the Kronecker product are connected by a theorem due to Roth (1934):

$$\text{vec}\,(\mathbf{ABC}) = \left(\mathbf{C}^\mathsf{T} \otimes \mathbf{A} \right) \text{vec}\,\mathbf{B}. \qquad (2.13)$$

We will often want to obtain the vec of a matrix that appears in the middle of a product; we will use Roth's theorem repeatedly.

2.3   Defining Matrix Derivatives

The derivative of a scalar y with respect to a scalar x is familiar. What, however, does it mean to speak of the derivative of a scalar with respect to a vector, or of a vector with respect to another vector, or any other combination? These can be defined in more than one way and the choice is critical (Nel 1980; Magnus and Neudecker 1985). This book relies on the notation due to Magnus and Neudecker, because it makes certain operations possible and consistent.

- If x and y are scalars, the derivative of y with respect to x is the familiar derivative dy/dx.

- If $\mathbf{y}$ is a $n \times 1$ vector and x a scalar, the derivative of $\mathbf{y}$ with respect to x is the $n \times 1$ column vector

$$\frac{d\mathbf{y}}{dx} = \begin{pmatrix} \dfrac{dy_1}{dx} \\ \vdots \\ \dfrac{dy_n}{dx} \end{pmatrix}. \tag{2.14}$$

- If y is a scalar and $\mathbf{x}$ a $m \times 1$ vector, the derivative of y with respect to $\mathbf{x}$ is the $1 \times m$ row vector (called the gradient vector)

$$\frac{dy}{d\mathbf{x}^{\mathsf{T}}} = \begin{pmatrix} \dfrac{\partial y}{\partial x_1} & \cdots & \dfrac{\partial y}{\partial x_m} \end{pmatrix}. \tag{2.15}$$

Note the orientation of $d\mathbf{y}/dx$ as a column vector and $dy/d\mathbf{x}^{\mathsf{T}}$ as a row vector.
- If $\mathbf{y}$ is a $n \times 1$ vector and $\mathbf{x}$ a $m \times 1$ vector, the derivative of $\mathbf{y}$ with respect to $\mathbf{x}$ is defined to be the $n \times m$ matrix whose (i, j) entry is the derivative of y_i with respect to x_j, i.e.,

$$\frac{d\mathbf{y}}{d\mathbf{x}^{\mathsf{T}}} = \begin{pmatrix} \dfrac{dy_i}{dx_j} \end{pmatrix} \tag{2.16}$$

(this matrix is called the Jacobian matrix).
- Derivatives involving matrices are written by first transforming the matrices into vectors using the vec operator, and then applying the rules for vector differentiation to the resulting vectors. Thus, the derivative of the $m \times n$ matrix $\mathbf{Y}$ with respect to the $p \times q$ matrix $\mathbf{X}$ is the $mn \times pq$ matrix

$$\frac{d\,\mathrm{vec}\,\mathbf{Y}}{d\,(\mathrm{vec}\,\mathbf{X})^{\mathsf{T}}}. \tag{2.17}$$

From now on, I will write $\mathrm{vec}^{\mathsf{T}}\mathbf{X}$ for $(\mathrm{vec}\,\mathbf{X})^{\mathsf{T}}$.

2.4 The Chain Rule

The chain rule for differentiation is your friend. The Magnus-Neudecker notation, unlike some alternatives, extends the familiar scalar chain rule to derivatives of vectors and matrices (Nel 1980; Magnus and Neudecker 1985). If $\mathbf{u}$ (size $m \times 1$) is a function of $\mathbf{v}$ (size $n \times 1$) and $\mathbf{v}$ is a function of $\mathbf{x}$ (size $p \times 1$), then

$$\underbrace{\frac{d\mathbf{u}}{d\mathbf{x}^{\mathsf{T}}}}_{m \times p} = \underbrace{\left(\frac{d\mathbf{u}}{d\mathbf{v}^{\mathsf{T}}} \right)}_{m \times n} \underbrace{\left(\frac{d\mathbf{v}}{d\mathbf{x}^{\mathsf{T}}} \right)}_{n \times p} \tag{2.18}$$

Notice that the dimensions are correct, and that the order of the multiplication matters. Checking dimensional consistency in this way is a useful way to find errors.

2.5   Derivatives from Differentials

The key to the matrix calculus of Magnus and Neudecker (1988) is the relationship between the differential and the derivative of a function. Experience suggests that, for many readers of this book, this relationship is shrouded in the mists of long-ago calculus classes.

2.5.1   Differentials of Scalar Function

Start with scalars. Suppose that $y = f(x)$ is a differentiable function at $x = x_0$. Then the derivative of y with respect to x at the point x_0 is defined as

$$f'(x_0) = \lim_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{h}. \tag{2.19}$$

Now define the differential of y. This is customarily denoted dy, but for the moment, I will denote it by cy. The differential of y at x_0 is a function of h, defined by

$$cy(x_0, h) = f'(x_0)h. \tag{2.20}$$

There is no requirement that h be "small." Since x is a function of itself, $x = g(x)$, with $g'(x) = 1$, we also have $cx(x_0, h) = g'(x_0)h = h$. Thus the ratio of the differential of y and the differential of x is

$$\frac{cy(x_0, h)}{cx(x_0, h)} = \frac{f'(x_0)h}{h} = f'(x_0). \tag{2.21}$$

That is, *the derivative is equal to the ratio of the differentials.*

Now, return to the standard notation of dy for the differential of y. This gives two meanings to the familiar notation for derivatives,

$$\left. \frac{dy}{dx} \right|_{x_0} = f'(x_0). \tag{2.22}$$

The left hand side can be regarded either as equivalent to the limit (2.19) or the ratio of the differentials given by (2.21). Mathematicians are strangely unconcerned with this ambiguity (e.g., Hardy 1952).

All this leads to a set of familiar rules for calculating differentials that guarantee that they can be used to create derivatives. A few of these, for scalars, are

$$d(u + v) = du + dv \qquad (2.23)$$

$$d(cu) = c\,du \qquad (2.24)$$

$$d(uv) = u(dv) + (du)v \qquad (2.25)$$

$$d(e^u) = e^u du \qquad (2.26)$$

$$d(\log u) = \frac{1}{u}du \qquad (2.27)$$

If $y = f(x_1, x_2)$, then the total differential is

$$dy = \frac{\partial f}{\partial x_1}dx_1 + \frac{\partial f}{\partial x_2}dx_2. \qquad (2.28)$$

Derivatives can be constructed from these expressions at will by dividing by differentials. For example, dividing (2.23) by dx gives $d(u + v)/dx = du/dx + dv/dx$. From (2.28), we have

$$\frac{dy}{dx_1} = \frac{\partial f}{\partial x_1} + \frac{\partial f}{\partial x_2}\frac{dx_2}{dx_1} \qquad (2.29)$$

$$\frac{dy}{dx_2} = \frac{\partial f}{\partial x_1}\frac{dx_1}{dx_2} + \frac{\partial f}{\partial x_2}. \qquad (2.30)$$

2.5.2   Differentials of Vectors and Matrices

To extend these concepts to matrices, we define the differential of a matrix (or vector) as the matrix (or vector) of differentials of the elements; i.e.,

$$d\mathbf{X} = \left(dx_{ij} \right). \qquad (2.31)$$

This definition leads to some basic rules for differentials of matrices:

$$d(c\mathbf{U}) = c(d\mathbf{U}) \qquad (2.32)$$

$$d(\mathbf{U} + \mathbf{V}) = d\mathbf{U} + d\mathbf{V} \qquad (2.33)$$

$$d(\mathbf{UV}) = (d\mathbf{U})\mathbf{V} + \mathbf{U}(d\mathbf{V}) \qquad (2.34)$$

$$d(\mathbf{U} \otimes \mathbf{V}) = (d\mathbf{U}) \otimes \mathbf{V} + \mathbf{U} \otimes (d\mathbf{V}) \qquad (2.35)$$

$$d(\mathbf{U} \circ \mathbf{V}) = (d\mathbf{U}) \circ \mathbf{V} + \mathbf{U} \circ (d\mathbf{V}) \tag{2.36}$$

$$d\operatorname{vec}\mathbf{U} = \operatorname{vec} d\mathbf{U} \tag{2.37}$$

where c is a constant, and, of course, the dimensions of $\mathbf{U}$ and $\mathbf{V}$ must be conformable. The differentials of an operators applied elementwise to a vector can be obtained from the differentials of the elements. For example, suppose $\mathbf{u}$ is a $s \times 1$ vector, and the exponential is applied elementwise. Then

$$d(\exp(\mathbf{u})) = \begin{pmatrix} e^{u_1} du_1 \\ \vdots \\ e^{u_s} du_s \end{pmatrix} \tag{2.38}$$

$$= \mathcal{D}\left[\exp(\mathbf{u})\right] d\mathbf{u}. \tag{2.39}$$

If $\mathbf{y}$ is a function of $\mathbf{x}_1$ and $\mathbf{x}_2$, the total differential is given just as in (2.28), by

$$d\mathbf{y} = \frac{\partial \mathbf{y}}{\partial \mathbf{x}_1^{\mathsf{T}}} d\mathbf{x}_1 + \frac{\partial \mathbf{y}}{\partial \mathbf{x}_2^{\mathsf{T}}} d\mathbf{x}_2 \tag{2.40}$$

2.6   The First Identification Theorem

For scalar y and x,

$$dy = qdx \implies \frac{dy}{dx} = q. \tag{2.41}$$

That much is easy. But, suppose that $\mathbf{y}$ is a $n \times 1$ vector function of the $m \times 1$ vector $\mathbf{x}$. The differential $d\mathbf{y}$ is the $n \times 1$ vector

$$d\mathbf{y} = \begin{pmatrix} dy_1 \\ \vdots \\ dy_n \end{pmatrix} \tag{2.42}$$

which, by the total derivative rule, is

$$d\mathbf{y} = \begin{pmatrix} \dfrac{\partial y_1}{\partial x_1} dx_1 + \cdots + \dfrac{\partial y_1}{\partial x_m} dx_m \\ \vdots \\ \dfrac{\partial y_n}{\partial x_1} dx_1 + \cdots + \dfrac{\partial y_n}{\partial x_m} dx_m \end{pmatrix} \tag{2.43}$$

$$= \begin{pmatrix} \frac{\partial y_1}{\partial x_1} & \cdots & \frac{\partial y_1}{\partial x_m} \\ \vdots & & \vdots \\ \frac{\partial y_n}{\partial x_1} & \cdots & \frac{\partial y_n}{\partial x_m} \end{pmatrix} \begin{pmatrix} dx_1 \\ \vdots \\ dx_m \end{pmatrix} \tag{2.44}$$

$$= \mathbf{Q}\, d\mathbf{x}. \tag{2.45}$$

If these were scalars, dividing both sides by $d\mathbf{x}$ would give $\mathbf{Q}$ as the derivative of $\mathbf{y}$ with respect to $\mathbf{x}$. But, one cannot divide by a vector. Instead, Magnus and Neudecker proved that if it can be shown that

$$d\mathbf{y} = \mathbf{Q}\, d\mathbf{x} \tag{2.46}$$

then the derivative is

$$\frac{d\mathbf{y}}{d\mathbf{x}^\mathsf{T}} = \mathbf{Q}. \tag{2.47}$$

This is the First Identification Theorem of Magnus and Neudecker (1988).[1]

2.6.1   The Chain Rule and the First Identification Theorem

Suppose that $d\mathbf{y}$ is given by (2.46), and that $\mathbf{x}$ is in turn a function of some vector $\boldsymbol{\theta}$. Then

$$d\mathbf{x} = \frac{d\mathbf{x}}{d\boldsymbol{\theta}^\mathsf{T}} d\boldsymbol{\theta} \tag{2.48}$$

and

$$\frac{d\mathbf{y}}{d\boldsymbol{\theta}^\mathsf{T}} = \mathbf{Q}\frac{d\mathbf{x}}{d\boldsymbol{\theta}^\mathsf{T}}. \tag{2.49}$$

In other words, the differential expression (2.46) can be transformed into a derivative with respect to any vector by careful use of the chain rule. This applies equally to more complicated expressions for the differential. Suppose that

$$d\mathbf{y} = \mathbf{Q}d\mathbf{x} + \mathbf{R}d\mathbf{z}. \tag{2.50}$$

---

[1]There is also a second identification theorem that provides the second derivatives of matrix functions. See Shyu and Caswell (2014) for applications of this theory to the second derivatives of measures of population growth rate.

Applying the chain rule to the differentials on the right hand side gives

$$dy = Q\frac{dx}{d\theta^{\mathsf{T}}}d\theta + R\frac{dz}{d\theta^{\mathsf{T}}}d\theta \tag{2.51}$$

for any vector θ. Thus

$$dy = \left(Q\frac{dx}{d\theta^{\mathsf{T}}} + R\frac{dz}{d\theta^{\mathsf{T}}}\right)d\theta, \tag{2.52}$$

and the First Identification Theorem gives

$$\frac{dy}{d\theta^{\mathsf{T}}} = \left(Q\frac{dx}{d\theta^{\mathsf{T}}} + R\frac{dz}{d\theta^{\mathsf{T}}}\right). \tag{2.53}$$

2.7 Elasticity

When parameters are measured on different scales, it is sometimes helpful to calculate proportional effects of proportional perturbations, also called elasticities. The elasticity of y_i to θ_j is

$$\frac{\epsilon y_i}{\epsilon \theta_j} = \frac{\theta_j}{y_i}\frac{dy_i}{d\theta_j}. \tag{2.54}$$

For vectors y and θ, this becomes

$$\frac{\epsilon y}{\epsilon \theta^{\mathsf{T}}} = \mathcal{D}(y)^{-1}\frac{dy}{d\theta^{\mathsf{T}}}\mathcal{D}(\theta). \tag{2.55}$$

There seems to be no accepted notation for elasticities; the notation used here is adapted from that in Samuelson (1947).

2.8 Some Useful Matrix Calculus Results

Several matrix calculus results will be used repeatedly. Many more can be found in Magnus and Neudecker (1988) and Abadir and Magnus (2005).

1. The matrix product $Y = AB$. Differentiate,

$$dY = (dA)B + A(dB). \tag{2.56}$$

Then write (or imagine writing; with practice one does not actually need this step explicitly)

$$(d\mathbf{A})\,\mathbf{B} = \mathbf{I}\,(d\mathbf{A})\,\mathbf{B} \tag{2.57}$$

$$\mathbf{A}\,(d\mathbf{B}) = \mathbf{A}\,(d\mathbf{B})\,\mathbf{I} \tag{2.58}$$

and apply the vec operator and Roth's theorem, to obtain

$$d\text{vec}\,\mathbf{Y} = \left(\mathbf{B}^{\mathsf{T}} \otimes \mathbf{I}\right) d\text{vec}\,\mathbf{A} + (\mathbf{I} \otimes \mathbf{A})\,d\text{vec}\,\mathbf{B}. \tag{2.59}$$

The chain rule gives, for any vector variable $\boldsymbol{\theta}$

$$\frac{d\text{vec}\,\mathbf{Y}}{d\boldsymbol{\theta}^{\mathsf{T}}} = \left(\mathbf{B}^{\mathsf{T}} \otimes \mathbf{I}\right)\frac{d\text{vec}\,\mathbf{A}}{d\boldsymbol{\theta}^{\mathsf{T}}} + (\mathbf{I} \otimes \mathbf{A})\frac{d\text{vec}\,\mathbf{B}}{d\boldsymbol{\theta}^{\mathsf{T}}}. \tag{2.60}$$

2. The Hadamard product $\mathbf{Y} = \mathbf{A} \circ \mathbf{B}$. Differentiate the product,

$$d\mathbf{Y} = d\mathbf{A} \circ \mathbf{B} + \mathbf{A} \circ d\mathbf{B}, \tag{2.61}$$

then vec

$$d\text{vec}\,\mathbf{Y} = d\text{vec}\,\mathbf{A} \circ \text{vec}\,\mathbf{B} + \text{vec}\,\mathbf{A} \circ d\text{vec}\,\mathbf{B}. \tag{2.62}$$

It will be useful to replace the Hadamard products, which we do using the fact that $\mathbf{x} \circ \mathbf{y} = \mathcal{D}\,(\mathbf{x})\mathbf{y}$, to get

$$d\text{vec}\,\mathbf{Y} = \mathcal{D}\,(\text{vec}\,\mathbf{B})d\text{vec}\,\mathbf{A} + \mathcal{D}\,(\text{vec}\,\mathbf{A})d\text{vec}\,\mathbf{B}. \tag{2.63}$$

The chain rule gives the derivative from the differential,

$$\frac{d\text{vec}\,\mathbf{Y}}{d\boldsymbol{\theta}^{\mathsf{T}}} = \mathcal{D}\,(\text{vec}\,\mathbf{B})\frac{d\text{vec}\,\mathbf{A}}{d\boldsymbol{\theta}^{\mathsf{T}}} + \mathcal{D}\,(\text{vec}\,\mathbf{A})\frac{d\text{vec}\,\mathbf{B}}{d\boldsymbol{\theta}^{\mathsf{T}}}. \tag{2.64}$$

3. Diagonal matrices. The diagonal matrix $\mathcal{D}\,(\mathbf{x})$, with the vector $\mathbf{x}$ on the diagonal and zeros elsewhere, can be written

$$\mathcal{D}\,(\mathbf{x}) = \mathbf{I} \circ \left(\mathbf{1}\,\mathbf{x}^{\mathsf{T}}\right) \tag{2.65}$$

Differentiate both sides,

$$d\mathcal{D}\,(\mathbf{x}) = \mathbf{I} \circ \left(\mathbf{1}\,d\mathbf{x}^{\mathsf{T}}\right) \tag{2.66}$$

and vec the result

$$dvec \, \mathcal{D} (\mathbf{x}) = \mathcal{D} (vec \, \mathbf{I}) vec \left(\mathbf{1} \, d\mathbf{x}^\mathsf{T} \right) \tag{2.67}$$

$$= \mathcal{D} (vec \, \mathbf{I}) (\mathbf{I} \otimes \mathbf{1}) \, d\mathbf{x} \tag{2.68}$$

The First Identification Theorem gives

$$\frac{dvec \, \mathcal{D} (\mathbf{x})}{d\boldsymbol{\theta}^\mathsf{T}} = \mathcal{D} (vec \, \mathbf{I}) (\mathbf{I} \otimes \mathbf{1}) \frac{d\mathbf{x}}{d\boldsymbol{\theta}^\mathsf{T}}. \tag{2.69}$$

The identity matrix in (2.65) *masks* the matrix $\left(\mathbf{1} \, \mathbf{x}^\mathsf{T} \right)$, setting to zero all but the diagonal elements. Matrices other than $\mathbf{I}$ can be used in this way to mask entries of a matrix. For example, the transition matrix for a Leslie matrix, with a vector of survival probabilities $\mathbf{p}$ on the subdiagonal, is obtained by setting $\mathbf{x} = \mathbf{p}$ and replacing $\mathbf{I}$ with a matrix $\mathbf{Z}$ that contains ones on the subdiagonal and zeros elsewhere (see, e.g., Chap. 4).

Some Markov chain calculations (Chaps. 5 and 11) involve a matrix $\mathbf{N}_{dg}$, which contains the diagonal elements of $\mathbf{N}$ on the diagonal and zeros elsewhere. This can be written

$$\mathbf{N}_{dg} = \mathbf{I} \circ \mathbf{N}. \tag{2.70}$$

Differentiating and applying the vec operator yields

$$dvec \, \mathbf{N}_{dg} = \mathcal{D} (vec \, \mathbf{I}) dvec \, \mathbf{N}. \tag{2.71}$$

4. The Kronecker product. Differentiating the Kronecker product is a bit more complicated (Magnus and Neudecker 1985, Theorem 11). We want an expression for the differential of the product in terms of the differentials of the components, something of the form

$$dvec \, (\mathbf{A} \otimes \mathbf{B}) = \mathbf{Z}_1 dvec \, \mathbf{A} + \mathbf{Z}_2 dvec \, \mathbf{B} \tag{2.72}$$

for some matrices $\mathbf{Z}_1$ and $\mathbf{Z}_2$.

This requires a result for the vec of the Kronecker product. Let $\mathbf{A}$ be of dimension $m \times p$ and $\mathbf{B}$ be $r \times s$. Then

$$vec \, (\mathbf{A} \otimes \mathbf{B}) = \left(\mathbf{I}_p \otimes \mathbf{K}_{s,m} \otimes \mathbf{I}_r \right) (vec \, \mathbf{A} \otimes vec \, \mathbf{B}) . \tag{2.73}$$

Let $\mathbf{Y} = \mathbf{A} \otimes \mathbf{B}$. Differentiate,

$$d\mathbf{Y} = (d\mathbf{A} \otimes \mathbf{B}) + (\mathbf{A} \otimes d\mathbf{B}) \tag{2.74}$$

and vec

$$d\text{vec}\,\mathbf{Y} = \left(\mathbf{I}_p \otimes \mathbf{K}_{s,m} \otimes \mathbf{I}_r\right) \left[(d\text{vec}\,\mathbf{A} \otimes \text{vec}\,\mathbf{B}) + (\text{vec}\,\mathbf{A} \otimes d\text{vec}\,\mathbf{B}) \right].$$
(2.75)

With some ingenious simplifications (Magnus and Neudecker 1985), this reduces to (2.72) with

$$\mathbf{Z}_1 = \left(\mathbf{I}_p \otimes \mathbf{K}_{s,m} \otimes \mathbf{I}_r\right) \left(\mathbf{I}_m \otimes \text{vec}\,\mathbf{B}\right)$$
(2.76)

$$\mathbf{Z}_2 = \left(\mathbf{I}_p \otimes \mathbf{K}_{s,m} \otimes \mathbf{I}_r\right) \left(\text{vec}\,\mathbf{A} \otimes \mathbf{I}_{rs}\right).$$
(2.77)

Substituting $\mathbf{Z}_1$ and $\mathbf{Z}_2$ into (2.72) gives the differential of the Kronecker product in terms of the differentials of its component matrices.

5. The matrix inverse. The inverse of $\mathbf{X}$ satisfies

$$\mathbf{X}\mathbf{X}^{-1} = \mathbf{I}.$$
(2.78)

Differentiate both sides

$$(d\mathbf{X})\,\mathbf{X}^{-1} + \mathbf{X}\left(d\mathbf{X}^{-1}\right) = \mathbf{0},$$
(2.79)

then vec

$$\left[\left(\mathbf{X}^{-1}\right)^{\mathsf{T}} \otimes \mathbf{I}\right] d\text{vec}\,\mathbf{X} + [\mathbf{I} \otimes \mathbf{X}]\,d\text{vec}\,\mathbf{X}^{-1} = \mathbf{0}$$
(2.80)

and finally solve for $d\text{vec}\,\mathbf{X}^{-1}$

$$d\text{vec}\,\mathbf{X}^{-1} = -[\mathbf{I} \otimes \mathbf{X}]^{-1}\left[\left(\mathbf{X}^{-1}\right)^{\mathsf{T}} \otimes \mathbf{I}\right] d\text{vec}\,\mathbf{X}$$
(2.81)

The properties (2.5) and (2.8) of the Kronecker product let this be simplified to

$$d\text{vec}\,\mathbf{X}^{-1} = -\left[\left(\mathbf{X}^{-1}\right)^{\mathsf{T}} \otimes \mathbf{X}^{-1}\right] d\text{vec}\,\mathbf{X}$$
(2.82)

6. The square root and ratios. In calculating standard deviations and coefficients of variation it is useful to calculate the elementwise square root and the elementwise ratio of two vectors. If $\mathbf{x}$ is a non-negative vector, and the square root $\sqrt{\mathbf{x}}$ is taken elementwise, then

$$d\sqrt{\mathbf{x}} = \frac{1}{2}\mathcal{D}\left(\sqrt{\mathbf{x}}\right)^{-1} d\mathbf{x}.$$
(2.83)

For the elementwise ratio, let $\mathbf{x}$ and $\mathbf{y}$ be $m \times 1$ vectors, with $\mathbf{y}$ nonzero. Let $\mathbf{w}$ be a vector whose ith element is x_i/y_i; i.e., $\mathbf{w} = \mathcal{D}(\mathbf{y})^{-1}\mathbf{x}$. Then

$$d\mathbf{w} = \mathcal{D}(\mathbf{y})^{-1}d\mathbf{x} - \left[\mathbf{x}^\mathsf{T}\mathcal{D}(\mathbf{y})^{-1} \otimes \mathcal{D}(\mathbf{y})^{-1}\right]\mathcal{D}(\operatorname{vec}\mathbf{I}_m)(\mathbf{I}_m \otimes \mathbf{1}_m)\,d\mathbf{y}. \tag{2.84}$$

This list could go on. The books by Magnus and Neudecker (1988) and Abadir and Magnus (2005) contain many other results, and demographically relevant derivations appear throughout this book, especially in Chap. 5.

2.9   LTRE Decomposition of Demographic Differences

The LTRE decomposition in Sect. 1.3.1 extends readily to matrix calculus. Suppose that a demographic outcome $\boldsymbol{\xi}$, dimension $(s \times 1)$, is a function of a vector $\boldsymbol{\theta}$ of parameters, dimension $(p \times 1)$. Suppose that results are obtained under two "conditions," with parameters $\boldsymbol{\theta}^{(1)}$ and $\boldsymbol{\theta}^{(2)}$. Define the parameter difference as $\Delta\boldsymbol{\theta} = \boldsymbol{\theta}^{(2)} - \boldsymbol{\theta}^{(1)}$ and the effect as $\Delta\boldsymbol{\xi} = \boldsymbol{\xi}^{(2)} - \boldsymbol{\xi}^{(1)}$. Then, to first order,

$$\Delta\boldsymbol{\xi} \approx \sum_{i=1}^{p} \frac{d\boldsymbol{\xi}}{d\theta_i}\Delta\theta_i \tag{2.85}$$

$$= \frac{d\boldsymbol{\xi}}{d\boldsymbol{\theta}^\mathsf{T}}\Delta\boldsymbol{\theta}. \tag{2.86}$$

Writing

$$\Delta\boldsymbol{\theta} = \mathcal{D}(\Delta\boldsymbol{\theta})\mathbf{1}_p, \tag{2.87}$$

we create a *contribution matrix* $\mathbf{C}$, of dimension $s \times p$,

$$\mathbf{C} = \frac{d\boldsymbol{\xi}}{d\boldsymbol{\theta}^\mathsf{T}}\,\mathcal{D}(\Delta\boldsymbol{\theta}). \tag{2.88}$$

The (i, j) entry of $\mathbf{C}$ is the contribution of $\Delta\theta_j$ to the difference ξ_i, for $i = 1, \ldots, s$ and $j = 1, \ldots, p$. The rows and columns of $\mathbf{C}$ give

$$\mathbf{C}(i, :) = \text{contributions of } \Delta\boldsymbol{\theta} \text{ to } \Delta\xi_i \tag{2.89}$$

$$\mathbf{C}(:, j) = \text{contributions of } \theta_j \text{ to } \Delta\boldsymbol{\xi} \tag{2.90}$$

When calculating $\mathbf{C}$, the derivative of $\boldsymbol{\xi}$ must be evaluated somewhere. Experience suggests that the evaluating it at the midpoint between $\boldsymbol{\theta}^{(1)}$ and $\boldsymbol{\theta}^{(2)}$ gives good results (Logofet and Lesnaya 1997; Caswell 2001).

2.10   A Protocol for Sensitivity Analysis

The calculations may grow to be complex, but the protocol is simple:

1. write a matrix expression for the outcome,
2. differentiate,
3. vec,
4. simplify,
5. calculate derivatives from the differentials, and
6. extend using the chain rule

The rest of this book shows what can be done with this simple procedure.

Bibliography

Abadir, K. M., and J. R. Magnus. 2005. Matrix algebra. Econometric exercises 1. Cambridge University Press, Cambridge, United Kingdom.

Caswell, H. 2001. Matrix Population Models: Construction, Analysis, and Interpretation. 2nd edition. Sinauer Associates, Sunderland, MA.

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. 2014. A matrix approach to the statistics of longevity in heterogeneous frailty models. Demographic Research **31**:553–592.

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.

Hardy, G. H. 1952. A course of pure mathematics. 10th edition. Cambridge University Press, Cambridge, United Kingdom.

Henderson, H. V., and S. R. Searle. 1981. The vec-permutation matrix, the vec operator and Kronecker products: a review. Linear and Multilinear Algebra **9**:271–288.

Logofet, D. O., and E. V. Lesnaya. 1997. Why are the middle points the most sensitive in the sensitivity experiments? Ecological Modelling **104**:303–306.

Magnus, J. R., and H. Neudecker. 1979. The commutation matrix: some properties and applications. Annals of Statistics **7**:381–394.

Magnus, J. R., and H. Neudecker. 1985. Matrix differential calculus with applications to simple, Hadamard, and Kronecker products. Journal of Mathematical Psychology **29**:474–492.

Magnus, J. R., and H. Neudecker. 1988. Matrix differential calculus with applications in statistics and econometrics. John Wiley and Sons, New York, New York.

Nel, D. G. 1980. On matrix differentiation in statistics. South African Statistical Journal **14**:137–193.

Roth, G., and H. Caswell. 2016. Hyperstate matrix models: extending demographic state spaces to higher dimensions. Methods in Ecology and Evolution **7**:1438–1450.

Roth, W. E. 1934. On direct product matrices. Bulletin of the American Mathematical Society **40**:461–468.

Samuelson, P. A. 1947. Foundations of economic analysis. Harvard University Press, Cambridge, Massachusetts, USA.

Shyu, E., and H. Caswell. 2014. Calculating second derivatives of population growth rates for ecology and evolution. Methods in Ecology and Evolution **5**:473–482.

Part II
Linear Models

Chapter 3
The Sensitivity of Population Growth Rate: Three Approaches

3.1 Introduction

The essence of stable population theory is the fact that a population subject to time-invariant vital rates will (with a few exceptions not of interest here) converge to a stable structure and grow exponentially at a constant rate (the population growth rate, or intrinsic rate of increase). The calculation of the population growth rate from the vital rates is one of the most important accomplishments of formal demography (Sharpe and Lotka 1911).[1] Ecologists recognized early on that, by integrating survival and fertility over the life course, the population growth rate provided a powerful tool for describing the population consequences of environmental conditions (e.g., Birch 1953). For the same reason, evolutionary biologists recognized it as a measure of fitness (Fisher 1930), although that concept requires careful consideration of both demographic and genetic processes (Charlesworth 1994; de Vries and Caswell 2018).

This makes the sensitivity analysis of population growth rate an important problem. It has been approached in three ways. The earliest approach (Hamilton 1966) is specific to age-classified models, and relies on differentiation of the characteristic equation. The second (Caswell 1978) applies to stage-classified as well as age-classified models, and uses eigenvalue perturbation theory. The third is based on matrix calculus and is more flexible than its predecessors.

---

[1]Leonard Euler had obtained the result in 1760, but his derivation rediscovered until 1970 (Keyfitz and Keyfitz 1970).

3.2 Hamilton's Equation for Age-Classified Populations

Consider an age-classified model, in which age x is a continuous variable, with mortality rate $\mu(x)$ and maternity function $m(x)$. The survivorship function is

$$\ell(x) = \exp\left(-\int_0^x \mu(a)da\right) \tag{3.1}$$

and the population growth rate r is the solution to the Euler-Lotka equation

$$1 = \int_0^\infty e^{-ra}\ell(a)m(a)da. \tag{3.2}$$

The stable age distribution, reproductive value function, birth rate, and generation time (mean age of reproduction in the stable population) are given by

$$c(x) = \frac{e^{-rx}\ell(x)}{\int_0^\infty e^{-ra}\ell(a)da} \qquad \text{stable age distribution} \tag{3.3}$$

$$v(x) = \frac{e^{rx}}{\ell(x)}\int_x^\infty e^{-ra}\ell(a)m(a)da \qquad \text{reproductive value} \tag{3.4}$$

$$b = \left[\int_0^\infty e^{-ra}\ell(a)da\right]^{-1} \qquad \text{birth rate} \tag{3.5}$$

$$\bar{A} = \int_0^\infty ae^{-ra}\ell(a)m(a)da \qquad \text{generation time} \tag{3.6}$$

Sensitivity of r Hamilton (1966) derived the sensitivities of r to changes in mortality and fertility at a specified age x. His results are equivalent to

$$\frac{dr}{d\mu(x)} = \frac{-c(x)v(x)}{b\bar{A}} \tag{3.7}$$

$$\frac{dr}{dm(x)} = \frac{c(x)}{b\bar{A}} \tag{3.8}$$

That is, the sensitivity of r to a change in mortality at age x is proportional to the product of the reproductive value at age x and the abundance of age x in the stable age distribution. The sensitivity of r to a change in fertility at age x is proportional to the stable age distribution (and the reproductive value at age 0, which equals 1). The proportionality constant in each case is the inverse of the product of the birth rate and the mean age of reproduction.

Derivation Hamilton's results are obtained by implicit differentiation of the Euler-Lotka equation (3.2). We will derive Hamilton's original formulation and then show how it reduces to the relation between the stable age distribution and the reproductive value distribution in (3.7) and (3.8).

First, introduce a perturbation parameter θ to measure the change in mortality or fertility at the specified age. Writing survival, fertility, and r as functions of θ gives the Euler-Lotka equation

$$1 = \int_0^\infty e^{-r(\theta)a} \ell(\theta, a) m(\theta, a) \, da. \tag{3.9}$$

Differentiating both sides of (3.9) with respect to θ gives

$$0 = -\frac{dr(\theta)}{d\theta} \int_0^\infty a e^{-r(\theta)a} \ell(\theta, a) m(\theta, a) \, da$$

$$+ \int_0^\infty e^{-r(\theta)a} \frac{d\ell(\theta, a)}{d\theta} m(\theta, a) \, da$$

$$+ \int_0^\infty e^{-r(\theta)a} \ell(\theta, a) \frac{dm(\theta, a)}{d\theta} \, da. \tag{3.10}$$

Solving (3.10) for $dr/d\theta$ gives

$$\frac{dr(\theta)}{d\theta} = \frac{1}{\bar{A}} \left(\underbrace{\int_0^\infty e^{-r(\theta)a} \frac{d\ell(\theta, a)}{d\theta} m(\theta, a) \, da}_{\text{mortality}} + \underbrace{\int_0^\infty e^{-r(\theta)a} \ell(\theta, a) \frac{dm(\theta, a)}{d\theta} \, da}_{\text{fertility}} \right)$$
$$\tag{3.11}$$

Equation (3.11) has two terms, one capturing effects of θ on mortality and the other capturing effects on fertility.

3.2.1 Effects of Changes in Mortality

We want to perturb mortality at one exact age x (remember that age and time are continuous), leaving mortality at all other ages unchanged. To do this, we use the unit impulse function, or Dirac delta function. This is a generalized function defined by

$$\delta(x) = 0 \qquad x \neq 0 \tag{3.12}$$

$$\int_{-\infty}^\infty \delta(s) ds = 1. \tag{3.13}$$

The unit impulse is used in signal processing (e.g., Kamen and Heck 1997, p. 7) to represent the limit of a perturbation of unit strength applied over a shorter and shorter time interval. Think of a normal distribution with mean 0, in the limit as the variance goes to 0, while the area under the curve remains at 1. The most useful properties of the unit impulse, for our application, are

$$\int_{-\infty}^{\infty} \delta(a - x)f(a)da = f(x) \tag{3.14}$$

and

$$\int_{-\infty}^{x} \delta(s)ds = H(x) \tag{3.15}$$

where $H(x)$ is the Heaviside function, or unit step function, which satisfies $H(x) = 0$ for $x < 0$ and $H(x) = 1$ for $x > 0$.

We write mortality as

$$\mu(\theta, a) = \mu(0, a) + \theta \delta(a - x) \tag{3.16}$$

where $\delta(x)$ is the unit impulse function. The sensitivity of r to $\mu(x)$ is obtained as the derivative of r with respect to θ, evaluated at $\theta = 0$,

$$\frac{dr}{d\mu(x)} = \left. \frac{dr}{d\theta} \right|_{\theta=0}. \tag{3.17}$$

Because only mortality is affected by θ

$$\frac{dm(\theta, a)}{d\theta} = 0 \tag{3.18}$$

$$\frac{d\mu(\theta, a)}{d\theta} = \delta(a - x). \tag{3.19}$$

From (3.1),

$$\frac{d\ell(\theta, a)}{d\theta} = -e^{-\int_0^a \mu(\theta, s)ds} \int_0^a \delta(a - x)da \tag{3.20}$$

$$= -\ell(\theta, a)H(a - x). \tag{3.21}$$

Substituting into (3.11) and evaluating at $\theta = 0$ gives

$$\frac{dr}{d\mu(x)} = \frac{-1}{\bar{A}} \left(\int_x^{\infty} e^{-ra} \ell(a)m(a)da \right). \tag{3.22}$$

The integral in (3.22) is close to the reproductive value $v(x)$ given by (3.4); specifically,

$$\int_x^\infty e^{-ra}\ell(a)m(a)da = \ell(x)e^{-rx}v(x).\qquad(3.23)$$

However, from (3.3) and (3.5), $\ell(x)e^{-rx} = c(x)/b$. Making these substitutions into (3.22) gives the formal relationship (3.7).

3.2.2 Effects of Changes in Fertility

Following the same approach, if the perturbation affects fertility at exact age x, we write

$$m(\theta, a) = m(0, a) + \theta\delta(a - x).\qquad(3.24)$$

Because only fertility is affected by θ, $d\mu(\theta, a)/d\theta = 0$ and $dm(\theta, a)/d\theta = \delta(a - x)$. Substituting these into (3.11) and evaluating the result at $\theta = 0$ gives

$$\frac{dr}{dm(x)} = \frac{1}{\bar{A}}\left(e^{-rx}\ell(x)\right).\qquad(3.25)$$

From (3.3) and (3.5) it can be seen that the numerator is $c(x)/b$, which leads to the formal relationship (3.8).

3.2.3 History and Perspectives

Hamilton (1966) obtained the relationship (3.22) in his analysis of the evolution of senescence. From (3.22) and (3.8) it is apparent that (provided $r \geq 0$) the magnitudes of the sensitivities of r to mortality and fertility decline with age. These sensitivities measure the selection gradients on age-specific mortality and fertility. Thus Hamilton concluded that the strength of selection against deleterious mutations would necessarily decline with their age of action, that small positive effects at early ages could easily compensate for much larger negative effects at later ages, and that the evolution of senescence was therefore inevitable.

In the years that followed Hamilton's paper, several other authors developed perturbation analysis for r, using related methods. Demetrius (1969) used a discrete age-classified model, and Emlen (1970) used Hamilton's results to derive the dynamics of gene frequencies resulting from the selection gradients on age-specific survival and fertility.

Keyfitz (1971) in a remarkable paper, used implicit differentiation to obtain the sensitivity of population growth rate, life expectancy, birth rates, death rates, and the stable age distribution, apparently independently of Hamilton. He noted the appearance of reproductive value in the sensitivity of r to mortality. Goodman (1971) was apparently the first to note that the sensitivities of r to mortality and fertility could be expressed in terms of the stable age distribution and reproductive value.

When Hamilton's paper appeared, it was regarded as difficult and esoteric, but it had a great impact. It provided the analytical machinery for examining trade-offs between opposing demographic traits, known as antagonistic pleiotropy (Williams 1957; Rose 1991). It also describes the accumulation of deleterious mutations due to the balance between mutation and selection (e.g., Steinsaltz et al. 2005). These ideas are fundamental to the analysis of human aging (e.g., Rose 1991; Wachter and Finch 1997; Carey and Tuljapurkar 2003; Baudisch 2008) and, more generally, the analysis of life history evolution in humans and other species (e.g., Charlesworth 1994; Stearns 1992).

3.3   Stage-Classified Populations: Eigenvalue Perturbations

Implicit in Hamilton's analysis is the assumption that the vital rates are functions of age. In many cases, they are not. In humans, characteristics such as education, marital status, health status, or spatial location, may provide important information in addition to age. In other species, the vital rates may depend on developmental stage or body size more than on age. Such populations are described by stage-classified demographic models, of which the age-classified theory is a special case.

Stage-classified demography can be analyzed using matrix population models (Leslie 1945; Caswell 2001). The discrete-time population growth rate λ is the dominant eigenvalue of the population projection matrix $\mathbf{A}$ (guaranteed to be real and positive by the Perron-Frobenius theorem). Let $\mathbf{n}(t)$ be the population vector at time t, and $\mathbf{A}$ the population projection matrix, with

$$\mathbf{n}(t+1) = \mathbf{An}(t) \tag{3.26}$$

and the population growth rate is given by the dominant eigenvalue λ of $\mathbf{A}$. The stable stage distribution is given by the corresponding right eigenvector $\mathbf{w}$ and the reproductive value function by the left eigenvector $\mathbf{v}$; they satisfy

$$\mathbf{Aw} = \lambda\mathbf{w} \tag{3.27}$$

$$\mathbf{v}^{\mathsf{T}}\mathbf{A} = \lambda\mathbf{v}^{\mathsf{T}} \tag{3.28}$$

Sensitivity of λ   The effects of perturbations on population growth are approached by looking for the sensitivity of an eigenvalue to changes in the entries of a matrix. We will see that the sensitivity of λ to a change in the entry a_{ij} of $\mathbf{A}$ is (Caswell 1978)

$$\frac{\partial \lambda}{\partial a_{ij}} = \frac{v_i w_j}{\mathbf{v}^\mathsf{T} \mathbf{w}}. \tag{3.29}$$

The entry a_{ij} measures the per-capita production of stage i by stage j. Thus the effect of a change in a_{ij} is proportional to the reproductive value of the destination stage and to the abundance of the origin stage in the stable population. This is a generalization of the relationships (3.7) and (3.8) obtained from Hamilton's analysis.

Derivation   The eigenvalue λ is a solution to the characteristic equation of $\mathbf{A}$, which generalizes the Euler-Lotka equation (3.2). Except in special cases, however, the characteristic equation cannot be written down explicitly, making the implicit differentiation approach used by Hamilton impossible. Instead, the relationship (3.29) is obtained by a perturbation expansion. Suppose that $\mathbf{A}$ is perturbed to $\mathbf{A} + \Delta\mathbf{A}$. This will result in perturbations of λ and of $\mathbf{w}$, which must satisfy

$$(\mathbf{A} + \Delta\mathbf{A})(\mathbf{w} + \Delta\mathbf{w}) = (\lambda + \Delta\lambda)(\mathbf{w} + \Delta\mathbf{w}). \tag{3.30}$$

Expanding the products, setting second order terms to zero, and remembering that $\mathbf{A}\mathbf{w} = \lambda\mathbf{w}$ gives

$$\mathbf{A}(\Delta\mathbf{w}) + (\Delta\mathbf{A})\mathbf{w} = \lambda(\Delta\mathbf{w}) + (\Delta\lambda)\mathbf{w}. \tag{3.31}$$

Multiply on the left by $\mathbf{v}^\mathsf{T}$ and simplify to obtain

$$(\Delta\lambda)\mathbf{v}^\mathsf{T}\mathbf{w} = \mathbf{v}^\mathsf{T}(\Delta\mathbf{A})\mathbf{w}. \tag{3.32}$$

If the perturbation affects only one entry, say a_{ij}, of $\mathbf{A}$, then

$$\Delta\lambda = \frac{v_i w_j \left(\Delta a_{ij}\right)}{\mathbf{v}^\mathsf{T}\mathbf{w}}. \tag{3.33}$$

Dividing both sides by Δa_{ij} and taking the limit as $\Delta a_{ij} \to 0$ gives the sensitivity result (3.29).

3.3.1   Age-Classified Models as a Special Case

To compare (3.29) with Hamilton's results (3.7) and (3.8), consider an age-classified matrix (a Leslie matrix) with fertilities F_i in the first row, survival probabilities P_i on

the subdiagonal, and zeros elsewhere (Leslie 1945; Keyfitz 1968). In this case (3.29) simplifies to

$$\frac{\partial \lambda}{\partial P_i} = \frac{v_{i+1} w_i}{\mathbf{v}^\mathsf{T} \mathbf{w}} \tag{3.34}$$

$$\frac{\partial \lambda}{\partial F_i} = \frac{v_1 w_i}{\mathbf{v}^\mathsf{T} \mathbf{w}}. \tag{3.35}$$

Equation (3.34) corresponds to (3.7); the sensitivity is proportional to the product of the reproductive value and the stable stage distribution. Equation (3.35) corresponds to (3.8), and shows why reproductive value is apparently missing from (3.8): reproductive value at birth [$v(0)$ in Hamilton's notation] is scaled to equal 1.

3.3.2   Sensitivity to Lower-Level Demographic Parameters

The entries of **A** are often functions of other, lower-level parameters. The sensitivity of λ to these parameters is obtained by the chain rule. For example, suppose that stage 1 may contribute individuals to stages 2 or 3 (Fig. 3.1). Write the transition probabilities as

$$a_{21} = \gamma \sigma \tag{3.36}$$

$$a_{31} = (1 - \gamma)\sigma \tag{3.37}$$

where σ is the survival probability and γ the probability that the individual moves to stage 2, conditional on survival. Then the sensitivities of λ to γ and to σ are given by

$$\frac{d\lambda}{d\sigma} = \frac{\partial \lambda}{\partial a_{21}} \frac{da_{21}}{d\sigma} + \frac{\partial \lambda}{\partial a_{31}} \frac{da_{31}}{d\sigma} \tag{3.38}$$

$$= \frac{w_1 \left[\gamma v_2 + (1 - \gamma) v_3 \right]}{\mathbf{v}^\mathsf{T} \mathbf{w}} \tag{3.39}$$

Fig. 3.1 An example of lower-level parameters appearing in a portion of a life cycle. Individuals in stage 1 survive with probability σ, and, conditional on survival, move to stage 2 with probability γ and to stage 3 with probability $1 - \gamma$
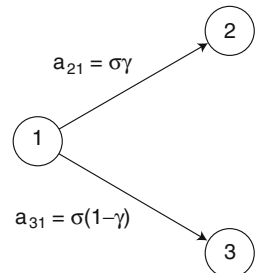
$$\frac{d\lambda}{d\gamma} = \frac{\partial\lambda}{\partial a_{21}} \frac{da_{21}}{d\gamma} + \frac{\partial\lambda}{\partial a_{31}} \frac{da_{31}}{d\gamma} \tag{3.40}$$

$$= \frac{\sigma\, w_1\, (v_2 - v_3)}{\mathbf{v}^{\mathsf{T}}\mathbf{w}}. \tag{3.41}$$

The sensitivity to survival is proportional to the weighted average of the reproductive values of the destination stages, and the sensitivity to the transition probability γ is proportional to the difference in reproductive value between the destination stages.

3.3.3   History

I first encountered the basis for this perturbation expansion in a paper by C.A. Desoer in the proceedings of an engineering conference (Desoer 1967).[2] Eigenvalue perturbations were of particular interest to engineers in the 1960s as part of a shift from frequency-domain methods to state-space methods in the study of linear systems (Zadeh and Desoer 1963). However, the result dates back to Jacobi (1846), and has been independently rediscovered many times (e.g., Faddeev 1959; Papoulis 1966; Franklin 1968). In population biology, this perturbation approach has been extended to many other sensitivity problems, including the sensitivity of subdominant eigenvalues and transient behavior, of growth rates in periodic and stochastic environments, of the eigenvectors, and of the spreading speed in biological or demographic invasions (see Caswell (2001) for reviews and references).

3.4   Growth Rate Sensitivity via Matrix Calculus

Matrix calculus provides a still more general approach to the sensitivity analysis of the population growth rate. Equation (3.29) perturbs only a single entry of $\mathbf{A}$; derivatives with respect to other parameters are assembled by summing their effects over all the entries of $\mathbf{A}$, as in (3.41). Using matrix calculus, we now consider λ as a scalar function of $\mathbf{A}$ and $\mathbf{A}$ as a matrix-valued function of a parameter vector $\boldsymbol{\theta}$.

Sensitivity of λ   We will show that the derivative of λ with respect to $\boldsymbol{\theta}$ is

$$\frac{d\lambda}{d\boldsymbol{\theta}^{\mathsf{T}}} = \left(\frac{\mathbf{w}^{\mathsf{T}} \otimes \mathbf{v}^{\mathsf{T}}}{\mathbf{v}^{\mathsf{T}}\mathbf{w}} \right) \left(\frac{d\operatorname{vec}\mathbf{A}}{d\boldsymbol{\theta}^{\mathsf{T}}} \right), \tag{3.42}$$

---

[2]By a fortunate accident; I was searching for something completely different. We may wonder whether the chances of such coincidences are higher or lower in the internet search era.

where $\otimes$ denotes the Kronecker product. If $\boldsymbol{\theta}$ is a $p \times 1$ vector of parameters, then $d\lambda/d\boldsymbol{\theta}^\mathsf{T}$ is a $1 \times p$ matrix whose ith entry is $d\lambda/d\theta_i$.

Derivation Following the steps in Chap. 2, begin by taking the differential of both sides of (3.27) to give

$$(d\mathbf{A})\mathbf{w} + \mathbf{A}(d\mathbf{w}) = (d\lambda)\mathbf{w} + \lambda(d\mathbf{w}). \tag{3.43}$$

Multiply both sides on the left by $\mathbf{v}^\mathsf{T}$ and simplify to obtain

$$(d\lambda)\mathbf{v}^\mathsf{T}\mathbf{w} = \mathbf{v}^\mathsf{T}(d\mathbf{A})\mathbf{w} \tag{3.44}$$

Next, apply the vec operator to both sides of (3.44). Since the left side is a scalar, the vec operator has no effect. The right side is a product of three quantities, so Roth's theorem implies that

$$(d\lambda)\mathbf{v}^\mathsf{T}\mathbf{w} = \left(\mathbf{w}^\mathsf{T} \otimes \mathbf{v}^\mathsf{T}\right) d\operatorname{vec}\mathbf{A}. \tag{3.45}$$

The First Identification Theorem then gives

$$\frac{d\lambda}{d\operatorname{vec}^\mathsf{T}\mathbf{A}} = \frac{\mathbf{w}^\mathsf{T} \otimes \mathbf{v}^\mathsf{T}}{\mathbf{v}^\mathsf{T}\mathbf{w}}. \tag{3.46}$$

Finally, the chain rule (2.18) gives us

$$\frac{d\lambda}{d\boldsymbol{\theta}^\mathsf{T}} = \frac{d\lambda}{d\operatorname{vec}^\mathsf{T}\mathbf{A}} \frac{d\operatorname{vec}\mathbf{A}}{d\boldsymbol{\theta}^\mathsf{T}}. \tag{3.47}$$

The matrix calculus approach is particularly powerful because of the flexibility in specifying the effect of $\boldsymbol{\theta}$ on the vital rates. Suppose that $\mathbf{A}$ depends on a vector $\boldsymbol{\sigma}$ of survival probabilities, which are a function of the concentration X of a pollutant, which in turn is changing as a function of time t. The rate of change of λ over time is

$$\frac{d\lambda}{dt} = \left(\frac{d\lambda}{d\operatorname{vec}^\mathsf{T}\mathbf{A}}\right)\left(\frac{d\operatorname{vec}\mathbf{A}}{d\boldsymbol{\sigma}^\mathsf{T}}\right)\left(\frac{d\boldsymbol{\sigma}}{dX}\right)\left(\frac{dX}{dt}\right) \tag{3.48}$$

Each of the terms in (3.48) can be evaluated separately; the matrix product gives the correct dimension for the final sensitivity result (a 1×1 scalar in this case).

3.5   Second Derivatives of Population Growth Rate

The second derivatives of λ measure the curvature of the response to changes in parameters. They have important applications in evolutionary demography, where they indicate the action of stabilizing, disruptive, or correlational selection on

fitness-related traits (e.g., Phillips and Arnold 1989; Caswell 2001), in adaptive dynamics, where they help determine the stability of evolutionary singular strategies (e.g., Diekmann 2004), and in extending sensitivity analysis to second-order effects.

Since the first derivatives of λ are written, in Eqs. (3.29) and (3.46), in terms of the right and left eigenvectors of $\mathbf{A}$, the second derivatives of λ require the first derivatives of those eigenvectors. Caswell (1996) derived the second derivatives of λ to entries of $\mathbf{A}$ by an extension of the method in Sect. 3.3. However, a more general and rigorous method is available using matrix calculus.

Consider a (scalar) variable ξ which is a function of a vector $\boldsymbol{\theta}$ of parameters. The complete set of second derivatives of ξ are given by the Hessian matrix

$$\mathbf{H} = \left(\frac{\partial^2 \xi}{\partial \theta_i \partial \theta_j} \right) \tag{3.49}$$

Magnus and Neudecker (1988) proved (their Second Identification Theorem) that if the second differential of ξ can be written as

$$d^2\xi = d\boldsymbol{\theta}^\mathsf{T} \mathbf{B} d\boldsymbol{\theta} \tag{3.50}$$

for some matrix $\mathbf{B}$, then

$$\mathbf{H} = \frac{1}{2} \left(\mathbf{B} + \mathbf{B}^\mathsf{T} \right). \tag{3.51}$$

Shyu and Caswell (2014) used this approach to derive the second derivatives of the population growth rate λ, the continuous-time population growth rate $r = \log \lambda$, and the net reproductive rate R_0, to changes in either the entries of $\mathbf{A}$ or to arbitrary lower-level parameters of which $\mathbf{A}$ is a function. We will not explore second derivatives in this book, but Shyu's other work (Shyu and Caswell 2016a,b) applies them to analyze the evolutionary demography of sex ratios, and Caswell and Shyu (2017) use them to analyze the effects of mortality on the selection gradients on senescence.

3.6 Conclusion

Each of the three approaches to growth rate sensitivity, leading to Eqs. (3.7), (3.8), (3.29), and (3.42), uses its own analytical methods. They agree, however, in showing how the sensitivity of population growth rate can be written in terms of the stable stage distribution and the reproductive value. In general, the effect of a change in the rate at which individuals move from stage j to stage i is proportional to the abundance of the origin stage (j) and the reproductive value of the destination stage (i). If a transition yields individuals with low reproductive value, or if few individuals are available to experience the change in the rate of transition, the effect on population growth will be small.

Bibliography

Baudisch, A. 2008. Inevitable aging? Contributions to evolutionary-demographic theory. Springer-Verlag, Berlin, Germany.

Birch, L. C. 1953. Experimental background to the study of the distribution and abundance of insects: I. The influence of temperature, moisture and food on the innate capacity for increase of three grain beetles. Ecology **34**:698–711.

Carey, J. R., and S. Tuljapurkar, editors, 2003. Life span: evolutionary, ecological, and demographic perspectives, volume 29. Population Council, New York, New York.

Caswell, H. 1978. A general formula for the sensitivity of population growth rate to changes in life history parameters. Theoretical Population Biology **14**:215–230.

Caswell, H. 1996. Second derivatives of population growth rate: calculation and applications. Ecology **77**:870–879.

Caswell, H. 2001. Matrix Population Models: Construction, Analysis, and Interpretation. 2nd edition. Sinauer Associates, Sunderland, MA.

Caswell, H., and E. Shyu, 2017. Senescence, selection gradients, and mortality. Chapter 4, pages 56–82 *in* R. P. Shefferson, O. R. Jones, and R. Salguero-Gómez, editors. The Evolution of Senescence in the Tree of Life. Cambridge University Press.

Charlesworth, B. 1994. Evolution in Age-Structured Populations. 2nd edition. Cambridge University Press, Cambridge, UK.

de Vries, C., and H. Caswell. 2018. Stage-stuctured evolutionary demography: linking life histories, population genetics, and ecological dynamics. (in prep).

Demetrius, L. 1969. The sensitivity of population growth rate to perturbations in the life cycle components. Mathematical Biosciences **4**:129–136.

Desoer, C. A., 1967. Perturbations of eigenvalues and eigenvectors of a network. Pages 8–11 *in* Fifth Annual Allerton Conference on Circuit and System Theory. University of Illinois, Urbana, Illinois, USA.

Diekmann, O. 2004. A beginner's guide to adaptive dynamics. Banach Center Publications **63**:47–86.

Emlen, J. M. 1970. Age specificity and ecological theory. Ecology **51**:588–601.

Faddeev, D. K. 1959. The conditionality of matrices ("Ob obuslovlennosti matrits"). Matematicheskii institut Steklov, Trudy No **53**:387–391.

Fisher, R. A. 1930. The Genetical Theory of Natural Selection. Revised 1956, Dover Publications edition. Oxford University Press, Oxford, U.K.

Franklin, J. N. 1968. Matrix theory. Prentice-Hall, Englewood Cliffs, New Jersey, USA.

Goodman, L. A. 1971. On the sensitivity of the intrinsic growth rate to changes in the age-specific birth and death rates. Theoretical Population Biology **2**:339–354.

Hamilton, W. D. 1966. The moulding of senescence by natural selection. Journal of Theoretical Biology **12**:12–45.

Jacobi, C. J. G. 1846. Uber ein leichtes Verfagren die in der Theorie der Sacularstorungen Vorkommenden Gleichungen numerisch aufsulosen. J. Reine Angew. Math **30**:51–95.

Kamen, E. W., and B. S. Heck. 1997. Fundamentals of signals and systems. Prentice Hall, Upper Saddle River, New Jersey, USA.

Keyfitz, N. 1968. Introduction to the Mathematics of Population. Addison-Wesley, Reading, Massachusetts, USA.

Keyfitz, N. 1971. Linkages of intrinsic to age-specific rates. Journal of the American Statistical Association **66**:275–281.

Keyfitz, N., and B. Keyfitz. 1970. Euler, L. 1760. Recherches générales sur la mortalité et la multiplication [General researches on mortality and multiplication]. Mémoires de l'Académie Royale des Sciences et Belles Lettres 16:144–164. Translated by N. Keyfitz and B. Keyfitz. Theoretical Population Biology **1**:307–314.

Leslie, P. H. 1945. On the use of matrices in certain population mathematics. Biometrika **33**:183–212.

Magnus, J. R., and H. Neudecker. 1988. Matrix differential calculus with applications in statistics and econometrics. John Wiley and Sons, New York, New York.

Papoulis, A. 1966. Perturbations of the natural frequencies and eigenvectors of a network. IEEE Transactions on Circuit Theory **CT-13**:188–195.

Phillips, P. C., and S. J. Arnold. 1989. Visualizing multivariate selection. Evolution **43**:1209–1222.

Rose, M. R. 1991. Evolutionary biology of aging. Oxford University Press, Oxford, U.K.

Sharpe, F. R., and A. J. Lotka. 1911. A problem in age-distribution. Philosophical Magazine **21**:435–438.

Shyu, E., and H. Caswell. 2014. Calculating second derivatives of population growth rates for ecology and evolution. Methods in Ecology and Evolution **5**:473–482.

Shyu, E., and H. Caswell. 2016*a*. A demographic model for sex ratio evolution and the effects of sex-biased offspring costs. Ecology and Evolution **6**:1470–1492.

Shyu, E., and H. Caswell. 2016*b*. Frequency-dependent two-sex models: a new approach to sex ratio evolution with multiple maternal conditions. Ecology and Evolution **6**:6855–6879.

Stearns, S. C. 1992. The evolution of life histories. Oxford University Press, Oxford, England.

Steinsaltz, D., S. N. Evans, and K. W. Wachter. 2005. A generalized model of mutation-selection balance with applications to aging. Advances in Applied Mathematics **35**:16–33.

Wachter, K. W., and C. E. Finch. 1997. Between Zeus and the salmon: the biodemography of longevity. National Academy Press, Washington DC, USA.

Williams, G. C. 1957. Pleiotropy, natural selection, and the evolution of senescence. Evolution **11**:398–411.

Zadeh, L. A., and C. A. Desoer. 1963. Linear system theory. McGraw-Hill, New York, New York, USA.

Chapter 4
Sensitivity Analysis of Longevity and Life Disparity

4.1 Introduction

The population growth rate (λ or r) analyzed in Chap. 3 is a population-level consequence of the individual-level vital rates. A similarly basic outcome, at the individual or cohort level, is longevity: the length of individual life. The most commonly encountered description of longevity is its expectation, the life expectancy. However, longevity is a random variable, differing among individuals (even when those individuals are subject to the same rates and hazards) because of the random vagaries of mortality and survival. Therefore, it is important to also consider its variance and higher moments. This chapter introduces the sensitivity analysis of longevity, which will be explored in more detail in Chaps. 5, 11, and 12.

As in Chap. 3, we will begin by reviewing a classic formula for the sensitivity of life expectancy in age-classified models. The we will use matrix calculus to derive more general formulas for the moments of longevity, the distribution of age or stage at death, and the life disparity, applicable to age- or stage-classified populations.

4.2 Life Expectancy in Age-Classified Populations

Notation It is customary to denote life expectancy by symbols like e_x^o or $e(x)$, but in general the symbol e plays too many roles in mathematics to be helpful for our purposes. So, when we make the transition to matrix formulations, I will use the symbol η, in various vector and scalar manifestations, to indicate longevity.

Perturbation analysis of longevity has been pursued mostly within the framework of age-classified life cycles (e.g., Canudas Romo 2003; Keyfitz 1971; Pollard 1982; Vaupel 1986; Vaupel and Canudas Romo 2003). The life expectancy at age x is given by

$$e(x) = \frac{1}{\ell(x)} \int_x^\infty \ell(s)ds \tag{4.1}$$

where the survivorship function $\ell(x)$ is the probability of survival to age x.

The classical result for the sensitivity of life expectancy at birth to a change in mortality at age a is

$$\frac{de(0)}{d\mu(a)} = -\ell(a)e(a). \tag{4.2}$$

That is, the sensitivity of life expectancy at birth to a change in mortality at age a is equal to the product of the probability of survival to age a and the life expectancy at age a. In other words, $e(0)$ is most sensitive to changes in mortality at ages to which lots of individuals survive (to experience the change in mortality) and beyond which there is lots of longevity remaining (so they can enjoy the change in mortality). The derivative is negative because increasing mortality reduces life expectancy.

The result was presented independently by Keyfitz (1971) who also referenced some earlier approaches (Wilson 1938; Irwin 1949) and by Pollard (1982). Keyfitz's derivation was sketchy, and Pollard simply stated that the result was well-known, and gave no derivation. From a general sensitivity analysis perspective, we can derive the result using the same approach applied in Chap. 3 to population growth rate.

4.2.1   Derivation

Differentiating (4.1) with respect to mortality at some specified age a gives

$$\frac{de(0)}{d\mu(a)} = \int_0^\infty \frac{d\ell(s)}{d\mu(a)} ds \tag{4.3}$$

and our problem reduces to finding the derivative of $\ell(s)$ with respect to $\mu(a)$. To do so, introduce a parameter θ to measure the size of the perturbation at age a, and write mortality as

$$\mu(x, \theta) = \mu(x, 0) + \theta \, \delta(x - a) \tag{4.4}$$

where $\delta(x - a)$ is the Dirac delta function.[1] The derivative with respect to $\mu(a)$ is obtained by differentiating with respect to θ and evaluating the result at $\theta = 0$.

---

[1] See Chap. 3 for a description of this generalized function.

Write survivorship as

$$\ell(x, \theta) = \exp\left[-\int_0^x \mu(z, \theta)dz\right] \tag{4.5}$$

so that

$$\frac{d\ell(x, \theta)}{d\theta} = -\ell(x, \theta)\int_0^x \frac{d\mu(z, \theta)}{d\theta}dz \tag{4.6}$$

From (4.4) we have

$$\frac{d\mu(z, \theta)}{d\theta} = \delta(z - a) \tag{4.7}$$

so that

$$\frac{d\ell(x, \theta)}{d\theta} = -\ell(x, \theta)\int_0^\infty \delta(z - a)dz \tag{4.8}$$

$$= -\ell(x, \theta)H(x - a) \tag{4.9}$$

where $H(\cdot)$ is the unit step function. Substituting this into (4.3) and evaluating at $\theta = 0$ gives

$$\frac{de(0)}{d\mu(a)} = -\int_0^\infty \ell(s)H(s - a)ds \tag{4.10}$$

$$= -\int_a^\infty \ell(s)ds \tag{4.11}$$

which, by (4.1) is equal to (4.2).

4.3 A Markov Chain Model for the Life Cycle

Age has a special status in demography because it is continuous, linear, and permits movement in only one direction and at one rate (age increases by one unit for every unit of time). All other demographic characteristics have the potential for much greater flexibility, and the operators that describe movement and development of individuals require an equal degree of flexibility. This book is devoted to matrix formulations of these problems, which have the great advantage of permitting both age and stage-classified models. The basic formulation, as far as longevity is concerned, is that of a finite-state absorbing Markov chain.

4.3.1   A Markov Chain Formulation of the Life Cycle

We describe the life cycle as an absorbing Markov chain. This approach was pioneered in demography by Feichtinger (1971) and Hoem (1969), and has been greatly extended in recent years (Caswell 2001, 2006, 2009; Horvitz and Tuljapurkar 2008; Tuljapurkar and Horvitz 2006; Steinsaltz and Evans 2004). Good sources for the basic theory of absorbing Markov chains are Kemeny and Snell (1976) and Iosifescu (1980).

These models will be explored in more detail in Chaps. 5 and 11. The sensitivity analysis of measures of variance in longevity has been developed by Van Raalte and Caswell (2013) and Engelman et al. (2014). An important extension of Markov chain models for longevity is the incorporation of "rewards" to represent the value, in some sense, of the length of life, extending methods developed for dynamic programming (Howard 1960). The rewards include the production of offspring (Caswell 2011; van Daalen and Caswell 2015, 2017), the accumulation of income and expenditures (Caswell and Kluge 2015) and healthy longevity (Caswell and Zarulli 2018). The sensitivity analysis of these important models is derived in van Daalen and Caswell (2017).

Markov chain theory distinguishes between *recurrent* and *transient* states. A recurrent state has the property that the probability of returning to that state at least once is 1. A transient state is one for which that probability is less than 1. If a Markov chain contains transient states, it will eventually leave those states and arrive in a recurrent state or class of states, where it will remain permanently. Such a chain is called *absorbing*. Absorbing chains are the basic model for the demography of individuals because life is inherently transient. Any individual will, with probability one, eventually leave the set of living states and be absorbed by death.

If a Markov chain consists of a single set of recurrent states that all communicate with each other, it is said to be ergodic. The transition matrix for an ergodic chain is irreducible and primitive. Ergodic Markov chains play a limited role in demographic contexts because they cannot include mortality. Chapter 11 will, however, present the sensitivity analysis of these models.

In demographic models, individuals move among a set of transient (i.e., living) states in their life cycle before they eventually reach an absorbing state (death). Transient states may represent age classes, developmental or life history stages, or states defined by health, employment, economic, or other kinds of status. In studying longevity, we are particularly interested in absorbing states representing death, or perhaps death classified by age or stage at death, or by cause of death. The analysis applies equally to other ways of leaving the life cycle (e.g., graduation in a model of educational states, discharge from treatment in model of health states).

Number the stages in the life cycle so that the transient states are $1, \ldots, s$ and the absorbing states are $s + 1, \ldots, s + a$. Then the transition matrix of the Markov chain is

$$\mathbf{P} = \left(\begin{array}{c|c} \mathbf{U} & 0 \\ \hline \mathbf{M} & \mathbf{I} \end{array} \right) \tag{4.12}$$

Here, $\mathbf{U}$ is the $s \times s$ matrix of transition probabilities among the transient states. The $a \times s$ matrix $\mathbf{M}$ gives the probabilities of absorption in each of the absorbing states. The columns of $\mathbf{P}$ sum to one. I assume that the spectral radius (the dominant eigenvalue) of $\mathbf{U}$ is strictly less than one; a sufficient condition for this is that there is a non-zero probability of ultimate death for every stage.

Age-classified models are a special case with survival probabilities on the subdiagonal (and possibly in the last diagonal entry); e.g., for $s = 3$ in which

$$\mathbf{U} = \begin{pmatrix} 0 & 0 & 0 \\ p_1 & 0 & 0 \\ 0 & p_2 & p_3 \end{pmatrix} \tag{4.13}$$

The age-specific survival probability is $p_i = e^{-\mu_i}$, with μ_i a mortality rate applying to age class i. The (s, s) entry of $\mathbf{U}$ is an age-independent survival probability for a final open-ended age class, with a remaining life expectancy of $1/(1 - p_s)$. If $p_s = 0$ no one survives beyond age class s. When the age-classified model is constructed from a life table, $p_i = 1 - q_{i-1}$; that is, the survival of age-class 1 is the complement of the probability of death between age 0 and 1.

The mortality matrix $\mathbf{M}$ gives the probabilities of transition from each of the transient states to each of the absorbing states. Figure 4.1 shows some examples of life cycle formulations that can arise, including both age and stage classification in the transient states, and absorbing states classified by age at death, grouped ages at death, stage at death, or cause of death. The resulting mortality matrices are

$$\text{Figure 4.1a} \qquad \mathbf{M} = \begin{pmatrix} 1 - P_1 & 1 - P_2 & 1 - P_3 & 1 \end{pmatrix} \tag{4.14}$$

$$\text{Figure 4.1b} \qquad \mathbf{M} = \begin{pmatrix} 1 - P_1 & 0 & 0 & 0 \\ 0 & 1 - P_2 & 0 & 0 \\ 0 & 0 & 1 - P_3 & 0 \\ 0 & 0 & 0 & 1 - P_4 \end{pmatrix} \tag{4.15}$$

$$\text{Figure 4.1c} \qquad \mathbf{M} = \begin{pmatrix} 1 - P_1 & 1 - P_2 & 0 & 0 \\ 0 & 0 & 1 - P_3 & 1 - P_4 \end{pmatrix} \tag{4.16}$$

$$\text{Figure 4.1d} \qquad \mathbf{M} = \begin{pmatrix} q_1 & 0 & 0 & 0 \\ 0 & q_2 & 0 & 0 \\ 0 & 0 & q_3 & 0 \\ 0 & 0 & 0 & q_4 \end{pmatrix} \tag{4.17}$$

(a) One dead state

(b) Age at death

(c) Grouped age at death

(d) Stage at death
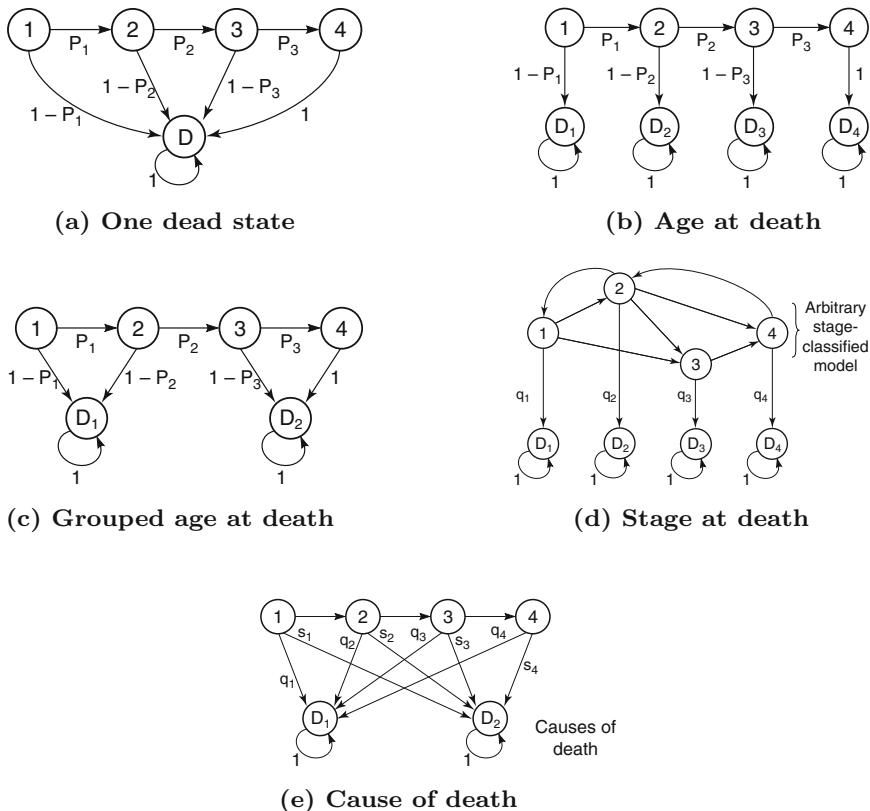
(e) Cause of death

Fig. 4.1 Life cycle graphs showing some alternative choices for structure of the absorbing state: death, age at death, stage at death, or cause of death. (**a**) Age-classified with one dead state. (**b**) Age-classified, age at death. (**c**) Age-classified, grouped ages at death. (**d**) Stage-classified, stage at death. (**e**) Age-classified, causes of death

$$\text{Figure 4.1e} \qquad \mathbf{M} = \begin{pmatrix} q_1 \; q_2 \; q_3 \; q_4 \\ s_1 \; s_2 \; s_3 \; s_4 \end{pmatrix} \qquad\qquad (4.18)$$

The beauty of formulating longevity as a Markov chain is that many statistics of longevity can be written in terms of the matrices $\mathbf{U}$ and $\mathbf{M}$ and sensitivity analysis can be carried out using matrix calculus.

4.3.2 Occupancy Times

Consider an individual in transient state j. Eventual absorption is certain. But before that, the individual will occupy various transient states. The number of such visits,

the occupancy time[2] is the basic unit of longevity. Occupancy is particularly central in studies of health demography, where it quantifies the parts of a life spent in different health states. But, even without the added dimension of something like health, occupancy of transient states is the basis of longevity analysis.

Let v_{ij} be the number of visits to transient state i by an individual in transient state j, prior to absorption. Its expectation is given by the fundamental matrix (e.g., Kemeny and Snell 1976; Iosifescu 1980)

$$\mathbf{N} = \left(E(v_{ij}) \right) \tag{4.19}$$

$$= (\mathbf{I} - \mathbf{U})^{-1} \tag{4.20}$$

More details, and examples, for the higher moments and variances of occupancy times are given in Chaps. 5 and 11.

4.3.3 Longevity

The longevity of an individual in state j can be equated to the total occupancy time of all transient states by that individual, prior to eventual absorption. Let η_j be this longevity; the expectation of η_j is the sum of the elements in column j of $\mathbf{N}$. We define $\boldsymbol{\eta}_1$ and $\boldsymbol{\eta}_2$ as the vectors containing the first and second moments of longevity, respectively. Then

$$E(\boldsymbol{\eta})^{\mathsf{T}} = \boldsymbol{\eta}_1^{\mathsf{T}} = \mathbf{1}^{\mathsf{T}}\mathbf{N} \tag{4.21}$$

Figure 4.2a shows the life expectancy for India in 1961 and Japan in 2006.

The vector of the second moments of longevity satisfies

$$\boldsymbol{\eta}_2^{\mathsf{T}} = \boldsymbol{\eta}_1^{\mathsf{T}} \left(2\mathbf{N} - \mathbf{I} \right) \tag{4.22}$$

(Iosifescu 1980). The variance and standard deviation of longevity are thus

$$V(\boldsymbol{\eta})^{\mathsf{T}} = \boldsymbol{\eta}_2 - \boldsymbol{\eta}_1 \circ \boldsymbol{\eta}_1 \tag{4.23}$$

$$SD(\boldsymbol{\eta}) = \sqrt{V(\boldsymbol{\eta})} \tag{4.24}$$

where the square root is taken element-wise.

---

[2]Because time is discrete here, the number of visits is equal to the number of time increments, which is the amount of time spent in the state. In continuous-time models, the number of visits to, and the length of time spent in, a transient state are different. The corresponding calculations for continuous-time models are given in Chap. 12.

(a) Life expectancy

(b) SD of longevity

(c) Sensitivity of life expectancy

(d) Sensitivity of variance in longevity

(e) Sensitivity of disparity
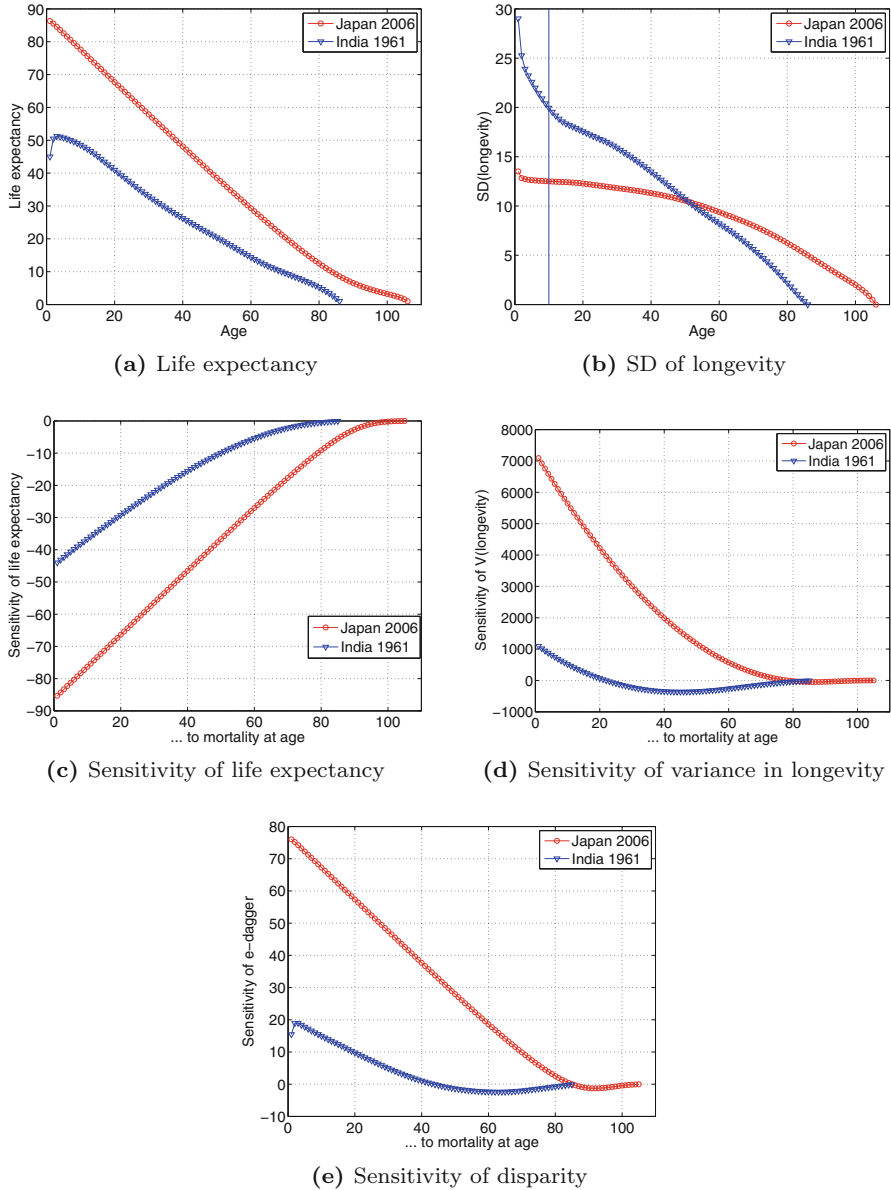
Fig. 4.2 Calculations for longevity of India (1961) and Japan (2006). (**a**) Remaining life expectancy as a function of age. (**b**) Standard deviation of remaining longevity as a function of age. Vertical line at age 10 indicates SD_{10}, sometimes used as a measure of lifespan disparity. (**c**) Sensitivity of life expectancy at birth to changes mortality at each age. (**d**) Sensitivity of variance in longevity at birth to changes in mortality at each age. (**e**) Sensitivity of life disparity $\eta^{\dagger}$ to changes in mortality at each age

Note that $V(\boldsymbol{\eta})$ and $SD(\boldsymbol{\eta})$ are vectors; their elements give the variance or standard deviation of longevity for individuals in each stage, making it easy to examine variation in remaining longevity conditional on the starting age. This conditioning can be important; Edwards and Tuljapurkar (2005) have made a strong case that $SD(\eta_{10})$, starting from age 10, is a good index to prevent infant and child mortality from obscuring patterns in old age longevity.

Figure 4.2b shows $SD(\boldsymbol{\eta})$ for India and Japan. The standard deviation at birth, $SD(\eta_1)$ is roughly twice as great in India as in Japan, a discrepancy that remains at $SD(\eta_{10})$. Eventually, beyond the age of 50, $SD(\boldsymbol{\eta})$ becomes greater in India than in Japan.

4.3.4 Age or Stage at Death

If the model contains more than one absorbing state (as in all the cases but the first in Fig. 4.1), the eventual fate of an individual is uncertain. The probability distributions of the eventual absorbing state are given by the columns of the matrix

$$\mathbf{B} = \mathbf{MN} \tag{4.25}$$

where b_{ij} is the probability of eventual absorption in absorbing state i for an individual starting in transient state j (Iosifescu 1980).

Suppose that the absorbing stages are defined as the age (or stage) at death, as in Fig. 4.1b, d. Then $\mathbf{M}$ is given by Eq. (4.17) and the jth column of $\mathbf{B}$ is the probability distribution of age at death for an individual starting in age class j:

$$\boldsymbol{\psi}_j = \mathbf{B}(:, j) = \mathbf{B}\mathbf{e}_j. \tag{4.26}$$

4.3.5 Life Lost and Life Disparity

When an individual dies, it loses the remaining life that it would have experienced, had it not died. This counterfactual proposition seems abstract, but we can make it concrete by asking for the expectation of that lost lifetime. An individual that dies at age x will lose, on average, an amount of life given by the life expectancy at age x. Averaging this remaining life expectancy over the distribution of age at death gives the mean life lost due to mortality. Vaupel and Canudas Romo (2003) denoted the life lost by $e^{\dagger}$. Here we define the vector $\boldsymbol{\eta}^{\dagger}$, whose ith entry is the expected life lost due to mortality by an individual starting in age class i; it is given by

$$\left(\boldsymbol{\eta}^{\dagger}\right)^{\mathsf{T}} = \boldsymbol{\eta}_1^{\mathsf{T}}\mathbf{B}. \tag{4.27}$$

Calculations of life lost from mortality due to specific causes of death play a central role in the calculations of disability-adjusted life years (DALYs) used in calculations of the burden of diseases (e.g., Devleesschauwer et al. 2014; GBD 2016 DALYs and HALE Collaborators 2017). See Caswell and Zarulli (2018) for the relationship between DALY calculations and Markov chain methods, and for a calculation of the variance in life lost.

The life lost $\eta^\dagger$ has an additional interpretation as a measure of disparity. Consider a population in which everyone dies at the same age. In such a situation, $\eta^\dagger = 0$, because at the age of death, there is no additional life expectancy. Thus $\eta^\dagger$ is a measure of "life disparity;" the larger its value, the more disparity there is among individuals in age at death (Vaupel et al. 2011).

The values of life disparity in age class 1, for Japan and India, in years, are

$$\eta_1^\dagger = \begin{cases} 10.1 & \text{Japan} \\ 23.9 & \text{India} \end{cases} \tag{4.28}$$

Just as India has a much larger variance in longevity than Japan, it also has a higher life disparity.

4.4 Sensitivity Analysis

Our goal is to obtain expressions for the derivatives of $E(\eta)$, $V(\eta)$, $SD(\eta)$, $\mathbf{B}$, and $\eta^\dagger$, with respect to changes in age specific-mortality rates. The calculations and some results (contrasting the mortality schedules of Japan and India) are given here. More details are presented in Chaps. 5 and 11. Results are presented in terms of an arbitrary vector $\boldsymbol{\theta}$ of parameters on which $\mathbf{U}$ and $\mathbf{M}$ depend. In the examples, $\boldsymbol{\theta}$ will be the vector $\boldsymbol{\mu}$ of age-specific mortality rates.

4.4.1 Sensitivity of the Fundamental Matrix

The fundamental matrix $\mathbf{N}$ appears in many of these formulas. Its sensitivity was first obtained by Caswell (2006). Suppose that $\mathbf{U}$ is a function of some vector $\boldsymbol{\theta}$ of parameters. Then

$$\frac{d \operatorname{vec} \mathbf{N}}{d \boldsymbol{\theta}^\mathsf{T}} = \left(\mathbf{N}^\mathsf{T} \otimes \mathbf{N} \right) \frac{d \operatorname{vec} \mathbf{U}}{d \boldsymbol{\theta}^\mathsf{T}} \tag{4.29}$$

(see Chap. 5).

4.4.2   Sensitivity of Life Expectancy

The sensitivity of the vector of life expectancy as a function of age is obtained by differentiating (4.21),

$$d\boldsymbol{\eta}_1^\mathsf{T} = \mathbf{1}^\mathsf{T}(d\mathbf{N}) \tag{4.30}$$

Applying the vec operator and Roth's theorem (2.13) gives

$$d\boldsymbol{\eta}_1 = \left(\mathbf{I} \otimes \mathbf{1}^\mathsf{T}\right) d\text{vec}\,\mathbf{N} \tag{4.31}$$

$$= \left(\mathbf{I} \otimes \mathbf{1}^\mathsf{T}\right)\left(\mathbf{N}^\mathsf{T} \otimes \mathbf{N}\right) d\text{vec}\,\mathbf{U} \tag{4.32}$$

$$= \left(\mathbf{N}^\mathsf{T} \otimes \boldsymbol{\eta}_1^\mathsf{T}\right) d\text{vec}\,\mathbf{U}. \tag{4.33}$$

The last step uses the fact that $(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = (\mathbf{A}\mathbf{C} \otimes \mathbf{B}\mathbf{D})$. Applying the chain rule and the first identification theorem gives the result

$$\frac{d\boldsymbol{\eta}_1}{d\boldsymbol{\theta}^\mathsf{T}} = \left(\mathbf{N}^\mathsf{T} \otimes \boldsymbol{\eta}_1^\mathsf{T}\right)\frac{d\text{vec}\,\mathbf{U}}{d\boldsymbol{\theta}^\mathsf{T}} \tag{4.34}$$

Sensitivity to mortality   If interest focuses on changes in age-specific mortality, so that $\boldsymbol{\theta} = \boldsymbol{\mu}$, then the sensitivity formula expands, using the chain rule, to

$$\frac{d\boldsymbol{\eta}_1}{d\boldsymbol{\mu}^\mathsf{T}} = \left(\mathbf{N}^\mathsf{T} \otimes \boldsymbol{\eta}_1^\mathsf{T}\right)\frac{d\text{vec}\,\mathbf{U}}{d\boldsymbol{\mu}^\mathsf{T}} \tag{4.35}$$

This can be evaluated in several ways, depending on how the matrix $\mathbf{U}$ is written as a function of mortality. One approach is used in Sect. 4.4.3, and a somewhat more widely useful approach in Sect. 4.4.4.

   The results for Japan and India are shown in Fig. 4.2. Life expectancy is more sensitive to changes in mortality in Japan than in India; the (absolute value of) sensitivity decreases almost linearly with age in Japan, and slightly less linearly in India (Fig. 4.2). On the other hand, life expectancy is more elastic to changes in mortality in India, and less so in Japan.

4.4.3   Generalizing the Keyfitz-Pollard Formula

The Keyfitz-Pollard formula for the sensitivity of life expectancy to changes in mortality rate, given in Eq. (4.2), has a clear interpretation: the sensitivity to mortality at age a depends on the probability of survival to age a and the remaining life expectancy at age a. We are now in a position to generalize this to stage-classified matrix models.

First, we derive the matrix version of the Keyfitz-Pollard result, for the sensitivity of life expectancy of age class 1, which is

$$dE\left(\eta_1\right) = \left(\mathbf{e}_1^\mathsf{T} \otimes \mathbf{1}^\mathsf{T}\right) d\text{vec}\,\mathbf{N} \tag{4.36}$$

$$= \left(\mathbf{e}_1^\mathsf{T} \otimes \mathbf{1}^\mathsf{T}\right)\left(\mathbf{N}^\mathsf{T} \otimes \mathbf{N}\right) d\text{vec}\,\mathbf{U} \tag{4.37}$$

Consider a population with s age classes and let μ_i be the mortality rate and $p_i = \exp(-\mu_i)$ the survival probability for age class i. The matrix $\mathbf{U}$ is given by (4.13), which can be written

$$\mathbf{U} = \sum_{k=1}^{s-1}\left(\mathbf{e}_{k+1}\mathbf{e}_k^\mathsf{T}\right) p_k \tag{4.38}$$

where $\mathbf{e}_k$ is the unit vector, of length s, with a 1 in the kth position and zeros elsewhere. Differentiating $\mathbf{U}$ and applying the vec operator gives

$$d\text{vec}\,\mathbf{U} = -\sum_{k=1}^{s-1} (\mathbf{e}_k \otimes \mathbf{e}_{k+1})\; p_k\,(d\mu_k) \tag{4.39}$$

Substitute (4.39) into (4.37) and consider a perturbation of mortality at age a; the result is

$$\frac{dE(\eta_1)}{d\mu_a} = -\left(\mathbf{e}_1^\mathsf{T} \otimes \mathbf{1}^\mathsf{T}\right)\left(\mathbf{N}^\mathsf{T} \otimes \mathbf{N}\right)(\mathbf{e}_a \otimes \mathbf{e}_{a+1})\; p_a. \tag{4.40}$$

This simplifies to

$$\frac{dE(\eta_1)}{d\mu_a} = -\left(\mathbf{e}_1^\mathsf{T}\mathbf{N}^\mathsf{T}\mathbf{e}_a \otimes \mathbf{1}^\mathsf{T}\mathbf{N}\mathbf{e}_{a+1}\right) p_a \tag{4.41}$$

$$= -\underbrace{E\left(\nu_a\right) p_a}_{\text{survival}}\; \underbrace{E\left(\eta_{a+1}\right)}_{\text{expectancy}} \qquad \text{age-classified} \tag{4.42}$$

In an age-classified model, ν_a is either 0 or 1 (you cannot occupy a year of age for more than 1 year); hence the $E\left(\nu_a\right)$ is the probability of survival to age a. Thus we have a matrix version of the Keyfitz-Pollard result: the sensitivity of life expectancy is the probability of survival to age a times the probability of survival from a to $a + 1$, times the life expectancy at age $a + 1$.

Now apply the same approach to a stage-classified model, in which $\mathbf{U}$ can be written as the product of a diagonal matrix $\boldsymbol{\Sigma}$ with survival probabilities on the diagonal, and a stochastic matrix $\mathbf{G}$ giving the transition probabilities conditional on survival:

$$\mathbf{U} = \mathbf{G}\boldsymbol{\Sigma} \tag{4.43}$$

$$= \mathbf{G} \begin{pmatrix} p_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & p_s \end{pmatrix} \tag{4.44}$$

$$= \mathbf{G} \sum_{k=1}^{s} \left(\mathbf{e}_k \mathbf{e}_k^\mathsf{T} \right) p_k \tag{4.45}$$

Differentiating and applying the vec operator gives

$$d\mathrm{vec}\,\mathbf{U} = \sum_{k=1}^{s} (\mathbf{e}_k \otimes \mathbf{G}\mathbf{e}_k)\; p_k\,(d\mu_k) \tag{4.46}$$

Substitute this into (4.37) and focus on a change in mortality at stage a; the result is

$$\frac{dE(\eta_1)}{d\mu_a} = -\left(\mathbf{e}_1^\mathsf{T} \otimes \mathbf{1}^\mathsf{T} \right) \left(\mathbf{N}^\mathsf{T} \otimes \mathbf{N} \right) (\mathbf{e}_a \otimes \mathbf{G}\mathbf{e}_a)\; p_a \tag{4.47}$$

which simplifies to

$$\frac{dE(\eta_1)}{d\mu_a} = -\left(\mathbf{e}_1^\mathsf{T} \mathbf{N}^\mathsf{T} \mathbf{e}_a \otimes \mathbf{1}^\mathsf{T} \mathbf{N}\mathbf{G}\mathbf{e}_a \right) p_a \tag{4.48}$$

$$= -E(v_{a1})\, E\left(\boldsymbol{\eta}^\mathsf{T} \right) \mathbf{G}(:,a) p_a \tag{4.49}$$

$$= -\underbrace{E(v_{a1})}_{\text{occupancy}} \sum_{h=1}^{s} \underbrace{p_a g_{ha}}_{\text{transitions}} \underbrace{E(\eta_h)}_{\text{expectancy}} \qquad \text{stage-classified} \tag{4.50}$$

Equation (4.50) is the stage-classified version of Keyfitz-Pollard: the sensitivity of life expectancy to a change in mortality in stage j is the product of the expected time spent in stage j and the remaining life expectancy, calculated as an average of the life expectancy of all stages k, weighted by the probability of transition from j to k. This can be simplified further by noting that, for either age or stage-classified populations, $\mathbf{G}(:,a) p_a = \mathbf{U}(:,a)$, so that a completely general expression is

$$\frac{dE(\eta_1)}{d\mu_a} = -E(v_{a1})\, E\left(\boldsymbol{\eta}^\mathsf{T} \right) \mathbf{U}(:,a) \qquad \text{age- or stage-classified} \tag{4.51}$$

4.4.4 Sensitivity of the Variance of Longevity

The sensitivity of the variance in longevity is obtained by differentiating (4.23)

$$dV(\boldsymbol{\eta}) = d\boldsymbol{\eta}_2 - 2\left(\boldsymbol{\eta}_1 \circ d\boldsymbol{\eta}_1\right) \tag{4.52}$$

and applying the vec operator (using results from Chap. 2 on the vec of the Hadamard product), to obtain

$$dV(\boldsymbol{\eta}) = d\boldsymbol{\eta}_2 - 2\mathcal{D}(\boldsymbol{\eta}_1)d\boldsymbol{\eta}_1. \tag{4.53}$$

The derivative of $\boldsymbol{\eta}_1$ is already given by (4.33):

$$d\boldsymbol{\eta}_1 = \left(\mathbf{N}^\mathsf{T} \otimes \boldsymbol{\eta}_1^\mathsf{T}\right) d\mathrm{vec}\,\mathbf{U}. \tag{4.54}$$

The derivative of $\boldsymbol{\eta}_2$ is obtained by differentiating (4.22):

$$d\boldsymbol{\eta}_2^\mathsf{T} = 2\left(d\boldsymbol{\eta}_1^\mathsf{T}\right)\mathbf{N} + d\boldsymbol{\eta}_1^\mathsf{T}(d\mathbf{N}) - d\boldsymbol{\eta}_1^\mathsf{T} \tag{4.55}$$

Applying the vec operator to both sides and substituting (4.29) for $d\mathrm{vec}\,\mathbf{N}$ gives

$$d\boldsymbol{\eta}_2 = \left(2\mathbf{N}^\mathsf{T} - \mathbf{I}\right)d\boldsymbol{\eta}_1 + 2\left(\mathbf{N}^\mathsf{T} \otimes \boldsymbol{\eta}_1^\mathsf{T}\mathbf{N}\right)d\mathrm{vec}\,\mathbf{U} \tag{4.56}$$

Inserting (4.54) for $d\boldsymbol{\eta}_1$ and (4.56) for $d\boldsymbol{\eta}_2$ into (4.53) gives the sensitivity of the variance in remaining longevity, for any starting age or stage, to changes in $\mathbf{U}$. The sensitivity of longevity to mortality is obtained by differentiating $\mathbf{U}$ with respect to $\boldsymbol{\mu}$.

Derivatives of U The derivative of $\mathbf{U}$ to the mortality vector $\boldsymbol{\mu}$ are obtained as follows. For an age-classified model, define an age-advancement matrix

$$\mathbf{L} = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & [1] \end{pmatrix} \tag{4.57}$$

(show here for three age classes, with the optional open-ended last age class). This matrix will mask the entries of a matrix $\mathbf{1p}^\mathsf{T}$, that contains $\mathbf{p}$ in each row, to obtain

$$\mathbf{U} = \mathbf{L} \circ \left(\mathbf{1p}^\mathsf{T}\right) \tag{4.58}$$

Differentiating and applying the vec operator gives

$$d\mathbf{U} = \mathbf{L} \circ \left(\mathbf{1}\left(d\mathbf{p}^\mathsf{T}\right)\right) \tag{4.59}$$

$$d\mathrm{vec}\,\mathbf{U} = \mathcal{D}(\mathrm{vec}\,\mathbf{L})(\mathbf{I} \otimes \mathbf{1})d\mathbf{p}. \tag{4.60}$$

Since $\mathbf{p} = \exp(-\boldsymbol{\mu})$,

$$d\mathbf{p} = -\mathcal{D}(\mathbf{p})d\boldsymbol{\mu}, \tag{4.61}$$

and hence

$$d\text{vec}\,\mathbf{U} = -\mathcal{D}(\text{vec}\,\mathbf{L})(\mathbf{I} \otimes \mathbf{1})\mathcal{D}(\mathbf{p})d\boldsymbol{\mu} \qquad \text{age-classified} \tag{4.62}$$

For a stage-classified model, write $\mathbf{U} = \mathbf{G}\boldsymbol{\Sigma}$, as in (4.44) as

$$\mathbf{U} = \mathbf{G}\left[\mathbf{I} \circ \left(\mathbf{1}\mathbf{p}^{\mathsf{T}}\right)\right] \tag{4.63}$$

Differentiating and applying the vec operator, following the strategy of (4.60), gives

$$d\text{vec}\,\mathbf{U} = -(\mathbf{I} \otimes \mathbf{G})\mathcal{D}(\text{vec}\,\mathbf{I})(\mathbf{I} \otimes \mathbf{1})\mathcal{D}(\mathbf{p})d\boldsymbol{\mu} \qquad \text{stage-classified} \tag{4.64}$$

Substituting (4.62) and (4.64) into the expressions for $d\boldsymbol{\eta}_1$ and $d\boldsymbol{\eta}_2$, and substituting those into (4.53) gives the sensitivity of the variance in longevity to age- or stage-specific mortality. It is possible to carry out the substitutions and to arrive at a single (large) expression for $dV(\boldsymbol{\eta})$; see Chap. 5.

Figure 4.2d shows the sensitivity and elasticity of variance of longevity to changes in age-specific mortality. The variance is more sensitive to mortality changes in Japan than in India, and the sensitivities are highest at young ages. Both life tables have the property that sensitivities are positive at early ages ($\approx$0–20 for India, $\approx$0–80 for Japan) and then become negative. Before this age, reductions in mortality will reduce variance; after this age, reductions in mortality increase the variance. See Sect. 4.4.6 for more on this.

4.4.5   Sensitivity of the Distribution of Age at Death

The sensitivity of the distribution of age or stage at death is obtained by differentiating (4.25) and applying the vec operator,

$$d\text{vec}\,\mathbf{B} = \left(\mathbf{N}^{\mathsf{T}} \otimes \mathbf{I}\right)d\text{vec}\,\mathbf{M} + (\mathbf{I} \otimes \mathbf{M})d\text{vec}\,\mathbf{N}. \tag{4.65}$$

We already know $d\text{vec}\,\mathbf{N}$. To obtain $d\text{vec}\,\mathbf{M}$, note that when the absorbing states are defined in terms of stage at death

$$\mathbf{M} = \mathbf{I} - \mathcal{D}(\mathbf{p}) \tag{4.66}$$

and thus

$$d\text{vec}\,\mathbf{M} = -\mathcal{D}\,(\text{vec}\,\mathbf{I})\,(\mathbf{I} \otimes \mathbf{1})\,d\mathbf{p} \tag{4.67}$$

It is revealing to write the sensitivity of $\mathbf{B}$ to changes in mortality using the chain rule,

$$\frac{d\text{vec}\,\mathbf{B}}{d\boldsymbol{\mu}^\mathsf{T}} = \left(\mathbf{N}^\mathsf{T} \otimes \mathbf{I}\right)\frac{d\text{vec}\,\mathbf{M}}{d\mathbf{p}^\mathsf{T}}\frac{d\mathbf{p}}{d\boldsymbol{\mu}^\mathsf{T}} + (\mathbf{I} \otimes \mathbf{M})\frac{d\text{vec}\,\mathbf{N}}{d\text{vec}^\mathsf{T}\mathbf{U}}\frac{d\text{vec}\,\mathbf{U}}{d\mathbf{p}^\mathsf{T}}\frac{d\mathbf{p}}{d\boldsymbol{\mu}^\mathsf{T}} \tag{4.68}$$

and to recognize how many of the pieces we have already obtained.

The distribution of stage at death for individuals starting in stage j is given by column j of $\mathbf{B}$; i.e., $\boldsymbol{\psi}_j = \mathbf{B}(:, j)$. The sensitivity of ψ_j to changes in mortality is

$$\frac{d\boldsymbol{\psi}_j}{d\boldsymbol{\mu}^\mathsf{T}} = \left(\mathbf{e}_j \otimes \mathbf{I}\right)\frac{d\text{vec}\,\mathbf{B}}{d\boldsymbol{\mu}^\mathsf{T}} \tag{4.69}$$

for any age or stage j of interest.

4.4.6   Sensitivity of Life Disparity

To get the sensitivity of the vector $\boldsymbol{\eta}^\dagger$, differentiate and apply the vec operator to Eq. (4.27), which gives

$$d\boldsymbol{\eta}^\dagger = \mathbf{B}^\mathsf{T}d\boldsymbol{\eta}_1 + \left(\mathbf{I} \otimes \boldsymbol{\eta}_1^\mathsf{T}\right)d\text{vec}\,\mathbf{B}. \tag{4.70}$$

Evaluating this expression for the data on India and Japan, we see that the sensitivity of $\eta^\dagger$ shows a pattern similar to that of the sensitivity of $V(\boldsymbol{\eta})$ (Fig. 4.2), confirming that these indices are measuring similar aspects of disparity in longevity.

In particular, they show the existence of a critical age, before which reductions in mortality reduce disparity and after which they have the opposite effect. Zhang and Vaupel (2009) showed that this critical age, which they describe as separating "early" from "late" deaths is a general property of $\eta^\dagger$. Although the details depend on which index of disparity one uses, the existence of a critical age separating positive and negative sensitivities is also a property of other measures of variation in longevity (Van Raalte and Caswell 2013). Vaupel et al. (2011) have used the critical age to decompose historical changes in lifespan disparity into components due to early and late mortality.

4.5   A Time-Series LTRE Decomposition: Life Disparity

The LTRE decomposition analysis in Sect. 2.9 can be used to decompose time series such as these into their components. We apply it here to calculate the contributions, to a long trajectory of changes in $\boldsymbol{\eta}^{\dagger}$, of changes in early and late mortality.

Suppose that some demographic outcome $\boldsymbol{\xi}(t)$ (dimension $s \times 1$) is measured as a function of a parameter vector $\boldsymbol{\theta}$ (dimension $p \times 1$), at times $1, 2, \ldots T$. The changes in $\boldsymbol{\xi}(t)$ over time result from the changes in the parameters,

$$\Delta \boldsymbol{\xi}(t) = \boldsymbol{\xi}(t+1) - \boldsymbol{\xi}(t) \tag{4.71}$$

$$\Delta \boldsymbol{\theta}(t) = \boldsymbol{\theta}(t+1) - \boldsymbol{\theta}(t) \tag{4.72}$$

The decomposition analysis for such sequences was introduced as a "regression LTRE" method in the context of ecotoxicology and response to environmental factors (e.g., Caswell 1996; Knight et al. 2009). The same approach was introduced independently by Horiuchi et al. (2008) to decompose differences between two conditions by imagining a continuous path from one to the other.

The analysis starts by considering the change in $\boldsymbol{\xi}$ over time,

$$\frac{d\boldsymbol{\xi}(t)}{dt} = \frac{d\boldsymbol{\xi}(t)}{d\boldsymbol{\theta}^{\mathsf{T}}(t)} \frac{d\boldsymbol{\theta}(t)}{dt} \tag{4.73}$$

If the time series is evaluated at discrete times $t = 1, \ldots, T$, then to first order

$$\Delta \boldsymbol{\xi}(t) \approx \frac{d\boldsymbol{\xi}(t)}{d\boldsymbol{\theta}^{\mathsf{T}}(t)} \Delta \boldsymbol{\theta}(t) \qquad s \times 1 \tag{4.74}$$

The contributions to $\Delta \boldsymbol{\xi}(t)$ are displayed separately in a contribution matrix

$$\mathbf{C}(t) = \frac{d\boldsymbol{\xi}(t)}{d\boldsymbol{\theta}^{\mathsf{T}}(t)} \mathcal{D}[\Delta \boldsymbol{\theta}(t)] \qquad s \times p \tag{4.75}$$

the (i, j) entry of $\mathbf{C}(t)$ is the contribution of $\Delta \theta_j(t)$ to $\Delta \xi_i(t)$. The contributions additive over time, so the contributions of all the changes, integrated from t_1 to t_2, are given by the entries of

$$\mathbf{C}(t_1, t_2) = \sum_{t=t_1}^{t_2} \mathbf{C}(t) \tag{4.76}$$

Suppose the dependent variable is $\boldsymbol{\xi} = \boldsymbol{\eta}^{\dagger}$ and the parameter vector is $\boldsymbol{\theta} = \boldsymbol{\mu}$.

At each time and for each age, we aggregate the contributions from early and late mortality. Let $\mathbf{X}$ be an indicator matrix whose entries define whether a particular entry of $\mathbf{C}(t)$ is to be counted as early or late:

$$x_{ij} = \begin{cases} 1 & \theta_j \text{ contributes to } \Delta\xi_i \\ 0 & \text{otherwise} \end{cases} \qquad (4.77)$$

Then

$$\mathbf{c}(t) = (\mathbf{C}(t) \circ \mathbf{X})\, \mathbf{1} \qquad (4.78)$$

is a vector giving the contributions to the change in $\boldsymbol{\xi}$ from the parameters chosen in $\mathbf{X}$. Defining $\mathbf{X}_{\text{early}}$ and $\mathbf{X}_{\text{late}}$ gives changes at time t due to early and late mortality. The LTRE analysis is then

$$\mathbf{c}_{\text{early}}(t_1, t_2) = \sum_{t_1}^{t_2} \mathbf{c}_{\text{early}}(t) \qquad (4.79)$$

and similarly for $\mathbf{c}_{\text{late}}(t_1, t_2)$.

As an example, Fig. 4.3a, b shows a time series of life expectancy (increasing from about 40–80 years between 1800 and 2010) and life disparity for Swedish

(a) Life expectancy $\bar{\eta}_1$

(b) Life disparity $\eta_i^\dagger$

(c) Contributions at age 0

(d) Contributions at age 50
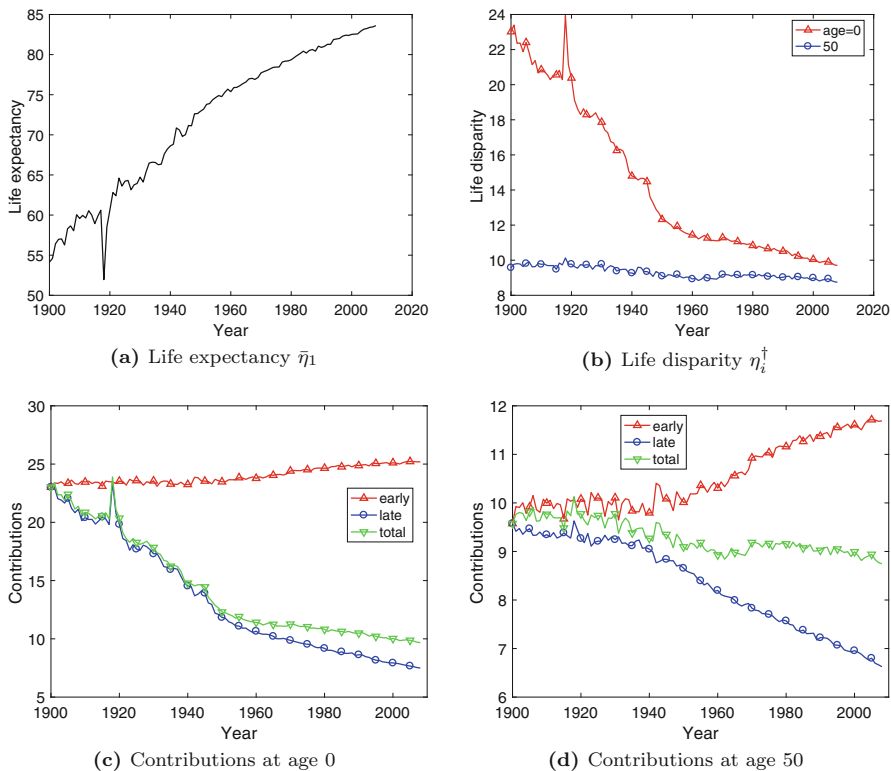
Fig. 4.3 **(a)** Historical trends in life expectancy at birth from 1800 to 2010. **(b)** Historical trends in life disparity (mean years of life lost due to mortality) for ages 0 and 50 years. **(c)** Contributions from early and late mortality improvement to the change in disparity at age 0. **(d)** The contributions for disparity at age 50. (Data for Swedish females, from the Human Mortality Database)

females, based on data from Human Mortality Database (2016). As in most developed countries, life disparity at birth dropped dramatically from 1850 to about 1950 (e.g., Edwards 2011; Vaupel et al. 2011). Declines at later ages were less dramatic, and remaining life disparity conditional on survival to age 50 has been almost flat (Engelman et al. 2014). How did changes in early and late mortality contribute to these patterns?

Figure 4.3c, d show the cumulative sums of the contributions $\mathbf{c}_{early}$ and $\mathbf{c}_{late}$, and their total, for ages 0 and 50. The decline in life disparity at birth was driven almost completely by improvements in early mortality, which completely overshadowed a small increase in disparity that was generated by improvements in late life mortality. The picture for remaining life disparity at age 50 is different: the contributions from changes in early and late life mortality almost completely cancel each other out. These patterns, looking at the details of a single time series, agree with the much more general exploration of multiple countries, using a different approach, by Vaupel et al. (2011).

The accuracy of the decomposition can be evaluated by comparing the time series calculated from the total contributions, as shown in Fig. 4.3c, d, with the observed series, as shown in Fig. 4.3b. The agreement is extremely close; the LTRE decomposition captures the end result of the historical changes from 1800 to 2010 with an error of less than 0.1%.

4.6 Conclusion

This chapter and Chap. 3 contain examples of different approaches to the sensitivity analysis, of population growth rate and longevity, respectively. The power and flexibility of matrix calculus methods is apparent: the models are not restricted to age- or stage-classification, the absorbing states may be a single category of death or some more diverse set, the demographic outcomes are not limited to expectations, and the independent variables, the parameters that are being perturbed, can be anything of interest. The only requirement is that a chain of functional dependence can be followed: the outcome $\boldsymbol{\xi}$ depends on $\mathbf{U}$, which depends on $\mathbf{p}$, which depends on $\boldsymbol{\mu}$, ... and so on. Mortality might depend on health status, which might depend on income level, which might depend on education, ..., and so on. The sensitivity of $\boldsymbol{\xi}$ to any of these parameters is a application of the chain rule.

Bibliography

Canudas Romo, V. 2003. Decomposition methods in demography. Population Studies, Rozenberg Publishers, Amsterdam, Netherlands.

Caswell, H. 1996. Analysis of life table response experiments. II. Alternative parameterizations for size- and stage-structured models. Ecological Modelling **88**:73–82.

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. 2009. Stage, age and individual stochasticity in demography. Oikos **118**:1763–1782.

Caswell, H. 2011. Beyond R_0: Demographic models for variability of lifetime reproductive output. PloS ONE **6**:e20809.

Caswell, H., and F. A. Kluge. 2015. Demography and the statistics of lifetime economic transfers under individual stochasticity. Demographic Research **32**:563–588.

Caswell, H., and V. Zarulli. 2018. Matrix methods in health demography: a new approach to the stochastic analysis of healthy longevity and DALYs. Population Health Metrics **16**:8.

Devleesschauwer, B., A. H. Havelaar, C. M. De Noordhout, J. A. Haagsma, N. Praet, P. Dorny, L. Duchateau, P. R. Torgerson, H. Van Oyen, and N. Speybroeck. 2014. DALY calculation in practice: a stepwise approach. International Journal of Public Health **59**:571–574.

Edwards, R. D. 2011. Changes in world inequality in length of life: 1970–2000. Population and Development Review **37**:499–528.

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.

Engelman, M., H. Caswell, and E. M. Agree. 2014. Why do lifespan variability trends for the young and old diverge? Demographic Research **30**:1367–1396.

Feichtinger, G. 1971. Stochastische Modelle Demographischer Prozesse. Lecture notes in operations research and mathematical systems, Springer-Verlag, Berlin, Germany.

GBD 2016 DALYs and HALE Collaborators. 2017. Global, regional, and national disability-adjusted life-years (DALYs) for 333 diseases and injuries and healthy life expectancy (HALE) for 195 countries and territories, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet **390**:1260–1344.

Hoem, J. M. 1969. Markov chain models in life insurance. Sonderdruck aus Blatter der Deutschen Gelellschaft fur Versicherungsmathematik **9**:91–107.

Horiuchi, S., J. R. Wilmoth, and S. D. Pletcher. 2008. A decomposition method based on a model of continuous change. Demography **45**:785–801.

Horvitz, C. C., and S. Tuljapurkar. 2008. Stage dynamics, period survival, and mortality plateaus. American Naturalist **172**:203–215.

Howard, R. A. 1960. Dynamic programming and Markov processes. Wiley, New York, New York.

Human Mortality Database. 2016. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). www.mortality.org URL www.mortality.org.

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

Irwin, J. 1949. The standard error of an estimate of expectation of life, with special reference to expectation of tumourless life in experiments with mice. The Journal of Hygiene **47**:188–189.

Kemeny, J. G., and J. L. Snell. 1976. Finite Markov Chains. Second edition. Undergraduate Texts in Mathematics, Springer-Verlag, New York, New York, USA.

Keyfitz, N. 1971. Linkages of intrinsic to age-specific rates. Journal of the American Statistical Association **66**:275–281.

Knight, T., H. Caswell, and S. Kalisz. 2009. Population growth rate of a common understory herb decreases non-linearly across a gradient of deer herbivory. Forest Ecology and Management **257**:1095–1103.

Pollard, J. H. 1982. The expectation of life and its relationship to mortality. Journal of the Institute of Actuaries **109**:225–240.

Steinsaltz, D., and S. N. Evans. 2004. Markov mortality models: implications of quasistationarity and varying initial distributions. Theoretical Population Biology **65**:319–337.

Tuljapurkar, S., and C. C. Horvitz. 2006. From stage to age in variable environments: life expectancy and survivorship. Ecology **87**:1497–1509.

van Daalen, S., and H. Caswell. 2015. Lifetime reproduction and the second demographic transition: Stochasticity and individual variation. Demographic Research **33**:561–588.

van Daalen, S. F., and H. Caswell. 2017. Lifetime reproductive output: individual stochasticity, variance, and sensitivity analysis. Theoretical Ecology **10**:355–374.

Van Raalte, A. A., and H. Caswell. 2013. Perturbation analysis of indices of lifespan variability. Demography **50**:1615–1640.

Vaupel, J. W. 1986. How change in age-specific mortality affects life expectancy. Population Studies **40**:147–157.

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.

Vaupel, J. W., Z. Zhang, and A. A. van Raalte. 2011. Life expectancy and disparity: an international comparison of life table data. BMJ Open **1**:e000128.

Wilson, E. B. 1938. The standard deviation of sampling for life expectancy. Journal of the American Statistical Association **33**:705–708.

Zhang, Z., and J. W. Vaupel. 2009. The age separating early deaths from late deaths. Demographic Research **20**:721–730.

Chapter 5
Individual Stochasticity and Implicit Age Dependence

5.1 Introduction

Demography is the study of the population consequences of the fates of individuals. As an individual organism develops through its life cycle it may increase in size, change its morphology, develop new physiological functions, exhibit new behaviors, or move to new locations. It may marry and divorce, become ill and recover, or change its employment status. It may change sex and/or change its reproductive status. These changes can be dramatic. This developmental process, and its attendant risks of death and opportunities for reproduction, determine the rates of birth and death that, in turn, determine population growth or decline.

Individuals are differentiated on the basis of age or, in general, life cycle stages. The movement of an individual through its life cycle is a random process, and although the eventual destination (death) is certain, the pathways taken to that destination are stochastic and will differ even between identical individuals; this is *individual stochasticity*. A stage-classified demographic model contains implicit age-specific information, which can be analyzed using Markov chain methods. The living stages in the life cycles are transient states in an absorbing Markov chain, in which death is an absorbing state.

This chapter presents Markov chain methods for computing the mean and variance of the lifetime number of visits to any transient state, the mean and variance of longevity, the net reproductive rate R_0, and the cohort generation time. It presents the matrix calculus methods needed to calculate the sensitivity and elasticity of all these indices to any life history parameters.

The Markov chain approach is then generalized to variable environments (deterministic environmental sequences, periodic environments, iid random environments, Markovian environments). Variable environments are analyzed using the vec-permutation method to create a model that classifies individuals jointly by the stage and environmental condition. Throughout, examples are presented using the North Atlantic right whale (*Eubaleana glacialis*) and an endangered prairie plant (*Lomatium bradshawii*) in a stochastic fire environment.

5.1.1   Age and Stage, Implicit and Explicit

The essence of demography is the connection between the fates of individual organisms and the dynamics of populations. There exist diverse mathematical frameworks in which this connection can be studied (Keyfitz 1967; Metz and Diekmann 1986; Nisbet and Gurney 1982; Caswell 1989; Tuljapurkar and Caswell 1997; Caswell et al. 1997; DeAngelis and Gross 1992; Ellner et al. 2016). Regardless of the type of equations used, demographic analysis must account for differences among individuals, and the ways in which those differences affect the vital rates.

Among the many ways that individuals may differ, age has long had a kind of conceptual priority. Age is universal in the sense that every organism becomes one minute older with the passage of one minute of time. Age is also often associated with predictable changes in the vital rates. However, in some organisms characteristics other than age provide more and better information about an individual. Ecologists recognized this long ago, and have developed demographic theory based on size, maturity, physiological condition, instar, spatial location, etc.—referred to in general as "stage-classified" demography. Human demographers, who were responsible for the classical age-classified theory, by no means deny the importance of other properties, such as employment, parity, or health status; see Land and Rogers (1982), Goldman (1994), Robine et al. (2003), and Willekens (2014) for a sample of the kinds of issues that arise.

Even when the demographic model is entirely stage-classified, however, age is still implicitly present. Individuals in a given stage may differ in age, and individuals of a given age may be found in many different stages, but each individual still becomes one unit of age older with the passage of each unit of time. Extracting this implicit age-dependent information makes it possible to calculate interesting age-specific properties, such as survivorship, longevity, life expectancy, generation time, and net reproductive rate (Cochran and Ellner 1992; Caswell 2001, 2006; Tuljapurkar and Horvitz 2006; Horvitz and Tuljapurkar 2008).[1]

---

[1]*Explicit* age and stage dependence is explored in Chap. 6; see also Caswell and Salguero-Gómez (2013) and Caswell et al. (2018).

In this chapter, I show how to calculate some of these implicit age-specific properties from any stage-classified model. The trick is to formulate the life cycle as a Markov chain, and to generalize the "life" cycle to include death as a stage. Because death is permanent, it is called an *absorbing state*, and the theory of absorbing Markov chains provides the starting point for our analysis (Feichtinger 1971; Caswell 2001).

A Markov chain is a stochastic model for the movement of a particle among a set of states (e.g., Kemeny and Snell 1976; Iosifescu 1980). The probability distribution of the next state of the particle may depend on the current state, but not on earlier states. In our context, a "particle" is an individual organism. The states correspond to the stages of the life cycle, plus death (or perhaps multiple types of death, for example deaths due to different causes). This structure is ideally suited to asking questions about individual stochasticity, because it accounts for all the possible pathways, and their probabilities, that an individual can follow through its life. I will focus on discrete-time models, but much of the theory can no doubt be generalized to continuous-time models.

The use of Markov chains in demographic analysis is not new. As far as I know, Feichtinger (1971, 1973) was the first to use discrete-time absorbing Markov chains in demography, paying particular attention to competing risks and multiple causes of death. At around the same time, Hoem (1969) applied continuous-time Markov chains in the analysis of insurance systems (with states such as "active," "disabled," and "dead"). Later, Cochran and Ellner (1992) independently proposed the use of Markov chains to generate age-classified statistics from stage-classified models, but minimized the use of matrix notation in their presentation. Influenced by Feichtinger's work, and relying heavily on Iosifescu's (1980) treatment of absorbing Markov chains, I extended the calculations using matrix notation (Caswell 2001; Keyfitz and Caswell 2005), introduced sensitivity analysis (Caswell 2006), and presented results for both time-invariant and time-varying models. At the same time, Tuljapurkar and Horvitz (2006) and Horvitz and Tuljapurkar (2008) developed the same approaches and presented a more extensive investigation of time variation.

5.1.2   *Individual Stochasticity and Heterogeneity*

Consider a newborn individual. As it develops through the stages of its life cycle, it may grow, shrink, mature, move, reproduce, and allocate resources among its biological processes. At each moment, it is exposed to various mortality risks. At each moment, it has some chance of reproducing. Because these processes are stochastic, the lives of any two individuals may differ. These random outcomes—this *individual stochasticity*—imply that the age-specific properties of an individual (say, longevity) are random variables—there is a distribution among individuals that should be characterized by its mean, moments, etc. (Caswell 2009).

It is critical to notice that the calculation of these moments explicitly assumes that every individual in a given stage experiences exactly the same rates and hazards. There is no heterogeneity among the individuals (or, at least, none that matters demographically), even though there is variation in their lifetime properties. Empirical studies of longevity or lifetime reproductive output find that the variation among individuals is usually large, but it is a mistake to jump to the conclusion that it is due to heterogeneity among individuals without first examining the variance that is inevitably created by individual stochasticity (e.g., Tuljapurkar et al. 2009; Steiner and Tuljapurkar 2012; Caswell 2011; Caswell and Kluge 2015; Caswell and Vindenes 2018; Hartemink et al. 2017; Hartemink and Caswell 2018; van Daalen and Caswell 2017).

5.1.3   Examples

The calculations will be demonstrated by means of two case studies. The first is a stage-classified model for the North Atlantic right whale (*Eubaleana glacialis*). Later, in Sect. 5.5.4, a stochastic matrix model for the threatened prairie plant *Lomatium bradshawii* will appear as part of a study of variable environments.

The North Atlantic right whale is a large, highly endangered baleen whale (Kraus and Rolland 2007). Once abundant in the north Atlantic, it was decimated by whaling, beginning as much as a thousand years ago (Reeves et al. 2007). By 1900 the eastern North Atlantic stock had been effectively eliminated, and the western North Atlantic stock hunted to near extinction. The population has recovered only slowly since receiving at least nominal protection in 1935, and now numbers only about 300 individuals. Right whales migrate along the Atlantic coast of North America, from summer feeding grounds in the Gulf of Maine and Bay of Fundy to winter calving grounds off the Southeastern U.S. They are killed by ship collisions and entanglement in fishing gear (Kraus et al. 2005), and may also be affected by pollution of coastal waters.

Individual right whales are photographically identifiable by scars and callosity patterns. Since 1980, the New England Aquarium has surveyed the population, accumulating a database of over 10,000 sightings (Crone and Kraus 1990). Treating the first year of identification of an individual as marking, and each year of resighting as a recapture, permits the use of mark-recapture statistics to estimate demographic parameters of this endangered population (Caswell et al. 1999; Fujiwara and Caswell 2001, 2002; Caswell and Fujiwara 2004).

Figure 5.1 shows a life cycle graph used by Caswell and Fujiwara (2004) as the basis of a stage-structured matrix population model for the right whale. The stages are calves, immature females, mature but non-reproductive females, mothers, and "resting" mothers (because of the long period of parental care and gestation, right whales do not reproduce in the year after giving birth). This life cycle is typical of large, long-lived monovular mammals and birds.
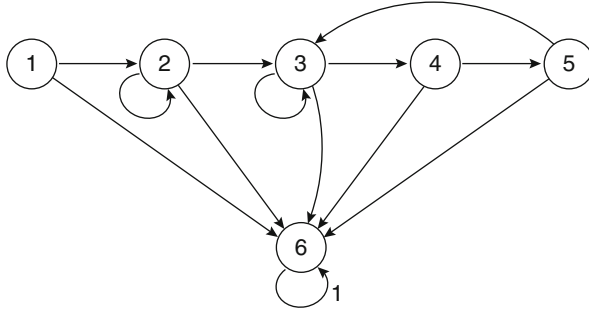
Fig. 5.1 Absorbing Markov chain transition graph for females of the North Atlantic right whale (*Eubalaena glacialis*). Projection interval is 1 year. Stages: 1 = calf, 2 = immature, 3 = mature, 4 = mother, 5 = post-breeding female, 6 = death. See Caswell and Fujiwara (2004) for explanation and parameter estimates

The model is parameterized in terms of survival probabilities $\sigma_1, \ldots, \sigma_5$, the probability of maturation γ_2, and the birth probability γ_3. The projection matrix is

$$\mathbf{A} = \begin{pmatrix} 0 & 0 & F & 0 & 0 \\ \sigma_1 & \sigma_2(1-\gamma_2) & 0 & 0 & 0 \\ 0 & \sigma_2\gamma_2 & \sigma_3(1-\gamma_3) & 0 & \sigma_5 \\ 0 & 0 & \sigma_3\gamma_3 & 0 & 0 \\ 0 & 0 & 0 & \sigma_4 & 0 \end{pmatrix} \tag{5.1}$$

The fertility term in the (1, 3) position is $F = 0.5\sigma_3\gamma_3\sqrt{\sigma_4}$, accounting for the sex ratio, the survival of mature females, their probability of giving birth if they survive, and the effect of survival of the mother on survival of the calf. For reasons related to parameter estimation, σ_5 is constrained to equal σ_3.

5.2   Markov Chains

The familiar life cycle graph (e.g., Fig. 5.1) corresponds to a projection matrix **A**, in which a_{ij} gives the per-capita production of stage i individuals at $t + 1$ by a stage j individual at t. This production may occur by the transition of an individual from stage j to stage i, or by the production of one or more new individuals (by reproduction, fragmentation, etc.). So, we partition **A** into a matrix **U** describing transition probabilities of extant individuals and a matrix **F** describing the production of new individuals

$$\mathbf{A} = \mathbf{U} + \mathbf{F} \tag{5.2}$$

The column sums of **U** are all less than or equal to 1. Because individuals eventually die and pass out of the stages contained in **U**, those stages are called transient states.

5.2.1   An Absorbing Markov Chain

If we include death explicitly (Fig. 5.1) and remove the arcs representing reproduction, we obtain the graph corresponding to the transition matrix for an absorbing Markov chain

$$\mathbf{P} = \left(\begin{array}{c|c} \mathbf{U} & 0 \\ \hline \mathbf{m} & 1 \end{array} \right) \tag{5.3}$$

The element m_j of the vector $\mathbf{m}$ is the probability of mortality of an individual in stage j. Death is an absorbing state. I will assume that at least one absorbing state is accessible from any transient state in $\mathbf{U}$, and that the spectral radius of $\mathbf{U}$ is strictly less than 1. This guarantees that, with probability 1, every individual ends up in the absorbing state.

The right whale Fujiwara estimated $\mathbf{U}$ by applying multi-stage mark-recapture methods to the photographic identification catalog. Although the best model, out of a large number evaluated, included significant time variation in survival and birth rates, here I will analyze a single matrix obtained from a time-invariant model. The complete transient matrix $\mathbf{U}$ and the fertility matrix $\mathbf{F}$ are

$$\mathbf{U} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0.90 & 0.85 & 0 & 0 & 0 \\ 0 & 0.12 & 0.71 & 0 & 1.00 \\ 0 & 0 & 0.29 & 0 & 0 \\ 0 & 0 & 0 & 0.85 & 0 \end{pmatrix} \tag{5.4}$$

$$\mathbf{F} = \begin{pmatrix} 0 & 0 & 0.13 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \tag{5.5}$$

5.2.2   Occupancy Times and the Fundamental Matrix

As the syllogism asserts, all men are mortal; absorbtion is certain. Our question is, how long does absorbtion take and what happens en route? From a demographic perspective, this is asking about the lifespan of an individual and the events that happen during that lifetime. The key to these questions is the *fundamental matrix* of the absorbing Markov chain. Consider an individual presently in transient state j. As time passes, it will visit other transient states, repeating some, skipping others,

until it eventually dies. Let v_{ij} denote the number of visits to, or the occupancy time in, transient state i that our individual, starting in transient state j, makes before being absorbed. The v_{ij} are random variables, reflecting individual stochasticity.

The entries of the matrix $\mathbf{U}$ give the probabilities of visiting each of the transient states after one time step. The entries of $\mathbf{U}^2$ give the probabilities of visiting each of the transient states after two time steps. Adding the powers of $\mathbf{U}$ gives the expected number of visits to each transient state, over a lifetime, in a matrix $\mathbf{N}$; i.e.,

$$\mathbf{N} = \left(E(v_{ij}) \right)$$

$$= \sum_{t=0}^{\infty} \mathbf{U}^t$$

$$= (\mathbf{I} - \mathbf{U})^{-1}. \tag{5.6}$$

The right whale The fundamental matrix for the right whale is calculated from (5.6) to be

$$\mathbf{N} = \begin{pmatrix} 1.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 5.88 & 6.52 & 0.00 & 0.00 & 0.00 \\ 16.35 & 18.11 & 22.94 & 19.49 & 22.94 \\ \mathbf{4.74} & 5.25 & 6.65 & 6.65 & 6.65 \\ 4.02 & 4.46 & 5.65 & 5.65 & 6.65 \end{pmatrix}. \tag{5.7}$$

The first column corresponds to calves. On average, a calf will spend 1 year as a calf, 5.9 years as a juvenile, 16.3 years as a mature but non-breeding female, etc. Row 4 of $\mathbf{N}$ is of particular interest. Stage 4 represents mothers, so n_{4j} is the expected number of reproductive events that a female in stage j will experience during her remaining lifetime. Based on this model, a newborn calf could expect to give birth $n_{41} = 4.74$ times. A mature female could expect to give birth $n_{43} = 6.65$ times; the difference reflects the likelihood of mortality between birth and maturity.[2]

We would like to know how the entries of $\mathbf{N}$ vary in response to changes in the vital rates. To accomplish this, we need matrix calculus, which is the topic of the next section.

---

[2]Note that $n_{43} = n_{44} = n_{45} = n_{55}$ and $n_{53} = n_{54}$. This seems to be due to the fact, specific to these data, that the survival probability of stages 3 and 5 is indistinguishable from 1.0, and influences the results below.

5.2.3 Sensitivity of the Fundamental Matrix

Let us apply matrix calculus to find the sensitivity of the fundamental matrix $\mathbf{N}$ (Caswell 2006). This result will appear in the sensitivity analysis of most other demographic quantities. Let $\boldsymbol{\theta}$ be a vector of parameters (of dimension $p \times 1$) on which the entries of the transition matrix $\mathbf{U}$ depend. The fundamental matrix satisfies

$$\mathbf{I} = \mathbf{N}\mathbf{N}^{-1}. \tag{5.8}$$

Differentiating both sides gives

$$\mathbf{0} = (d\mathbf{N})\mathbf{N}^{-1} + \mathbf{N}\left(d\mathbf{N}^{-1}\right). \tag{5.9}$$

Applying the vec operator and Roth's theorem to both sides gives

$$\operatorname{vec}\mathbf{0} = \left[\left(\mathbf{N}^{-1}\right)^{\mathsf{T}} \otimes \mathbf{I}_s\right] d\operatorname{vec}\mathbf{N} + (\mathbf{I}_s \otimes \mathbf{N})\, d\operatorname{vec}\mathbf{N}^{-1} \tag{5.10}$$

Solving for $d\operatorname{vec}\mathbf{N}$ gives

$$d\operatorname{vec}\mathbf{N} = \left[\left(\mathbf{N}^{-1}\right)^{\mathsf{T}} \otimes \mathbf{I}_s\right]^{-1} (\mathbf{I}_s \otimes \mathbf{N})\, d\operatorname{vec}\mathbf{U} \tag{5.11}$$

To simplify this, it helps to know two facts about the Kronecker product:

$$(\mathbf{A} \otimes \mathbf{B})^{-1} = \mathbf{A}^{-1} \otimes \mathbf{B}^{-1} \tag{5.12}$$

$$(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = (\mathbf{A}\mathbf{C} \otimes \mathbf{B}\mathbf{D}) \tag{5.13}$$

provided that the sizes of the matrices permit the indicated operations. Thus $d\operatorname{vec}\mathbf{N}$ in (5.11) simplifies to

$$d\operatorname{vec}\mathbf{N} = \left(\mathbf{N}^{\mathsf{T}} \otimes \mathbf{N}\right) d\operatorname{vec}\mathbf{U} \tag{5.14}$$

The identification theorem (2.47) implies

$$\frac{d\operatorname{vec}\mathbf{N}}{d\operatorname{vec}^{\mathsf{T}}\mathbf{U}} = \mathbf{N}^{\mathsf{T}} \otimes \mathbf{N} \tag{5.15}$$

and the chain rule permits us to write

$$\frac{d\operatorname{vec}\mathbf{N}}{d\boldsymbol{\theta}^{\mathsf{T}}} = \left(\mathbf{N}^{\mathsf{T}} \otimes \mathbf{N}\right) \frac{d\operatorname{vec}\mathbf{U}}{d\boldsymbol{\theta}^{\mathsf{T}}} \tag{5.16}$$
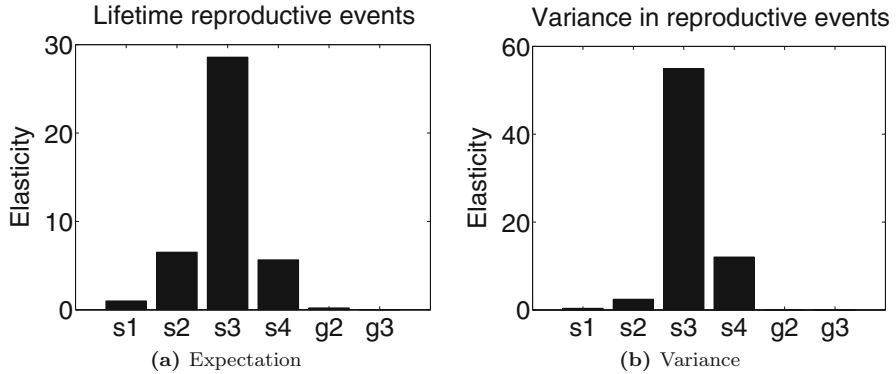
Fig. 5.2 Elasticity, to each of the vital rates, of reproductive outcomes in the right whale. (**a**) The elasticity of the expected lifetime number of reproductive events ($E(\nu_{41})$). (**b**) The elasticity of the variance in the lifetime number of reproductive events, $V(\nu_{41})$. Vital rates: s_1–s_4 are survival probabilities ($s_5 = s_3$ by assumption in this model); g_2 is the probability of maturation, and g_3 is the probability of reproduction

The left-hand side of (5.16) is a matrix, of dimension $s^2 \times p$, containing the sensitivity of every entry of $\mathbf{N}$ to every parameter in $\boldsymbol{\theta}$. The matrix $d\text{vec}\,\mathbf{U}/d\boldsymbol{\theta}^{\mathsf{T}}$ is an $s^2 \times p$ matrix containing the sensitivities of all the elements of $\mathbf{U}$ to all the elements of $\boldsymbol{\theta}$. From (2.55), the elasticity of the fundamental matrix is given by

$$\frac{\epsilon \text{vec}\,\mathbf{N}}{\epsilon \boldsymbol{\theta}^{\mathsf{T}}} = \mathcal{D}\,(\text{vec}\,\mathbf{N})^{-1}\,\frac{d\text{vec}\,\mathbf{N}}{d\boldsymbol{\theta}^{\mathsf{T}}}\,\mathcal{D}\,(\boldsymbol{\theta}) \qquad (5.17)$$

The right whale As an example, we use (5.16) and (5.17) to calculate the elasticity of the expected lifetime number of reproductive events, $E(\nu_{41}) = n_{41}$, with respect to the survival probabilities $\sigma_1, \ldots, \sigma_4$, the maturation probability γ_2, and the breeding probability γ_3. Figure 5.2 shows that the number of breeding events is most elastic to mature female survival (σ_3), and less so to the survival of mature females or mothers (σ_2 and σ_4). Changes in the probability of giving birth, γ_3, have, remarkably enough, no impact on the expected number of reproductive events.

The elasticity of n_{41} to σ_3 (survival of mature females) is approximately 30. This implies that a 1% increase in σ_3 would produce about a 30% increase in the expected number of reproductive events.

5.3 From Stage to Age

The fundamental matrix summarizes the age-specific information implicit in the transient matrix $\mathbf{U}$, even if the model is stage-classified and age does not appear explicitly. We now extend this, to explore a series of age-specific demographic

indices and their sensitivity analyses. Some are well known (R_0, generation time), others little explored (variance in longevity, for example). They can, however, all be easily calculated from any stage-classified model.

5.3.1   Variance in Occupancy Time

The occupancy time in any transient state is a random variable; the fundamental matrix $\mathbf{N}$ gives its mean. Some individuals will visit that state more often, some less often, some not at all. This basic property of individual stochasticity can be described by the variance of v_{ij}. Iosifescu (1980), Theorem 3.1 gives a formula for all the moments of the v_{ij}; from this we can calculate the matrix of variances

$$\mathbf{V} = \big(V(v_{ij}) \big) = \big(2\mathbf{N}_{\mathrm{dg}} - \mathbf{I} \big) \mathbf{N} - \mathbf{N} \circ \mathbf{N} \tag{5.18}$$

(Caswell 2006) where $\circ$ denotes the Hadamard, or element-by-element, product and $\mathbf{N}_{\mathrm{dg}}$ is a matrix with the diagonal elements of $\mathbf{N}$ on its diagonal and zeros elsewhere. The standard deviations of the occuancy times are the square roots of the elements of $\mathbf{V}$.

The right whale For the right whale, the matrix of variances calculated from (5.18) is

$$\mathbf{V} = \begin{pmatrix} 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 36.18 & 35.95 & 0.00 & 0.00 & 0.00 \\ 466.44 & 484.80 & 503.32 & 494.86 & 503.32 \\ 35.80 & 36.98 & 37.54 & 37.54 & 37.54 \\ 33.28 & 34.94 & 37.54 & 37.54 & 37.54 \end{pmatrix}, \tag{5.19}$$

and the corresponding standard deviations are

$$\big(SD(v_{ij}) \big) = \begin{pmatrix} 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 6.02 & 6.00 & 0.00 & 0.00 & 0.00 \\ 21.60 & 22.02 & 22.43 & 22.25 & 22.43 \\ 5.98 & 6.08 & 6.13 & 6.13 & 6.13 \\ 5.77 & 5.91 & 6.13 & 6.13 & 6.13 \end{pmatrix}. \tag{5.20}$$

The variance in the v_{ij} is the result of luck, not heterogeneity. That is, it is the variance among a group of individuals all experiencing exactly the same stage-specific transition and mortality probabilities in $\mathbf{U}$. As such, it can provide a null model for studies of heterogeneity in quantities such as the number of reproductive events. This idea has been explored independently, and in more detail, by Tuljapurkar and colleagues (Tuljapurkar et al. 2009; Steiner and Tuljapurkar 2012).

The sensitivity of the variance is derived in Appendix A.1 as

$$
\frac{d\operatorname{vec}\mathbf{V}}{d\boldsymbol{\theta}^{\mathsf{T}}} = \left[2\left(\mathbf{N}^{\mathsf{T}} \otimes \mathbf{I}_s\right)\mathcal{D}\left(\operatorname{vec}\mathbf{I}_s\right) + 2\left(\mathbf{I}_s \otimes \mathbf{N}_{\mathrm{dg}}\right) \right.
$$

$$
\left. -\mathbf{I}_{s^2} - 2\mathcal{D}\left(\operatorname{vec}\mathbf{N}\right) \right] \frac{d\operatorname{vec}\mathbf{N}}{d\boldsymbol{\theta}^{\mathsf{T}}} \tag{5.21}
$$

Elasticities of $\mathbf{V}$ are calculated using (2.55).

Hint Before looking at Appendix A.1, to derive (5.21), write $\mathbf{N}_{\mathrm{dg}} = \mathbf{I} \circ \mathbf{N}$, differentiate (5.18), and use the fact that $\operatorname{vec}(\mathbf{A} \circ \mathbf{B}) = \mathcal{D}(\operatorname{vec}\mathbf{A})\operatorname{vec}\mathbf{B} = \mathcal{D}(\operatorname{vec}\mathbf{B})\operatorname{vec}\mathbf{A}$.

The right whale The elasticities of $V(\nu_{41})$, calculated from (5.21) and (5.17), are shown in Fig. 5.2b. They are roughly proportional to the elasticities of $E(\nu_{41})$; that is, the vital rates that have large effects on the expected number of reproductive events also have large effects on the variance.

5.3.2   Longevity and Life Expectancy

Longevity is an important demographic characteristic (Carey 2003). Mean longevity, or life expectancy, it is one of the most widely reported demographic statistics, used to compare populations, species, countries, regions, historical periods, etc., and to examine the effects of evolutionary, management, medical, and social processes. The longevity of an individual is the sum of the time spent in all of the transient states before final absorption. Let the random variable η_j denote the longevity of an individual currently in stage j. Then

$$
\eta_j = \sum_i \nu_{ij}. \tag{5.22}
$$

A vector $E(\boldsymbol{\eta})$ of expected longevities, or life expectancies, is obtained by summing the columns of $\mathbf{N}$:

$$
E(\boldsymbol{\eta}^{\mathsf{T}}) = \mathbf{1}^{\mathsf{T}}\mathbf{N} \tag{5.23}
$$

where $\mathbf{1}$ is a vector of ones. Often, life expectancy at birth is of primary interest. If stages are numbered so that birth corresponds to stage 1, then life expectancy at birth is

$$
E(\eta_1) = \mathbf{1}^{\mathsf{T}}\mathbf{N}\mathbf{e}_1 \tag{5.24}
$$

where $\mathbf{e}_1$ is a vector with 1 in the first entry and zeros elsewhere.

The sensitivity of life expectancy in age-classified models has been studied by Pollard (1982) and Keyfitz (1971); see Keyfitz and Caswell (2005, Section 4.3), Vaupel (1986), and Vaupel and Canudas Romo (2003).

For more general stage-classified models, the sensitivity of $E(\boldsymbol{\eta})$ is (Caswell 2006)

$$\frac{dE(\boldsymbol{\eta})}{d\boldsymbol{\theta}^\mathsf{T}} = \left(\mathbf{I}_s \otimes \mathbf{1}^\mathsf{T}\right)\left(\mathbf{N}^\mathsf{T} \otimes \mathbf{N}\right)\frac{d\,\mathrm{vec}\,\mathbf{U}}{d\boldsymbol{\theta}^\mathsf{T}} \qquad (5.25)$$

Hint To obtain (5.25), differentiate both sides of (5.23), apply the vec operator, and use (5.16) for the derivative of $\mathbf{N}$. See Appendix A.2 for the derivation.

The right whale For the right whale, the vector of life expectancies is

$$E(\boldsymbol{\eta}^\mathsf{T}) = \left(32.0\ 34.3\ 35.2\ 31.8\ 36.2\right) \qquad (5.26)$$

Because mortality rates vary relatively little among stages, the life expectancies of the stages differ by only about 15%. Thus life expectancy for a calf implied by these data was 32 years. The elasticities of life expectancy to the vital rates are shown in Fig. 5.3. Life expectancy is most elastic to mature female survival σ_3, and less so to σ_2 and σ_3. This partly reflects the longer amount of time spent as a mature female, compared to an immature female or mother; see (5.7). The elasticity to the birth rate γ_3 is negative, because of the reduced survival of mothers. A 1% increase in γ_3 will lead to a 0.51% *decrease* in life expectancy. This is one possible measure of the cost of reproduction.
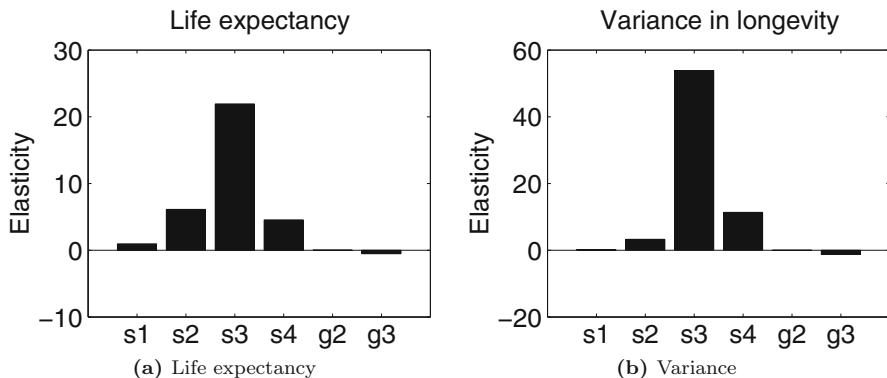
Fig. 5.3 Elasticities of longevity for the right whale. (**a**) The elasticity, to each of the vital rates, of life expectancy for a female right whale calf. (**b**) The elasticity of the variance in longevity for a female right whale calf. Parameters as in Fig. 5.2

5.3.3 Variance in Longevity

Like the occupancy time in a transient state, longevity is a random variable, the variability of which is a measure of individual stochasticity. Individuals differ in longevity depending on the pathways taken from birth to death. This variance has been explored by human demographers, using life table methods, as one way of studying the inequality in life span generated by a given mortality schedule, and how that inequality has changed over time (e.g., Wilmoth and Horiuchi 1999; Shkolnikov et al. 2003; Edwards and Tuljapurkar 2005; Van Raalte and Caswell 2013).

The variance of the time to absorbtion is

$$V(\boldsymbol{\eta}^\mathsf{T}) = \mathbf{1}^\mathsf{T}\mathbf{N}\,(2\mathbf{N} - \mathbf{I}) - E\left(\boldsymbol{\eta}^\mathsf{T}\right) \circ E\left(\boldsymbol{\eta}^\mathsf{T}\right). \tag{5.27}$$

(Caswell 2006; Iosifescu 1980).

The sensitivity of the variance in longevity is

$$\frac{dV(\boldsymbol{\eta})}{d\boldsymbol{\theta}^\mathsf{T}} = \left[2\left(\mathbf{N}^\mathsf{T} \otimes \mathbf{1}^\mathsf{T}\right) + 2\left(\mathbf{I}_s \otimes \mathbf{1}^\mathsf{T}\mathbf{N}\right) \right.$$

$$\left. - \left(\mathbf{I}_s \otimes \mathbf{1}^\mathsf{T}\right) - 2\mathcal{D}\left(E(\boldsymbol{\eta})\right)\left(\mathbf{I}_s \otimes \mathbf{1}^\mathsf{T}\right) \right]\left(\mathbf{N}^\mathsf{T} \otimes \mathbf{N}\right)\frac{d\mathbf{U}}{d\boldsymbol{\theta}^\mathsf{T}} \tag{5.28}$$

The first entry of (5.28) is the sensitivity of the variance in longevity starting in stage 1.

Hint To derive (5.28), differentiate (5.27) and apply the vec operator and Roth's theorem to each term, using (5.25) for the derivative of $E(\boldsymbol{\eta})$. See Sect. A.3 for details.

The right whale For the right whale, the variance and standard deviation of longevity are given by

$$V(\boldsymbol{\eta})^\mathsf{T} = \begin{pmatrix} 1157 & 1167 & 1172 & 1163 & 1172 \end{pmatrix} \tag{5.29}$$

$$SD(\boldsymbol{\eta})^\mathsf{T} = \begin{pmatrix} 34.0 & 34.2 & 34.2 & 34.1 & 34.2 \end{pmatrix} \tag{5.30}$$

The life expectancy at birth of 32 years has a standard deviation of about 34 years. Note that this result implies a very long positive tail of longevity. The interpretation of this result is tricky; I will return to it in Sect. 5.7.

The elasticities of the variance of longevity of a calf are shown in Fig. 5.3b. The variance in longevity is increased by increases in σ_3, less so by increases in σ_2 and σ_4. The pattern of the elasticities is strikingly similar to that of the elasticities of $E(\boldsymbol{\eta})$.

5.3.4 Cohort Generation Time

Generation time measures the typical age at which offspring are produced, or the age at which the typical offspring is produced. It appears in the IUCN criteria for classifying threatened species (IUCN Species Survival Commission 2001) as well as in various evolutionary considerations. There are several definitions of generation time (Coale 1972); here we will examine the cohort generation time, defined as the mean age of production of offspring in a cohort of newborn individuals. From the definition it is clear why calculation of generation time is a problem in stage-classified models, in which the age of parents does not appear. Moreover, in stage-classified models, individuals may be born into several stages (e.g., cleisthogamous vs. chasmogamous seeds; LeCorff and Horvitz 2005), each with a different subsequent pattern of development, survival, and fertility. There could be a different generation time for each type of offspring, and if individuals may produce more than one type of offspring, the average age at which they are produced could differ from one kind of offspring to another.

Thus, we expect to have a generation time that measures the mean age of production of offspring of type i by an individual born in stage j. Write this as a vector $\boldsymbol{\mu}^{(j)}$. Then it can be shown (Sect. A.5) that

$$\boldsymbol{\mu}^{(j)} = \mathcal{D}\left(\mathbf{FNe}_j\right)^{-1} \mathbf{FNUNe}_j \tag{5.31}$$

The sensitivity of $\boldsymbol{\mu}^{(j)}$ is obtained by a methodical application of matrix calculus to (5.31). To simplify notation, define

$$\mathbf{X} = \mathcal{D}\left(\mathbf{FNe}_j\right) \tag{5.32}$$

$$\mathbf{r} = \mathbf{FNUNe}_j \tag{5.33}$$

The resulting sensitivity of $\boldsymbol{\mu}^{(j)}$ is

$$\frac{d\boldsymbol{\mu}^{(j)}}{d\boldsymbol{\theta}^{\mathsf{T}}} = -\left(\mathbf{r}^{\mathsf{T}} \otimes \mathbf{I}\right)\left(\mathbf{X}^{-1} \otimes \mathbf{X}^{-1}\right)\mathcal{D}\left(\operatorname{vec}\mathbf{I}\right)$$

$$\times \left[\left(\mathbf{1e}_j^{\mathsf{T}}\mathbf{N}^{\mathsf{T}} \otimes \mathbf{I}\right)\frac{d\operatorname{vec}\mathbf{F}}{d\boldsymbol{\theta}^{\mathsf{T}}} + \left(\mathbf{1e}_j \otimes \mathbf{F}\right)\frac{d\operatorname{vec}\mathbf{N}}{d\boldsymbol{\theta}^{\mathsf{T}}}\right]$$

$$+ \left\{\left[\left(\mathbf{NUNe}_j\right)^{\mathsf{T}} \otimes \mathbf{I}\right]\frac{d\operatorname{vec}\mathbf{F}}{d\boldsymbol{\theta}^{\mathsf{T}}} + \left[\left(\mathbf{UNe}_j\right)^{\mathsf{T}} \otimes \mathbf{F}\right]\frac{d\operatorname{vec}\mathbf{N}}{d\boldsymbol{\theta}^{\mathsf{T}}}\right.$$

$$\left. + \left[\left(\mathbf{Ne}_j\right)^{\mathsf{T}} \otimes \mathbf{FN}\right]\frac{d\operatorname{vec}\mathbf{U}}{d\boldsymbol{\theta}^{\mathsf{T}}} + \left[\mathbf{e}_j^{\mathsf{T}} \otimes \mathbf{FNU}\right]\frac{d\operatorname{vec}\mathbf{N}}{d\boldsymbol{\theta}^{\mathsf{T}}}\right\} \tag{5.34}$$

Fig. 5.4 The elasticity, to each of the vital rates, of the cohort generation time for a newborn calf right whale. Parameters as in Fig. 5.2

Generation time
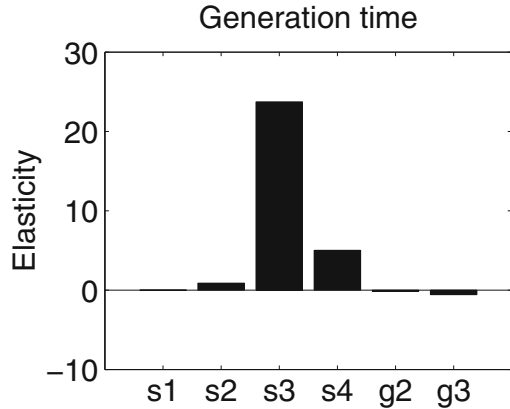
Hint To derive (5.34), it helps to note that, for any vector $\mathbf{z}$, one can write $\mathcal{D}(\mathbf{z}) = \mathbf{I} \circ \mathbf{z}\mathbf{1}^{\mathsf{T}}$. Apply this to $\mathbf{X}$, differentiate all the terms in $\mu^{(j)}$, and apply the vec operator. With any luck, you will come out to this answer. See Sect. A.5.1 for derivation.

The right whale The elasticities of the generation time $\boldsymbol{\mu}^{(1)}$ of a calf are shown in Fig. 5.4. Changes in early survival (σ_1 and σ_2) have little effect. Adult survival σ_3 and, to a lesser extent, σ_4 increase the generation time by extending the reproductive lifespan. The maturation probability γ_2 and the birth probability γ_3 have negative effects on generation time, because they speed up reproduction.

5.4 The Net Reproductive Rate

In age-classified demography, the net reproductive rate R_0 measures lifetime reproductive output. It also appears in epidemiology, where it measures the potential of a disease to spread (e.g., Diekmann et al. 1990; van den Driessche and Watmough 2002). The classical net reproductive rate satisfies three conditions:

C_1: R_0 measures the expected lifetime production of offspring.
C_2: R_0 measures the rate of increase per generation (in contrast to the rate of increase per unit of time, which is given by λ or r).
C_3: R_0 is an indicator function for population persistence. If $R_0 > 1$ then an individual will, on average, produce more than enough offspring to replace itself, the next generation will be larger than the present generation, and the population will grow. If $R_0 < 1$, each generation is smaller than the one before, and the population will decline to extinction.

In classical demography (Lotka 1939; Rhodes 1940),

$$R_0 = \int_0^{\infty} \ell(x)m(x)dx \tag{5.35}$$

where $\ell(x)$ is survivorship to age x and $m(x)$ is the maternity function. It is not difficult to show that R_0 defined in this way satisfies conditions C_1, C_2, and C_3.

In stage-classified models, however, the calculation of R_0 must account for the multiple pathways that an individual may follow through the life cycle, and the production of multiple kinds of offspring along each of these pathways. Rogers (1974; see also Lebreton 1996) considered R_0 in the context of an age-classified population distributed across a set of spatial regions. However, these calculations assume that age-specific survival and fertility schedules are available for each region. A more general solution was provided by Cushing and Zhou (1994) for stage-classified populations with no age-specific information. Their analysis produces an index that satisfies as many as possible of the conditions C_1, C_2, and C_3. de Camino-Beck and Lewis (2007, 2008) have derived graph-theoretic ways to calculate R_0.

Consider an initial cohort at $t = 0$ with structure $\mathbf{x}_0$, and call this the first generation. This cohort will produce offspring according to $\mathbf{Fx}_0$. The survivors of the cohort at $t = 1$ will produce offspring according to $\mathbf{FUx}_0$. The survivors at $t = 2$ will produce offspring $\mathbf{FU}^2\mathbf{x}_0$, and so on. The second generation is composed of all the offspring of the first generation, obtained by summing over the lifetime of the cohort

$$\mathbf{x}(1) = \left(\mathbf{F} \sum_{i=0}^{\infty} \mathbf{U}^i \right) \mathbf{x}_0$$

$$= (\mathbf{FN})\,\mathbf{x}_0 \tag{5.36}$$

Iterating this process leads to a model for the growth from one generation to the next

$$\mathbf{x}(k+1) = \mathbf{FNx}(k) \tag{5.37}$$

Cushing and Zhou (1994) define R_0 as the per-generation growth rate, given by the dominant eigenvalue ρ of $\mathbf{FN}$,

$$R_0 = \rho[\mathbf{FN}] \tag{5.38}$$

Thus the Cushing-Zhou measure of R_0 clearly satisfies condition C_2. Cushing and Zhou (1994) also prove (their Theorem 3) that R_0 defined in this way is less than, equal to, or greater than 1 if and only if λ is less than, equal to, or greater than one, respectively, thus satisfying condition C_3.

The relation between lifetime offspring production and R_0 (condition C_1) is more complicated when the life cycle contains multiple types of offspring. If only a single type of offspring is produced (call it stage 1), then $\mathbf{F}$ will have nonzero entries only in its first row, and $\mathbf{FN}$ will be upper triangular, with its dominant eigenvalue appearing in the $(1, 1)$ position. i.e., the sum of the fertilities of each stage weighted by the expected time spent in that stage. This is precisely the expected lifetime offspring production, so for the case of a single type of offspring, the Cushing-Zhou R_0 also satisfies C_1.

However, if the life cycle contains multiple types of offspring (say stages $1, \ldots, h$), the upper left $h \times h$ corner of $\mathbf{FN}$ will contain the expected lifetime production of offspring of types $1, \ldots, h$ by individuals starting life as types $1, \ldots, h$. Since such a life cycle contains more than one kind of expected lifetime production of offspring, R_0 cannot satisfy C_1 in the sense of being *the* expected lifetime reproduction. Instead, R_0 is calculated from all these expectations (as the dominant eigenvalue of this $h \times h$ submatrix). It determines per-generation growth and population persistence as a function of the expected lifetime production of all types of offspring in a way that satisfies C_2 and C_3.

The right whale The right whale produces only a single type of offspring. The fundamental matrix $\mathbf{N}$ is given by (5.7), the fertility matrix is given by (5.5), and the generation growth matrix is

$$\mathbf{FN} = \begin{pmatrix} 2.18 & 2.42 & 3.06 & 2.60 & 3.06 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \tag{5.39}$$

The dominant eigenvalue of $\mathbf{FN}$ is its $(1, 1)$ entry

$$R_0 = \sum_j f_{1j} E(\nu_{j1}) = 2.18 \tag{5.40}$$

It is interesting to compare $R_0 = 2.18$ with $E(\nu_{14}) = 4.74$. Only female offspring are counted in R_0, whereas $E(\nu_{14})$ counts reproductive events regardless of the sex of the offspring produced. Still, R_0 is less than half of $E(\nu_{14})$, because of the less than perfect survival of calves from t to $t + 1$.

5.4.1 Net Reproductive Rate in Periodic Environments

Periodic time-varying models (Caswell 2001, Chapter 13) are an interesting special case of the multiple offspring type problem. In a periodic model, apparently identical offspring (e.g., seeds) produced at different phases of the cycle (e.g., seasons) are, in effect, of different types of. To the extent that they face different environments, they will differ in their expected offspring production, and R_0 will differ depending on the phase of the cycle in which it is calculated.

The net reproductive rate in a periodic environment was calculated by Hunter and Caswell (2005a) in a study of the sooty shearwater, a pelagic seabird nesting on offshore islands in New Zealand. In that study, the year was divided into two short phases, during which breeding and harvest of chicks occur, and a longer phase encompassing the rest of the year. Let $\mathbf{B}_i = \mathbf{U}_i + \mathbf{F}_i$ be the projection matrix in phase i of the cycle. Without loss of generality, consider an environment with a period of 2

(e.g., winter and summer). The population is projected over a year, starting in phase 1, by

$$\mathbf{A}_1 = \mathbf{B}_2\mathbf{B}_1 \tag{5.41}$$

which is decomposed as

$$\begin{aligned} \mathbf{A}_1 &= (\mathbf{U}_2 + \mathbf{F}_2)\,(\mathbf{U}_1 + \mathbf{F}_1) \\ &= \mathbf{U}_2\mathbf{U}_1 + \mathbf{U}_2\mathbf{F}_1 + \mathbf{F}_2\mathbf{U}_1 + \mathbf{F}_2\mathbf{F}_1 \end{aligned} \tag{5.42}$$

The first term includes only transitions, whereas the last three terms all describe some aspect of reproduction. Thus the annual matrix is $\mathbf{A}_1 = \widehat{\mathbf{U}} + \widehat{\mathbf{F}}$, where

$$\widehat{\mathbf{U}}_1 = \mathbf{U}_2\mathbf{U}_1 \tag{5.43}$$

$$\widehat{\mathbf{F}}_1 = \mathbf{U}_2\mathbf{F}_1 + \mathbf{F}_2\mathbf{U}_1 + \mathbf{F}_2\mathbf{F}_1 \tag{5.44}$$

and

$$R_0^{(1)} = \rho\left[\widehat{\mathbf{F}}_1\left(\mathbf{I} - \widehat{\mathbf{U}}_1\right)^{-1}\right] \tag{5.45}$$

where the superscript 1 indicates that this is the net reproductive rate of a generation beginning in season 1. The corresponding matrices for a generation starting in season 2 are obtained from

$$\mathbf{A}_2 = \mathbf{B}_1\mathbf{B}_2 \tag{5.46}$$

and lead to a net reproductive rate $R_0^{(2)}$. It is easily verified that $R_0^{(1)} \neq R_0^{(2)}$ in general. This contrasts with the population growth rate λ, which is independent of cyclic permutation of the seasons. However, since λ is the same for $\mathbf{A}_1$ and $\mathbf{A}_2$, it must be the case that $R_0^{(1)}$ and $R_0^{(2)}$ are both greater than or less than 1 together.

An alternative formulation of R_0 in periodic environments was published at the same time as Caswell (2009), by Bacaër (2009). He wrote the model, using methods equivalent to those in Sect. 5.5 below, by jointly classifying individuals by stage and by their phase within a seasonal cycle. Let $\mathbf{A}_i = \mathbf{U}_i + \mathbf{F}_i$ be the projection matrix in season i. Then, for example with three seasons, the projection matrix would take the block-circulant form

$$\tilde{\mathbf{A}} = \begin{pmatrix} 0 & 0 & \mathbf{A}_3 \\ \mathbf{A}_1 & 0 & 0 \\ 0 & \mathbf{A}_2 & 0 \end{pmatrix} \tag{5.47}$$

(with similar formulations for $\tilde{\mathbf{U}}$ and $\tilde{\mathbf{F}}$). After some manipulations, Bacaër shows that $\mathbf{R}_0$ is the dominant eigenvalue of the matrix[3]

―――――――――

[3]It might be easier to apply the Cushing-Zhou theorem directly to $\tilde{\mathbf{A}}$ and write

$$R_0 = \rho\left(\tilde{\mathbf{F}}\left(\mathbf{I} - \tilde{\mathbf{U}}\right)^{-1}\right) \tag{5.48}$$

but Bacaër does not do this.

$$\begin{pmatrix} \mathbf{F}_1 & 0 & 0 \\ 0 & \mathbf{F}_2 & 0 \\ 0 & 0 & \mathbf{F}_3 \end{pmatrix} \begin{pmatrix} -\mathbf{U}_1 & \mathbf{I} & 0 \\ 0 & -\mathbf{U}_2 & \mathbf{I} \\ \mathbf{I} & 0 & -\mathbf{U}_3 \end{pmatrix}^{-1}. \tag{5.49}$$

Bacaër (2009) proves that R_0 calculated in this way satisfies condition C_3, providing an indicator for population growth ($R_0 > 1$) or decline ($R_0 < 1$). However, this definition of R_0 does not satisfy C_1 because it does not distinguish the different lifetime reproductive output of individuals born in different seasons.

Cushing and Ackleh (2012) returned to this issue. They argue that the standard approach for studying dynamics of periodic models is to study the "periodic composite map", which is the map for the entire cycle composed of the product of the phase-specific matrices, as in (5.41), which projects over the entire cycle, rather than from one season to the next. They separate transitions and reproduction as in Eqs. (5.43) and (5.44), and prove that R_0 calculated in this way satisfies C_1 (with a different lifetime reproductive output for each starting season) and C_3 (so that the values of R_0 in each season agree in their determination of positive or negative growth). Cushing and Ackleh (2012) also explore the net reproductive rate in nonlinear models, in which R_0 calculated at zero density determines whether the extinction equilibrium is stable.

In the end, it is valuable to have two different ways of calculating R_0, but it highlights the need to carefully specify which properties one wants the index to have.

5.4.2 Sensitivity of the Net Reproductive Rate

Since R_0 is obtained as an eigenvalue, its sensitivity to parameter changes is easy to derive. Let $\mathbf{x}$ and $\mathbf{y}$ be the right and left eigenvectors of $\mathbf{FN}$ corresponding to R_0. Then (Caswell 2006) the sensitivity of R_0 is

$$\frac{dR_0}{d\boldsymbol{\theta}^\mathsf{T}} = \left(\mathbf{y}^\mathsf{T}\mathbf{N}^\mathsf{T} \otimes \mathbf{x}^\mathsf{T} \right) \frac{d\operatorname{vec}\mathbf{F}}{d\boldsymbol{\theta}^\mathsf{T}} + \left(\mathbf{y}^\mathsf{T}\mathbf{N}^\mathsf{T} \otimes \mathbf{x}^\mathsf{T}\mathbf{FN} \right) \frac{d\operatorname{vec}\mathbf{U}}{d\boldsymbol{\theta}^\mathsf{T}} \tag{5.50}$$

The first term captures the effects of changing fertility, the second term captures effects of changes in survival and transitions. The derivation of (5.50) is given in Appendix A.4.

Hint To derive (5.50), write $R_0 = \rho[\mathbf{FN}]$ and write dR_0 in terms of the right and left eigenvectors of $\mathbf{FN}$ and the differential of $\mathbf{FN}$. Then expand $d(\mathbf{FN}) = (d\mathbf{F})\mathbf{N} + \mathbf{F}d(\mathbf{N})$ and apply the vec operator and the chain rule.

The right whale The elasticity of R_0 is shown in Fig. 5.5; R_0 is most elastic to σ_3, less so to σ_2 and σ_4. Remarkably, the elasticity of R_0 to the birth probability γ_3 is zero (actually, $\sim 10^{-9}$). This is a case where lifetime reproductive output is

Fig. 5.5 The elasticity, to
each of the vital rates, of the
net reproductive rate (R_0) for
the right whale. Parameters as
in Fig. 5.2

Fig. 5.5 The elasticity, to each of the vital rates, of the net reproductive rate (R_0) for the right whale. Parameters as in Fig. 5.2
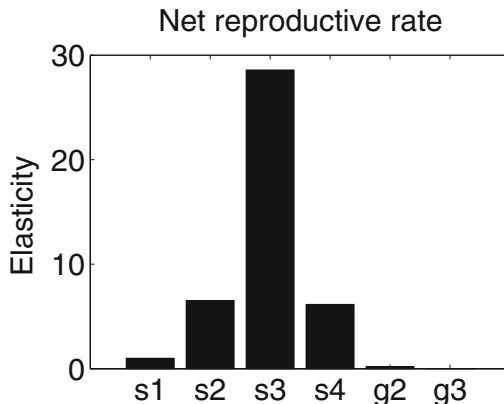
affected strongly by survival, slightly by maturation, but not at all by the probability
of breeding given survival. This seems to be a consequence of the lower survival
probability of mothers; an increase in γ_3 increases the probability of reproduction,
but reduces the lifetime over which that reproduction will be realized.

5.4.3 Invasion Exponents, Selection Gradients, and R_0

Selection on life history traits can be studied in terms of the invasion exponent,
which measures the rate at which a mutation, introduced at low densities, will
increase in the environment created by a resident phenotype (Metz et al. 1992;
Ferriére and Gatto 1993); for a recent introduction see Otto and Day (2007). The
selection gradient on a trait is the derivative of the invasion exponent with respect to
the value of the trait. If the derivative is positive, selection favors an increase in the
trait, and vice-versa. The invasion exponent in a density-independent model is given
by $\log \lambda$. In a density-dependent model, the invasion exponent is given by the growth
rate at equilibrium, $\lambda[\hat{\mathbf{n}}]$. The net reproductive rate R_0 is not, strictly speaking,
an invasion exponent, but because it measures expected lifetime reproduction, it is
attractive as a measure of fitness (see, e.g., the discussion in Kozlowski 1999). Using
R_0 as a measure of fitness will lead to erroneous conclusions unless the selection
gradients, measured in terms of λ and of R_0, give the same answers, i.e., unless
$dR_0/d\theta \propto d \log \lambda/d\theta$.

  For an age-classified model, we write R_0 in terms of the net maternity function
$\phi(x, \theta) = \ell(x, \theta)m(x, \theta)$ where both survival and reproduction depend on some
parameter θ. Then

$$R_0(\theta) = \int_0^\infty \phi(x, \theta)dx \tag{5.51}$$

The growth rate $r = \log \lambda$ is the solution to

$$1 = \int_0^\infty \phi(x, \theta) e^{-r(\theta)x} dx \tag{5.52}$$

Differentiating (5.51) and (5.52) gives

$$\frac{dR_0}{d\theta} = \int_0^\infty \frac{d\phi(x, \theta)}{d\theta} dx \tag{5.53}$$

$$\frac{dr}{d\theta} = \frac{\int_0^\infty e^{-rx} \frac{d\phi(x,\theta)}{d\theta} dx}{\int_0^\infty x\phi(x, \theta) e^{-rx} dx} \tag{5.54}$$

Equation (5.54) is Hamilton's (1966) famous result; the denominator is the generation time measured as the average age of reproduction in the stable age distribution (see Chap. 3).

When $R_0 = 1$ and $r = 0$, it follows from (5.53) and (5.54) that the gradients $dr/d\theta$ and $dR_0/d\theta$ are proportional. Use of either will lead to the same conclusions about selection. But when $r \neq 0$, this is not the case. If $r > 0$, then $dr/d\theta$ is reduced for traits that operate at later ages, because $d\phi/dx$ is weighted by e^{-rx}. It is an open problem to generalize this result to stage-classified models, and prove that

$$\frac{d \log \lambda}{d\theta^\mathsf{T}} \propto \frac{dR_0}{d\theta^\mathsf{T}} \tag{5.55}$$

when $\lambda = R_0 = 1$. In a few cases I have examined, it appears to be true numerically. As the following example shows, it is certainly the case that when $\lambda \neq 1$, the derivatives are not generally proportional.

The right whale The lack of proportionality between the selection gradients in terms of λ and of R_0 means that evolutionary conclusions will differ depending on which is used, especially when tradeoffs exist between two or more traits. For example, for the right whale, $\lambda = 1.025$ and $R_0 = 2.183$. Figure 5.6 shows the sensitivity of λ and of R_0; while the patterns are similar, they are *not* proportional, and the use of R_0 as an invasion exponent would result in erroneous predictions. Suppose a trait existed that would increase the birth probability γ_3 at the cost of a reduction in calf survival σ_1, with the cost measured by $c = -d\sigma_1/d\gamma_3$. An increase in this trait would be favored by selection provided that

$$c < \frac{\partial\lambda/\partial\gamma_3}{\partial\lambda/\partial\sigma_1} = 0.96 \tag{5.56}$$
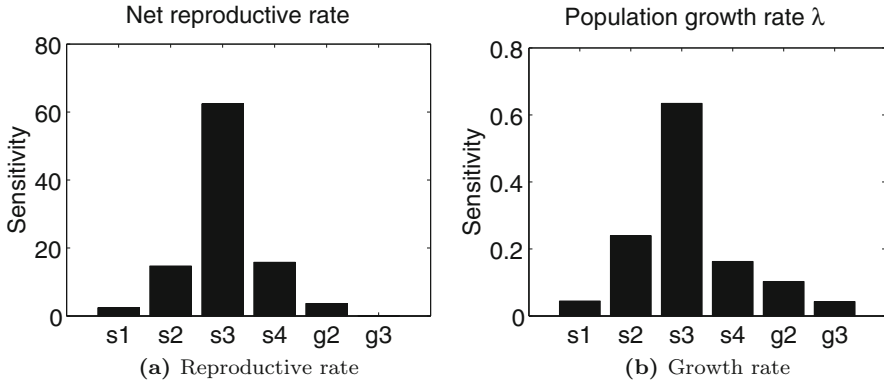
Fig. 5.6 (**a**) The sensitivity, to each of the vital rates, of the net reproductive rate R_0 for the right whale. (**b**) The sensitivity of population growth rate λ. The derivative of λ is the selection gradient; use of the derivative of R_0 leads to erroneous predictions unless the population is at equilibrium. Parameters as in Fig. 5.2

But if expected lifetime reproduction was used as an invasion exponent, the analysis would conclude that selection would favor an increase in the trait only if

$$c < \frac{\partial R_0/\partial \gamma_3}{\partial R_0/\partial \sigma_1} = 0.0 \qquad\qquad (5.57)$$

That is, according to R_0, any cost whatsoever of increased birth rate would prevent selection from favoring it. According to λ (and correctly, in this case), selection would favor increased birth rate provided that the cost was not too great. In spite of the superficial similarity of the patterns in Fig. 5.6, the evolutionary implications are quite different, reflecting the impact of *timing* of life history events on λ. The sensitivities of λ to σ_2 and γ_2, which influence early survival and the age at maturity, are larger than the sensitivities of R_0 to the same parameters.

5.4.4 Beyond R_0: Individual Stochasticity in Lifetime Reproduction

Variation among individuals is fundamental to population biology. As argued here, two sources of variation must be distinguished: heterogeneity and individual stochasticity Heterogeneity refers to genuine differences among individuals, because of which the individuals experience different vital rates. Individual stochasticity refers to the apparent differences that result from the random outcome of identical vital rates, applied to identical individuals. We have seen above that individual stochasticity is always present. That is particularly true of lifetime

reproductive output (LRO). The net reproductive rate is the *expectation* of LRO, but what can we say about the variance among individuals.

Empirical measurement shows that LRO is usually highly variable among individuals and positively skewed. Typically, a few individuals produce many offspring while most produce few, or none at all (Clutton-Brock 1988; Newton 1989). If this variance reflected heterogeneity among individual properties, and if the heterogeneity had a genetic basis, the variance would provide material for natural selection (the "opportunity for selection" of Crow 1958). Population and quantitative genetics are replete with methods to measure such genetic variation; e.g., Lande and Arnold (1983) and Endler (1986).

However, variance among individuals in LRO is not evidence of heterogeneity, genetic or otherwise; some is due to individual stochasticity. Only after evaluating the extent of individual stochasticity can data on LRO be interpreted as evidence for heterogeneity (Caswell 2011; Tuljapurkar et al. 2009; Steiner et al. 2010; Steiner and Tuljapurkar 2012). Caswell (2011) developed a method to calculate the mean, variance, and higher moments of lifetime reproductive output for any age- or stage-classified life cycle, using Markov chains with rewards; see van Daalen and Caswell (2015, 2017) for full details. In these models[4] the movement of the individual through its life cycle is described by an absorbing Markov chain; mortality appears as transitions to an absorbing (dead) state. At each step, the individual accumulates a "reward." In our context, the reward is the production of offspring. The reproductive reward is a random variable with a specified set of moments. The reward accumulated by the inevitable death of the individual is its LRO. Although every individual experiences the same vital rates—there is no heterogeneity—each individual may experience a different life and thus a different lifetime reproductive output.

Stage-specific reproductive output is specified by a set of reward matrices $\mathbf{R}_k$. The (i, j) element of $\mathbf{R}_k$ is the kth moment of the reproductive output associated with the transition from stage j to stage i. Given the reward matrices, the Markov chain transition matrix $\mathbf{P}$, and the reasonable assumption that the dead do not reproduce, all the moments of LRO can be calculated (van Daalen and Caswell 2017).

Let $\tilde{\boldsymbol{\rho}}_k$ be a vector containing the kth moments of LRO for individuals starting in each transient (living) stage. Then, it has been shown (van Daalen and Caswell 2017) that, e.g., the first two moments of LRO are

$$\tilde{\boldsymbol{\rho}}_1 = \mathbf{N}^\mathsf{T}\mathbf{Z}\left(\mathbf{P} \circ \mathbf{R}_1\right)^\mathsf{T}\mathbf{1}_{s+1} \tag{5.58}$$

$$\tilde{\boldsymbol{\rho}}_2 = \mathbf{N}^\mathsf{T}\left[\mathbf{Z}(\mathbf{P} \circ \mathbf{R}_2)^\mathsf{T}\mathbf{1}_{s+1} + 2(\mathbf{U} \circ \mathbf{R}_1)^\mathsf{T}\tilde{\boldsymbol{\rho}}_1\right] \tag{5.59}$$

---

[4]Markov chains with rewards have a long history in stochastic process theory; see Howard (1960), Puterman (1994), and Sheskin (2010).

where s is the number of stages in the life cycle, $\circ$ denotes the Hadamard product, $\mathbf{N} = (\mathbf{I} - \mathbf{U})^{-1}$ is the fundamental matrix of the Markov chain and $\mathbf{Z}$ is a matrix that selects the living states. From these moment vectors we can calculate all the statistics of LRO. In addition, the full sensitivity analysis, calculating the derivatives of any of the moments of LRO to any parameters affecting any of the transition, mortality, or reward matrices, has been presented by van Daalen and Caswell (2017).

One of the most significant findings of this line of research has been that, in many cases, individual stochasticity can account for most or all of the observed phenotypic variance in LRO (Steiner and Tuljapurkar 2012; van Daalen and Caswell 2017). It appears that the contribution of stochasticity to variance in lifetime reproductive output has been underappreciated.

5.5 Variable and Stochastic Environments

The variance due to individual stochasticity can be examined in the case of variable environments (Caswell 2006; Tuljapurkar and Horvitz 2006; Horvitz and Tuljapurkar 2008; see also Chap. 8). Several cases can be considered:

- Deterministic aperiodic environments. These usually appear as specific historical sequences; e.g., the specific sequence of vital rates exhibited by the right whale between 1980 and 1998 (Caswell 2006). That sequence is fixed, and is neither random nor periodic.
- Periodic environments. A periodic model may describe seasonal variation within a year, or may approximate inter-annual variability in events such as floods, fires, or hurricanes.
- Stochastic iid environments. In such environments, successive states are drawn independently from a fixed probability distribution; hence the identifier iid, short for "independent and identically distributed."
- Markovian stochastic environments. In a Markovian environment the probability distribution of the next environmental state may depend on the current state. This permits study of the effects of environmental autocorrelation. Markovian environments include periodic and iid environments as special cases.

See Tuljapurkar (1990) for a thorough discussion of types of stochastic environments.

When studying variable environments, it is important to distinguish *period* and *cohort* calculations. Period calculations are based on the vital rates in a given year. They describe the results of the hypothetical situation where the conditions of year t are maintained indefinitely, and compare those to the results for conditions in year $t + 1$, etc. Period calculations are a way to summarize the effects of changing environment. But an individual born in year t does not live its life under the conditions of year t. It spends its first year of life under the conditions in year t, its second year under the conditions of year $t + 1$, and so on. Results calculated in this way are called *cohort* calculations, because they describe a cohort born in

year t and living through the environmental sequence starting then. Period-specific calculations are easy; simply apply the time-invariant calculation to the vital rates of each year and tabulate the results. Cohort calculations, however, must account for all the possible environmental sequences through which a cohort may pass. Caswell (2006) and Tuljapurkar and Horvitz (2006) independently introduced two different, complementary approaches to doing so. I will present the former approach here.

5.5.1 A Model for Variable Environments

In a variable environment, the transient matrix $\mathbf{U}$ is a time-varying matrix $\mathbf{U}(t)$. We can define a fundamental matrix by

$$\mathbf{N} = \mathbf{I} + \mathbf{U}(0) + \mathbf{U}(1)\mathbf{U}(0) + \mathbf{U}(2)\mathbf{U}(1)\mathbf{U}(0) + \cdots \tag{5.60}$$

The (i, j) element of $\mathbf{N}$ is the expected occupancy time in transient state i by an individual starting in transient state j at time 0, and experiencing the specific sequence of environments $\mathbf{U}(0), \mathbf{U}(1), \ldots$. Thus there will be a different matrix $\mathbf{N}$ for each possible environmental sequence.

Tuljapurkar and Horvitz (2006), whose paper I highly recommend, work directly from (5.60) to develop the means and variances of $\mathbf{N}$, $\boldsymbol{\eta}$, and survivorship, in periodic, iid, and Markovian environments. Here, we consider an approach in which an individual is jointly classified by stage and environment, using the vec-permutation model developed by Hunter and Caswell (2005b).

Suppose that there are q environmental states $\epsilon = 1, \ldots, q$ and s stages, $g = 1, \ldots, s$. Corresponding to environment i is a $s \times s$ transient matrix $\mathbf{U}_i$. Assemble the matrices $\mathbf{U}_i$ into a block-diagonal matrix

$$\mathbb{U} = \begin{pmatrix} \mathbf{U}_1 & & \\ & \ddots & \\ & & \mathbf{U}_q \end{pmatrix} \tag{5.61}$$

of dimension $sq \times sq$.

The transitions among environmental states are defined by a $q \times q$ column-stochastic matrix $\mathbf{D}$. Use the matrix $\mathbf{D}$ to construct a block-diagonal environmental transition matrix

$$\mathbb{D} = \begin{pmatrix} \mathbf{D} & 0 & \cdots & 0 \\ 0 & \mathbf{D} & \cdots & 0 \\ & & \ddots & \\ 0 & 0 & \cdots & \mathbf{D} \end{pmatrix} \tag{5.62}$$

of dimension $sq \times sq$.

Suppose that there are 4 environmental states. In an aperiodic deterministic environment,

$$\mathbf{D} = \begin{pmatrix} 0\ 0\ 0\ 0 \\ 1\ 0\ 0\ 0 \\ 0\ 1\ 0\ 0 \\ 0\ 0\ 1\ 1 \end{pmatrix} \tag{5.63}$$

That is, the environment moves deterministically from state 1 to state 2 to state 3 to state 4. Setting $d_{44} = 1$ solves the problem of what to do at the end of the sequence, by the (possibly satisfactory) trick of letting the final state repeat indefinitely. In a periodic environment,

$$\mathbf{D} = \begin{pmatrix} 0\ 0\ 0\ 1 \\ 1\ 0\ 0\ 0 \\ 0\ 1\ 0\ 0 \\ 0\ 0\ 1\ 0 \end{pmatrix} \tag{5.64}$$

In an iid environment in which environment i occurs with probability π_i,

$$\mathbf{D} = \begin{pmatrix} \pi_1\ \pi_1\ \pi_1\ \pi_1 \\ \pi_2\ \pi_2\ \pi_2\ \pi_2 \\ \pi_3\ \pi_3\ \pi_3\ \pi_3 \\ \pi_4\ \pi_4\ \pi_4\ \pi_4 \end{pmatrix} \tag{5.65}$$

In a Markovian environment, $\mathbf{D}$ is a column stochastic transition matrix describing the transition probabilities. I will assume that the environmental Markov chain is ergodic, with a stationary probability distribution denoted by $\boldsymbol{\pi}$. This gives the long-term frequency of occurrence of each environmental state.

The state of the cohort can be specified by a matrix $\mathbf{X}$, of dimension $s \times q$, with rows corresponding to stages and columns to environments, and where $x_{ij}(t)$ is the expected number of individuals in stage i and environmental state j at time t.

$$\mathbf{X}(t) = \begin{pmatrix} x_{11} & \cdots & x_{1q} \\ \vdots & & \vdots \\ x_{s1} & \cdots & x_{sq} \end{pmatrix} \tag{5.66}$$

We rearrange $\mathbf{X}$ into a vector by applying the vec operator to $\mathbf{X}^{\mathsf{T}}$,

$$\text{vec}\,\mathbf{X}^{\mathsf{T}} = \begin{pmatrix} x_{11} & \cdots & x_{1q} \big| \cdots \big| x_{s1} & \cdots & x_{sq} \end{pmatrix}^{\mathsf{T}} \tag{5.67}$$

The first block of entries gives stage 1 individuals in environments 1 through q. The second block gives stage 2 individuals in environments 1 through q, and so on.

To describe the dynamics of the cohort, suppose that individuals first move among stages, according to the vital rates determined by the current environment, and then the environment changes to a new state according to **D**. Then

$$\text{vec}^\mathsf{T}\mathbf{X}(t+1) = \mathbb{D}\,\mathbf{K}_{s,q}\,\mathbb{U}\,\mathbf{K}_{s,q}^\mathsf{T}\,\text{vec}^\mathsf{T}\mathbf{X}(t) \qquad (5.68)$$

The matrix $\mathbf{K}_{s,q}$ is the vec-permutation matrix (Henderson and Searle 1981; Hunter and Caswell 2005b), commutation matrix (Magnus and Neudecker 1979), which permutes the entries of a vector so that

$$\text{vec}^\mathsf{T}\mathbf{X} = \mathbf{K}_{s,q}\,\text{vec}\,\mathbf{X} \qquad (5.69)$$

(see Sect. 2.2.3). Like all permutation matrices, its transpose is also its inverse. Its role here is to rearrange the population vector into a form appropriate for multiplication by the block-diagonal matrices $\mathbb{B}$ and $\mathbb{D}$.

Working from right to left, (5.68) first rearranges the vector, then applies the block-transition matrix $\mathbb{U}$, then reverses the rearrangement of the vector, and finally applies the environmental transition block matrix $\mathbb{D}$ to obtain the expected cohort at $t+1$. This gives a transition matrix for the joint process,

$$\widetilde{\mathbf{U}} = \mathbb{D}\,\mathbf{K}_{s,q}\,\mathbb{U}\,\mathbf{K}_{s,q}^\mathsf{T} \qquad (5.70)$$

that incorporates the demographic transitions within each environment and the patterns of time variation among environments.[5] Here and in what follows, the tilde distinguishes the matrix from the environment-specific matrices.

Matrices of similar form, but not using this formalism, were introduced by Horvitz to study populations in habitat patches where the habitat patches change state over time, for example in recovering from disturbance (Horvitz and Schemske 1986; Pascarella and Horvitz 1998). Horvitz introduced the term "megamatrix" to describe these models. A megamatrix, in the sense of Horvitz, is a special case of (5.70) when the population is classified by stages within environmental states, the demographic matrices are applied first, and the environmental transition matrices $\mathbf{D}_i$ are identical for all stages, as is the case in (5.62).

---

[5]Note that (5.68) computes the *expected* population at $t+1$ from the *expected* population at t. It might be tempting to do this with the projection matrix $\mathbb{A}$ and use the eigenvalues of $\tilde{\mathbf{A}}$ to calculate the stochastic population growth rate. However, this would give the growth rate of the mean population, but not the stochastic growth rate (which is always less than or equal to the growth rate of the mean population). For calculations such as moments of longevity, which are explicitly properties of the expected population, the difference does not arise.

5.5.2   The Fundamental Matrix

Since $\widetilde{\mathbf{U}}$ is the transient matrix of an absorbing Markov chain, the fundamental matrix in the time-varying environment is

$$\widetilde{\mathbf{N}} = \left(\mathbf{I}_{sp} - \widetilde{\mathbf{U}}\right)^{-1} \tag{5.71}$$

The elements of $\widetilde{\mathbf{N}}$ give the expected occupancy times in each stage, in each environment, as a function of the starting stage and starting environment.

Notation alert Developing a complete system of notation for $\widetilde{\mathbf{N}}$ would obscure more than it would clarify. Pictures can help. As I present the fundamental matrix and some of the properties calculated from it, I will use diagrams for a simple case with three stages and two environments. I will often indicate the dimension of matrices and vectors with subscripts. I will use g to denote stages ($g = 1, 2, \ldots, s$) and ϵ to denote environments ($\epsilon = 1, \ldots, q$). I will use superscripts on $\widetilde{\mathbf{N}}$ and quantities derived from it, to distinguish different ways of combining information across environmental states (see Table 5.1).

Recall that in a constant environment, v_{ij} was the number of visits to stage i, starting in stage j. Now we must consider the visits to stage i in environment ϵ, starting in stage j and environment ϵ_0, so we write

$$\widetilde{\mathbf{N}} = E\left(v_{ij,\epsilon}|\epsilon_0\right) \tag{5.72}$$

Table 5.1 Superscript notation for time-varying models. The tilde indicates quantities calculated from the complete transient matrix $\widetilde{\mathbf{U}}$ in (5.70). Occupancy and times to absorbtion depend on the initial and final demographic and environmental states. The superscripts ($\ddagger$, $\S$, $\heartsuit$) indicate choices of summing and averaging over the environmental states. The superscripts are shown here for the fundamental matrix $\widetilde{\mathbf{N}}$

Symbol	Definition	Description	Equation
$\widetilde{\mathbf{N}}$	$E\left(v_{ij,\epsilon}|\epsilon_0\right)$	Expected visits to state i in environment ϵ, starting from state j in environment ϵ_0	(5.72)
$\widetilde{\mathbf{N}}^{\ddagger}$	$E\left(v_{ij}|\epsilon_0\right)$	Expected visits to state i, summed over environments, starting from state j in environment ϵ_0	(5.73)
$\widetilde{\mathbf{N}}^{\ddagger\ddagger}$		Rearrangement of the rows and columns of $\widetilde{\mathbf{N}}^{\ddagger}$	(5.74)
$\widetilde{\mathbf{N}}^{\S}$	$E\left(v_{ij,\epsilon}\right)$	Expected visits to state i and environmental state ϵ, averaged over initial environmental states	(5.75)
$\widetilde{\mathbf{N}}^{\S\S}$		Rearrangement of the rows and columns of $\widetilde{\mathbf{N}}^{\S}$	(5.76)
$\widetilde{\mathbf{N}}^{\heartsuit}$	$E\left(v_{ij}\right)$	Expected visits to state i summed over environments, starting from state j and averaged over initial environmental states	(5.77)

The structure of $\widetilde{\mathbf{N}}$ when $s = 3$ and $q = 2$ is

		$g = 1$		$g = 2$		$g = 3$	
		$\epsilon_0 = 1$	$\epsilon_0 = 2$	$\epsilon_0 = 1$	$\epsilon_0 = 2$	$\epsilon_0 = 1$	$\epsilon_0 = 2$
$g = 1$	$\epsilon = 1$						
	$\epsilon = 2$						
$g = 2$	$\epsilon = 1$						
	$\epsilon = 2$						
$g = 3$	$\epsilon = 1$						
	$\epsilon = 2$						

$\widetilde{\mathbf{N}}_{sq \times sq}$:

From $\widetilde{\mathbf{N}}$ we can obtain the expected occupancy time in each stage, regardless of the environment in which those visits occur, by aggregating rows. The resulting matrix $\widetilde{\mathbf{N}}^{\ddagger}$ is

$$\widetilde{\mathbf{N}}^{\ddagger} = E\left(\nu_{ij} | \epsilon_0\right)$$
$$= \left(\mathbf{I}_s \otimes \mathbf{1}_{q \times 1}^{\mathsf{T}}\right) \widetilde{\mathbf{N}} \tag{5.73}$$

where $\mathbf{1}_{q \times 1}$ is a vector of ones. The structure of $\widetilde{\mathbf{N}}^{\ddagger}$ is

		$g = 1$		$g = 2$		$g = 3$	
		$\epsilon_0 = 1$	$\epsilon_0 = 2$	$\epsilon_0 = 1$	$\epsilon_0 = 2$	$\epsilon_0 = 1$	$\epsilon_0 = 2$
$\widetilde{\mathbf{N}}^{\ddagger}_{s \times sq}$:	$g = 1$						
	$g = 2$						
	$g = 3$						

If it is useful to group stages within initial environments, rather than grouping environments within stages, $\widetilde{\mathbf{N}}^{\ddagger}$ can be rearranged as

$$\widetilde{\mathbf{N}}^{\ddagger\ddagger} = \widetilde{\mathbf{N}}^{\ddagger} \, \mathbf{K}_{s,q} \tag{5.74}$$

with the structure

		$\epsilon_0 = 1$			$\epsilon_0 = 2$		
		$g = 1$	$g = 2$	$g = 3$	$g = 1$	$g = 2$	$g = 3$
$\widetilde{\mathbf{N}}^{\ddagger\ddagger}_{s \times sq}$:	$g = 1$						
	$g = 2$						
	$g = 3$						

The matrices $\widetilde{\mathbf{N}}^{\ddagger}$ and $\widetilde{\mathbf{N}}^{\ddagger\ddagger}$ both display expected occupancy of each stage as a function of initial state and environment. To describe the fates of individuals without specifying their initial environment, we take an expectation over the stationary distribution $\boldsymbol{\pi}$ of initial environments. This gives

$$\widetilde{\mathbf{N}}^{\S} = E\left[v_{ij,\epsilon}\right]$$

$$= \widetilde{\mathbf{N}}\left(\mathbf{I}_s \otimes \boldsymbol{\pi}\right) \tag{5.75}$$

The structure of $\widetilde{\mathbf{N}}^{\S}$ is

$$
\widetilde{\mathbf{N}}^{\S}_{sq \times s} :
\begin{array}{cc}
 & \begin{array}{c} \epsilon_0 = \bar{\epsilon} \\ g=1 \; g=2 \; g=3 \end{array} \\
\begin{array}{cc}
g=1 & \epsilon=1 \\
 & \epsilon=2 \\
g=2 & \epsilon=1 \\
 & \epsilon=2 \\
g=3 & \epsilon=1 \\
 & \epsilon=2
\end{array} & \boxed{}
\end{array}
$$

The rows of $\widetilde{\mathbf{N}}^{\S}$ can be rearranged to display stages within environments, giving

$$\widetilde{\mathbf{N}}^{\S\S} = \mathbf{K}^{\mathsf{T}}_{s,q} \; \widetilde{\mathbf{N}}^{\S} \tag{5.76}$$

with the structure

$$
\widetilde{\mathbf{N}}^{\S\S}_{sq \times s} :
\begin{array}{cc}
 & \begin{array}{c} \epsilon_0 = \bar{\epsilon} \\ g=1 \; g=2 \; g=3 \end{array} \\
\begin{array}{cc}
\epsilon=1 & g=1 \\
 & g=2 \\
 & g=3 \\
\epsilon=2 & g=1 \\
 & g=2 \\
 & g=3
\end{array} & \boxed{}
\end{array}
$$

Finally, aggregating over destination environments *and* averaging over initial environments gives a matrix containing the expected occupancy of stages as a function of initial stage, averaged over environments

$$\widetilde{\mathbf{N}}^{\heartsuit} = E\left[v_{ij}\right]$$

$$= \left(\mathbf{I}_s \otimes \mathbf{1}^{\mathsf{T}}_{q \times 1}\right) \widetilde{\mathbf{N}} \left(\mathbf{I}_s \otimes \boldsymbol{\pi}\right) \tag{5.77}$$

The structure of $\widetilde{\mathbf{N}}^\heartsuit$ is

$$
\begin{array}{c}
\epsilon_0 = \bar{\epsilon} \\
g = 1 \quad g = 2 \quad g = 3
\end{array}
$$

$$
\widetilde{\mathbf{N}}^\heartsuit_{s \times s} : \quad
\begin{array}{c}
g = 1 \\
g = 2 \\
g = 3
\end{array}
\begin{array}{|c|}
\hline
 \\
\\
\\
\hline
\end{array}
$$

The matrix $\widetilde{\mathbf{N}}^\heartsuit$, obtained by the simple calculation (5.77), is "the" fundamental matrix for the variable environment. It could be compared directly to the fundamental matrix in a constant environment (e.g., the environment defined by one of the environmental states).

5.5.3   Longevity in a Variable Environment

Life expectancy, as a function of initial stage and initial environment is obtained by summing the columns of $\widetilde{\mathbf{N}}$,

$$
E\left(\widetilde{\boldsymbol{\eta}}^\mathsf{T}\right) = E\left[\boldsymbol{\eta}^\mathsf{T} | \epsilon_0\right]
$$
$$
= \mathbf{1}^\mathsf{T}_{sq \times 1} \widetilde{\mathbf{N}} \tag{5.78}
$$

The structure of $E\left(\widetilde{\boldsymbol{\eta}}^\mathsf{T}\right)$ is

$$
\begin{array}{cccccc}
 & g = 1 & & g = 2 & & g = 3 \\
E\left(\widetilde{\boldsymbol{\eta}}^\mathsf{T}\right): & \epsilon_0 = 1 \ \ \epsilon_0 = 2 & & \epsilon_0 = 1 \ \ \epsilon_0 = 2 & & \epsilon_0 = 1 \ \ \epsilon_0 = 2
\end{array}
$$

$$
\begin{array}{|c|c|c|}
\hline
 & & \\
\hline
\end{array}
$$

Averaging this conditional life expectancy over the stationary distribution $\boldsymbol{\pi}$ of initial environments gives

$$
E\left(\widetilde{\boldsymbol{\eta}}^\heartsuit\right) = E\left(\widetilde{\boldsymbol{\eta}}\right)\left(\mathbf{I}_s \otimes \boldsymbol{\pi}\right) \tag{5.79}
$$

This measure of life expectancy in a variable environment is directly comparable to $E(\boldsymbol{\eta})$ calculated from the same life history in a constant environment.

5.5.3.1   Variance in Longevity

In a constant environment, the variance among individuals in longevity is due to individual stochasticity. In a time-varying environment, the variance contains an

additional component due to differences among individuals as a function of their environment at birth. Applying (5.27) to $\widetilde{\mathbf{N}}$ we obtain the variances conditional on the initial environment:

$$V\left[\widetilde{\boldsymbol{\eta}}^{\mathsf{T}}|\epsilon_0\right] = E\left(\widetilde{\boldsymbol{\eta}}^{\mathsf{T}}\right)(2\widetilde{\mathbf{N}} - \mathbf{I}_{sq}) - E\left(\widetilde{\boldsymbol{\eta}}^{\mathsf{T}}\right) \circ E\left(\widetilde{\boldsymbol{\eta}}^{\mathsf{T}}\right) \tag{5.80}$$

As indicated by the notation, $V\left[\widetilde{\boldsymbol{\eta}}^{\mathsf{T}}|\epsilon_0\right]$ is a conditional variance of $\widetilde{\boldsymbol{\eta}}$, given the initial environment ϵ_0. The initial environment is distributed according to the stationary distribution $\boldsymbol{\pi}$, so the unconditional longevity $\boldsymbol{\eta}$ follows a finite mixture distribution with mixing distribution $\boldsymbol{\pi}$.

The unconditional variance of $\boldsymbol{\eta}$, taking account of both sources of variability, is

$$V\left[\widetilde{\boldsymbol{\eta}}^{\mathsf{T}}\right] = V\left[E(\widetilde{\boldsymbol{\eta}}^{\mathsf{T}}|\epsilon_0)\right] + E_\pi\left[V(\widetilde{\boldsymbol{\eta}}^{\mathsf{T}}|\epsilon_0)\right] \tag{5.81}$$

where E_π denotes the expectation over the stationary distribution $\boldsymbol{\pi}$ of initial environments (Rényi 1970, p. 275, Theorem 1). This can be rearranged as

$$V\left[\widetilde{\boldsymbol{\eta}}^{\mathsf{T}}\right] = E_\pi\left[\widetilde{\boldsymbol{\eta}}^{\mathsf{T}} \circ \widetilde{\boldsymbol{\eta}}^{\mathsf{T}}\right] - E_\pi\left[\widetilde{\boldsymbol{\eta}}^{\mathsf{T}}\right] \circ E_\pi\left[\widetilde{\boldsymbol{\eta}}^{\mathsf{T}}\right] + E_\pi\left[V(\boldsymbol{\eta}^{\mathsf{T}}|\epsilon_0)\right]$$

$$= \left[E\left(\widetilde{\boldsymbol{\eta}}^{\mathsf{T}}\right) \circ E\left(\widetilde{\boldsymbol{\eta}}^{\mathsf{T}}\right)\right](\mathbf{I}_s \otimes \boldsymbol{\pi}) - \left[E\left(\widetilde{\boldsymbol{\eta}}^{\heartsuit}\right) \circ E\left(\widetilde{\boldsymbol{\eta}}^{\heartsuit}\right)\right]^{\mathsf{T}}$$

$$+ V\left[\widetilde{\boldsymbol{\eta}}^{\mathsf{T}}|\epsilon_0\right](\mathbf{I}_s \otimes \boldsymbol{\pi}) \tag{5.82}$$

(e.g., Frühwirth-Schnatter 2006, p. 10). This variance decomposition has developed into a powerful tool for the analysis of heterogeneity in demography (Edwards 2011; Hartemink and Caswell 2018; Hartemink et al. 2017; Caswell et al. 2018; Jenouvrier et al. 2018).

The choice of the mixing distribution $\boldsymbol{\pi}$ is important. Hernandez-Suarez et al. (2012) present an alternative where $\boldsymbol{\pi}$ is the stationary distribution of births across environments, rather than the distribution of environments itself.

5.5.4 A Time-Varying Example: Lomatium bradshawii

Lomatium bradshawii is an endangered herbaceous perennial plant, found in only a few isolated populations in prairies of Oregon and Washington. These habitats were, until recent times, subject to natural and anthropogenic fires, to which *L. bradshawii* seems to have adapted. Fall-season fires increase plant size and seedling recruitment, but the effect fades within a few years. Populations in burned areas have higher growth rates and lower probabilities of extinction than unburned populations (Caswell and Kaye 2001).

A stochastic demographic model for *L. bradshawii* was developed by Caswell and Kaye (2001), Kaye et al. (2001), and Kaye and Pyke (2003) based on data from an experimental study using controlled burning. Individuals were classified into six stages based on size and reproductive status: yearlings, small and large vegetative plants, and small, medium, and large reproductive plants. The environment was classified into four states defined by fire history: the year of a fire and 1, 2, and 3+ years post-fire. Projection matrices were estimated in each environment; the example here is based on one of the two sites (Rose Prairie) in the original study. The matrices are given in Caswell and Kaye (2001).

L. bradshawii performs well under recently burned conditions, but less well in sites that have not been recently burned. For example, the values of λ are

$$
\begin{array}{lcccc}
\text{Years since fire:} & 0 & 1 & 2 & \geq 3 \\
\text{Growth rate } \lambda : & 1.18 & 1.12 & 0.48 & 0.88
\end{array}
$$

Caswell and Kaye (2001) found a minimum frequency of fire (0.4–0.5) below which the stochastic growth rate was negative and the population would be unable to persist. Effects of autocorrelation were small, but positive autocorrelation reduced the stochastic growth rate.

As an example of a time-varying analysis, let us examine *L. bradshawii* in a Markovian environment. Let f be the long-term frequency of fire, and ρ the temporal autocorrelation. Then the transition matrix for environmental states is

$$
\mathbf{D} = \begin{pmatrix}
p & q & q & q \\
1-p & 0 & 0 & 0 \\
0 & 1-q & 0 & 0 \\
0 & 0 & 1-q & 1-q
\end{pmatrix}
\tag{5.83}
$$

where $q = f(1 - \rho)$ and $p = \rho + q$.

Figure 5.7a shows the life expectancy $E(\widetilde{\boldsymbol{\eta}}|\epsilon_0)$ of *L. bradshawii* as a function of initial stage and initial environmental state, from (5.78). Life expectancy increases with the stage (size) of a plant. A seedling has its greatest life expectancy in the year of a fire, less in an environment three or more years post-fire. A large flowering plant, in contrast, has its greatest life expectancy in an environment three or more years post-fire. When the environment-dependence is averaged over the stationary distribution of environmental states, there is a smooth increase in life expectancy from ~2.5 years for a seedling to 8 years for a large flowering plant (Fig. 5.7b). The standard deviation of longevity also increases with stage, in a pattern very similar to that of the expectation.

These patterns in the mean and variance of longevity (Fig. 5.7) depend on the stochastic properties of the environment—in this case, the frequency f and autocorrelation ρ of fires. Even with an environmental model this simple, the effects of f and ρ can be complicated. I know of no previous attempts to examine their effects on longevity. To do so, I calculated life expectancy with $f = 0.5$ for autocorrelation $-1 < \rho < 1$, and with $\rho = 0$ for fire frequency $0 < f < 1$.

(a) Life expectancy $\widetilde{\eta}_1$

(b) Life expectancy $\widetilde{\eta}_1^\heartsuit$

(c) Standard deviation of $\widetilde{\eta}_1$

(d) Standard deviation of $\widetilde{\eta}_1^\heartsuit$
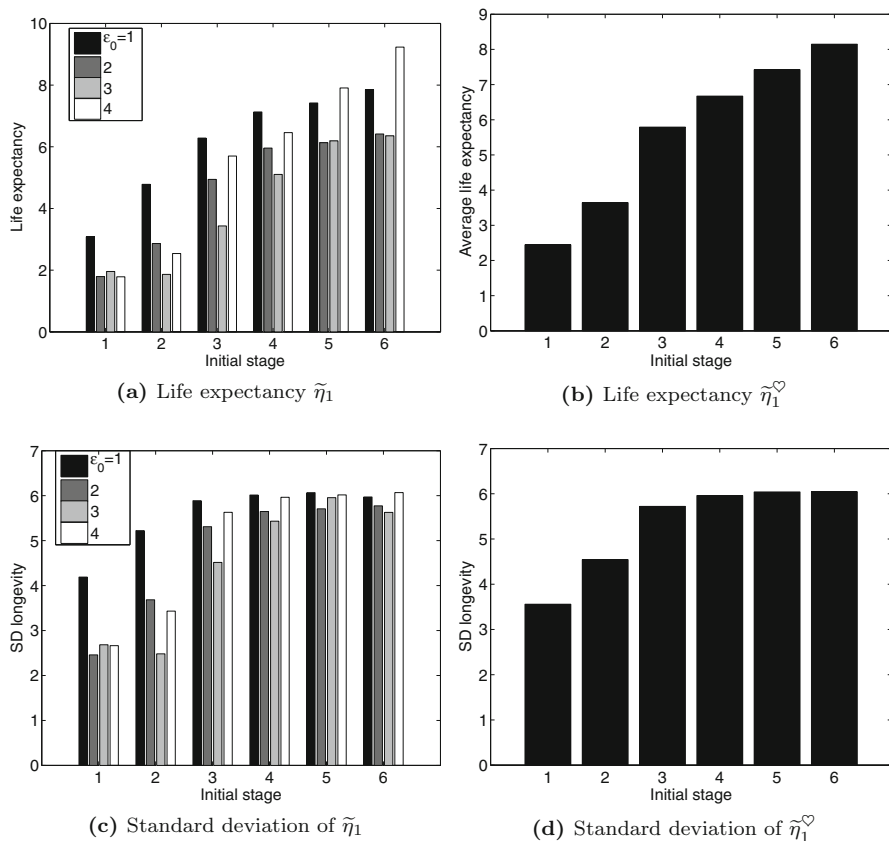
Fig. 5.7 The expectation and standard deviation of longevity for *Lomatium bradshawii* in a stochastic fire environment. (**a**) Expected longevity conditional on initial environment (ϵ_0). (**b**) Expected longevity averaged over the stationary distribution of initial environments. (**c**) The standard deviation of longevity conditional on initial environment. (**d**) The standard deviation of longevity over the stationary distribution of initial environments. The frequency of fire is 0.5 and the temporal autocorrelation $\rho = 0.7$

The life expectancy of early life cycle stages increases monotonically with fire frequency (Fig. 5.8a), but the life expectancy of large reproductive plants is greatest at either low or high fire frequencies. The standard deviation of longevity increases with f (Fig. 5.8b). As $f \to 1$, the standard deviation of longevity is approximately twice the mean.

The autocorrelation of fires has little effect on the life expectancy of seedlings, but a larger effect on that of large plants. For the latter, life expectancy is maximized as $\rho \to -1$ (alternating fire and non-fire years) or as $\rho \to 1$ (long periods of fires
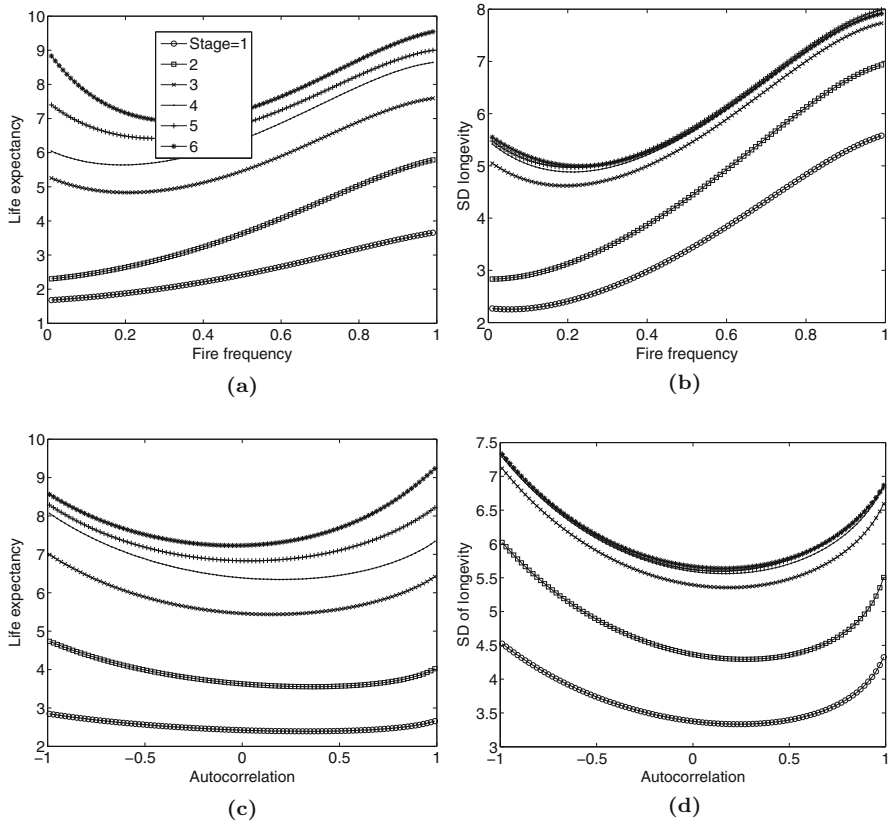
Fig. 5.8 The expectation and standard deviation of longevity, averaged over the stationary distribution of initial environments, for *Lomatium bradshawii*, as a function of the initial stage, the fire frequency f, and the temporal autocorrelation ρ. Parameters as in Fig. 5.7. (**a**) Life expectancy $\widetilde{\boldsymbol{\eta}}^{\heartsuit}$. (**b**) Standard deviation of longevity. (**c**) Life expectancy $\widetilde{\boldsymbol{\eta}}^{\heartsuit}$. (**d**) Standard deviation of longevity

alternating with long periods without fire). The standard deviation of longevity also shows a strong U-shaped response to ρ for all stages. The generality of this pattern is unknown.

5.6 The Importance of Individual Stochasticity

The concept of individual stochasticity strikes to the heart of one of the most fundamental problems in population biology: the sources of variability among individuals. Heterogeneity—genuine differences among individuals—translates into differences in the age- or stage-specific vital rates to which they are subject.

Heterogeneity may arise from genetics, from physiological effects, from health conditions, or from unknown causes ("frailty," "quality"). Stochasticity results from the random outcomes of probabilistic processes. Markov chains naturally treat individual trajectories (i.e., individual lives) as realizations of an underlying stochastic process, and so much of this chapter has been focused on the analysis of individual stochasticity. The distinction is particularly important in evolutionary demography, where variance in lifetime reproductive output is routinely treated as variance in fitness, or a component of fitness. See Sect. 5.4.4 for some recent work on this problem.

Individual stochasticity is an important component of demography, for both human and non-human populations. It complements environmental stochasticity (externally imposed random changes in vital rates) and demographic stochasticity (randomness in the growth of populations due to stochastic survival and reproduction) (Caswell and Vindenes 2018). Individual stochasticity reflects randomness in the pathways that individuals take through the life cycle. It expresses itself in inter-individual variation in occupancy times, longevity, lifetime reproductive output, and other outcomes. The availability of methods based on Markov chains promises to change the way population biologists approach the analysis of variance among individuals (Caswell 2011; Tuljapurkar et al. 2009; Steiner and Tuljapurkar 2012; van Daalen and Caswell 2015; van Daalen and Caswell 2017).

5.7 Discussion

Taking advantage of the Markov chain formulation of the life cycle opens up a wealth of demographic information. The age-classified information extracted from a stage-classified model can form a valuable component of behavioral studies, especially if the model (like the right whale example) includes reproductive behavior as part of the life cycle structure. Longevity provides a powerful way to compare mortality schedules among species, populations, or environmental conditions, but it has been inaccessible to stage-classified analysis prior to the development of Markov chain methods. The generation time characterizes an important population time scale, with implications in conservation (IUCN Species Survival Commission 2001), but there has been no way to compute it from stage-classified models.

Stage-classified life cycles may have consequences that are not yet appreciated, but must be considered when interpreting the results. For example, any stage-classified model eventually leads to an age-independent mortality rate (Horvitz and Tuljapurkar 2008), and so is of limited use in the study of senescence. This fact has consequences for life expectancy and variance in longevity that are not well understood (at least by me). For the right whale, expected longevity at birth is 32 years with a standard deviation of 34 years. It is unlikely that there are appreciable numbers of whales alive at even one standard deviation above this mean. The high survival probability and the assumption of age-independence lead to the high standard deviation. Those of us who work with stage-classified models are

accustomed to this, but discount its importance because it (often) has little effect on λ. It will be important to determine the stochastic consequences of simplifying assumptions in the life cycle graph.

This chapter does not begin to exhaust the information that can be extracted from the Markov chain formulation of a stage-classified model. Three examples of particular interest are the occupancy of sets of states, the problem of competing risks, and the calculation of passage times. It is often of interest to calculate the statistics of occupancy of sets of states (e.g., all reproductive classes, or all stages in some particular health condition). We have seen how to calculate the moments of the occupancy time of single states. The mean occupancy time of a set of states is the sum of the mean occupancy times of each state, but that is not true for the variance or higher moments. Roth and Caswell (2018) derived a general expression for all the statistics, and the complete distribution, of occupancy time for any set of states. If more than one absorbing state exists (e.g., death at different stages, or from different causes), then the risks of absorbtion compete, because an individual can only be absorbed (i.e., die) once. It is possible to calculate the probability of absorbtion in each state, and to explore the effects of changing one risk on the probability of experiencing another (Caswell and Ouellette 2018). Passage times refer to the time required to get from one stage to another in the life cycle. An important passage time is the birth interval: the time from one birth to the next. This can only be calculated for individuals that do reproduce a second time (otherwise the interval is infinite), and so it requires developing a chain that is conditional on successfully reaching the reproductive state (Caswell 2001). In species that produce only one or a few offspring, reproduction cannot be adjusted in response to the environment by changing offspring number, and so changes in the birth interval are particularly important in such species.

A Appendix: Derivations

This appendix contains step-by-step derivations of many of the results in this chapter, especially for sensitivities. Taking advantage of the freedom from length limits, I have tried to show the derivations step-by-step. Recall the definitions of the Hadamard product

$$\mathbf{A} \circ \mathbf{B} = \left(a_{ij} b_{ij} \right), \tag{5.84}$$

the Kronecker product

$$\mathbf{A} \otimes \mathbf{B} = \left(a_{ij} \mathbf{B} \right), \tag{5.85}$$

the vec operator

$$\text{vec}\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a \\ c \\ b \\ d \end{pmatrix}, \tag{5.86}$$

and Roth's theorem

$$\text{vec}\left(\mathbf{ABC}\right) = \left(\mathbf{C}^{\mathsf{T}} \otimes \mathbf{A}\right)\text{vec}\,\mathbf{B}. \tag{5.87}$$

A.1   Variance in Occupancy Times

The occupancy time in transient state i, starting from transient state j, is ν_{ij}. The matrix of variances of the ν_{ij} is

$$\mathbf{V} = \left(V(\nu_{ij})\right) = \left(2\mathbf{N}_{\text{dg}} - \mathbf{I}_s\right)\mathbf{N} - \mathbf{N} \circ \mathbf{N} \tag{5.88}$$

(Caswell 2006, derived from Theorem 3.1 of Iosifescu 1980) where $\mathbf{N}_{\text{dg}}$ is a matrix with the diagonal elements of $\mathbf{N}$ on its diagonal and zeros elsewhere; it can be written

$$\mathbf{N}_{\text{dg}} = \mathbf{I}_s \circ \mathbf{N} \tag{5.89}$$

Differentiating both sides of (5.88) gives

$$d\mathbf{V} = 2(\mathbf{I}_s \circ d\mathbf{N})\mathbf{N} + 2(\mathbf{I}_s \circ \mathbf{N})(d\mathbf{N}) - d\mathbf{N}$$
$$-(d\mathbf{N}) \circ \mathbf{N} - \mathbf{N} \circ (d\mathbf{N}) \tag{5.90}$$

The next step is to apply the vec operator to both sides. The vec of a Hadamard product can be written in two ways:

$$\text{vec}\left(\mathbf{A} \circ \mathbf{B}\right) = \mathcal{D}\left(\text{vec}\,\mathbf{A}\right)\text{vec}\,\mathbf{B} = \mathcal{D}\left(\text{vec}\,\mathbf{B}\right)\text{vec}\,\mathbf{A}. \tag{5.91}$$

Using this result and Roth's theorem (5.87) gives

$$d\text{vec}\,\mathbf{V} = 2\left(\mathbf{N}^{\mathsf{T}} \otimes \mathbf{I}_s\right)\mathcal{D}\left(\text{vec}\,\mathbf{I}_s\right)d\text{vec}\,\mathbf{N} + 2\left[\mathbf{I}_s \otimes (\mathbf{I} \circ \mathbf{N})\right]d\text{vec}\,\mathbf{N}$$

$$-d\text{vec}\,\mathbf{N} - 2\mathcal{D}\left(\text{vec}\,\mathbf{N}\right)d\text{vec}\,\mathbf{N} \tag{5.92}$$

Factoring out $d\text{vec}\,\mathbf{N}$ and using the chain rule gives the final result

$$\frac{d\text{vec}\,\mathbf{V}}{d\boldsymbol{\theta}^{\mathsf{T}}} = \left[2 \left(\mathbf{N}^{\mathsf{T}} \otimes \mathbf{I}_s \right) \mathcal{D} \left(\text{vec}\,\mathbf{I}_s \right) + 2 \left(\mathbf{I}_s \otimes \mathbf{N}_{\text{dg}} \right) \right.$$
$$\left. - \mathbf{I}_{s^2} - 2 \mathcal{D} \left(\text{vec}\,\mathbf{N} \right) \right] \frac{d\text{vec}\,\mathbf{N}}{d\boldsymbol{\theta}^{\mathsf{T}}} \tag{5.93}$$

A.2   Life Expectancy

Let η_i be the time to absorbtion (i.e., death) of an individual currently in stage i. The vector $E(\boldsymbol{\eta})$ of expected values of the η_i satisfies

$$E(\boldsymbol{\eta})^{\mathsf{T}} = \mathbf{1}^{\mathsf{T}}\mathbf{N} \tag{5.94}$$

where $\mathbf{1}$ is a vector of ones. Differentiating both sides gives

$$dE(\boldsymbol{\eta})^{\mathsf{T}} = \mathbf{1}^{\mathsf{T}}(d\mathbf{N}) \tag{5.95}$$

Applying the vec operator gives

$$dE(\boldsymbol{\eta}) = \left(\mathbf{I}_s \otimes \mathbf{1}^{\mathsf{T}} \right) d\text{vec}\,\mathbf{N} \tag{5.96}$$

Applying the identification theorem and the chain rule, and using (5.16) for the sensitivity of the fundamental matrix, gives

$$\frac{dE(\boldsymbol{\eta})}{d\boldsymbol{\theta}^{\mathsf{T}}} = \left(\mathbf{I}_s^{\mathsf{T}} \otimes \mathbf{1}^{\mathsf{T}} \right) \left(\mathbf{N}^{\mathsf{T}} \otimes \mathbf{N} \right) \frac{d\text{vec}\,\mathbf{U}}{d\boldsymbol{\theta}^{\mathsf{T}}} \tag{5.97}$$

This gives the derivative of the entire vector of life expectancies. Suppose that stage 1 corresponds to birth. The life expectancy at birth is then

$$E(\eta_1) = \mathbf{1}^{\mathsf{T}}\mathbf{N}\mathbf{e}_1 \tag{5.98}$$

where $\mathbf{e}_1$ is a vector with 1 in the first position and zeros elsewhere. Following the same derivation gives

$$\frac{dE(\eta_1)}{d\boldsymbol{\theta}^{\mathsf{T}}} = \left(\mathbf{e}_1^{\mathsf{T}} \otimes \mathbf{1}^{\mathsf{T}} \right) \left(\mathbf{N}^{\mathsf{T}} \otimes \mathbf{N} \right) \frac{d\text{vec}\,\mathbf{U}}{d\boldsymbol{\theta}^{\mathsf{T}}} \tag{5.99}$$

A.3   Variance in Longevity

The variance of the time to absorbtion satisfies

$$V(\boldsymbol{\eta})^{\mathsf{T}} = \mathbf{1}^{\mathsf{T}}\mathbf{N}\,(2\mathbf{N} - \mathbf{I}) - E\left(\boldsymbol{\eta}^{\mathsf{T}}\right) \circ E\left(\boldsymbol{\eta}^{\mathsf{T}}\right) \tag{5.100}$$

(Caswell 2006, derived from Theorem 3.2 of Iosifescu 1980). Differentiating gives

$$dV(\boldsymbol{\eta})^{\mathsf{T}} = 2\mathbf{1}^{\mathsf{T}}(d\mathbf{N})\mathbf{N} + 2\mathbf{1}^{\mathsf{T}}\mathbf{N}(d\mathbf{N}) - \mathbf{1}^{\mathsf{T}}(d\mathbf{N}) - 2E\left(\boldsymbol{\eta}^{\mathsf{T}}\right) \circ dE\left(\boldsymbol{\eta}^{\mathsf{T}}\right) \tag{5.101}$$

Applying the vec operator and Roth's theorem (5.87), using (5.91) for the vec of the Hadamard product, gives

$$dV(\boldsymbol{\eta}) = \left[2\left(\mathbf{N}^{\mathsf{T}} \otimes \mathbf{1}^{\mathsf{T}}\right) + 2\left(\mathbf{I}_s \otimes \mathbf{1}^{\mathsf{T}}\mathbf{N}\right)\right.$$
$$\left. - \left(\mathbf{I}_s \otimes \mathbf{1}^{\mathsf{T}}\right)\right]d\mathrm{vec}\,\mathbf{N} - 2\mathcal{D}\left(E(\boldsymbol{\eta})\right)\,dE(\boldsymbol{\eta}) \tag{5.102}$$

Substituting (5.96) for $dE(\boldsymbol{\eta})$ gives

$$dV(\boldsymbol{\eta}) = \left[2\left(\mathbf{N}^{\mathsf{T}} \otimes \mathbf{1}^{\mathsf{T}}\right) + 2\left(\mathbf{I}_s \otimes \mathbf{1}^{\mathsf{T}}\mathbf{N}\right)\right.$$
$$\left. - \left(\mathbf{I}_s \otimes \mathbf{1}^{\mathsf{T}}\right) - 2\mathcal{D}\left(E(\boldsymbol{\eta})\right)\left(\mathbf{I}_s \otimes \mathbf{1}^{\mathsf{T}}\right)\right]d\mathrm{vec}\,\mathbf{N} \tag{5.103}$$

Using (5.16) for the sensitivity of $\mathbf{N}$, the identification theorem, and the chain rule finally leads to

$$\frac{dV(\boldsymbol{\eta})}{d\boldsymbol{\theta}^{\mathsf{T}}} = \left[2\left(\mathbf{N}^{\mathsf{T}} \otimes \mathbf{1}^{\mathsf{T}}\right) + 2\left(\mathbf{I}_s \otimes \mathbf{1}^{\mathsf{T}}\mathbf{N}\right)\right.$$
$$\left. - \left(\mathbf{I}_s \otimes \mathbf{1}^{\mathsf{T}}\right) - 2\mathcal{D}\left(E(\boldsymbol{\eta})\right)\left(\mathbf{I}_s \otimes \mathbf{1}^{\mathsf{T}}\right)\right]\left(\mathbf{N}^{\mathsf{T}} \otimes \mathbf{N}\right)\frac{d\mathrm{vec}\,\mathbf{U}}{d\boldsymbol{\theta}^{\mathsf{T}}} \tag{5.104}$$

A.4   Net Reproductive Rate

The net reproductive rate R_0 is given by the dominant eigenvalue of $\mathbf{FN}$. Let $\mathbf{y}$ and $\mathbf{x}$ be the right and left eigenvectors, respectively, of $\mathbf{FN}$, corresponding to R_0. The matrix calculus version of the standard eigenvalue perturbation result (e.g., Caswell 1978) gives

$$dR_0 = \mathbf{x}^{\mathsf{T}}d(\mathbf{FN})\mathbf{y}$$
$$= \mathbf{x}^{\mathsf{T}}\left[(d\mathbf{F})\mathbf{N} + \mathbf{F}(d\mathbf{N})\right]\mathbf{y} \tag{5.105}$$

Applying the vec operator to both sides gives

$$d R_0 = \left(\mathbf{y}^\mathsf{T} \mathbf{N}^\mathsf{T} \otimes \mathbf{x}^\mathsf{T} \right) d\mathrm{vec}\, \mathbf{F} + \left(\mathbf{y}^\mathsf{T} \otimes \mathbf{x}^\mathsf{T} \mathbf{F} \right) d\mathrm{vec}\, \mathbf{N} \tag{5.106}$$

Applying the chain rule and the result (5.14) for $d\mathrm{vec}\, \mathbf{N}$ gives the sensitivity of R_0 in terms of effects of the parameter vector $\boldsymbol{\theta}$ on the fertility matrix $\mathbf{F}$ and the transient matrix $\mathbf{U}$:

$$\frac{d R_0}{d\boldsymbol{\theta}^\mathsf{T}} = \left(\mathbf{y}^\mathsf{T} \mathbf{N}^\mathsf{T} \otimes \mathbf{x}^\mathsf{T} \right) \frac{d\mathrm{vec}\, \mathbf{F}}{d\boldsymbol{\theta}^\mathsf{T}} + \left(\mathbf{y}^\mathsf{T} \otimes \mathbf{x}^\mathsf{T} \mathbf{F} \right) \left(\mathbf{N}^\mathsf{T} \otimes \mathbf{N} \right) \frac{d\mathrm{vec}\, \mathbf{U}}{d\boldsymbol{\theta}^\mathsf{T}} \tag{5.107}$$

A.5 Cohort Generation Time

To derive the cohort generation time, we begin at time $t = 0$ with an individual newly born in stage j. This tiny cohort is described by an initial vector $\mathbf{e}_j$. The expected survivors of this cohort at time t are $\mathbf{U}^t \mathbf{e}_j$. The expected offspring produced by these survivors at time t are $\mathbf{F} \mathbf{U}^t \mathbf{e}_j$. Summing over the lifetime of the cohort gives a vector of expected lifetime reproduction, of all types of offspring,

$$E(\text{total offspring}) = \sum_{t=0}^{\infty} \mathbf{F} \mathbf{U}^t \mathbf{e}_j$$

$$= \mathbf{F} \left(\sum_{t=0}^{\infty} \mathbf{U}^t \right) \mathbf{e}_j$$

$$= \mathbf{F} \mathbf{N} \mathbf{e}_j \tag{5.108}$$

Let $\mathbf{m}^{(j)}(t)$ be the vector of offspring production at time t, expressed as a *proportion* of the lifetime total of the individual starting in stage j. Then

$$\mathbf{m}^{(j)}(t) = \mathcal{D} \left(\mathbf{F} \mathbf{N} \mathbf{e}_j \right)^{-1} \left(\mathbf{F} \mathbf{U}^t \mathbf{e}_j \right) \tag{5.109}$$

If no offspring of some stage, say stage i, are produced, then set $m_i^{(j)}(t) = 0$.

The cohort generation time $\boldsymbol{\mu}^{(j)}$ is the expectation of the distribution defined by $\mathbf{m}^{(j)}(t)$:

$$\boldsymbol{\mu}^{(j)} = \sum_{x=0}^{\infty} x \mathbf{m}^{(j)}(x)$$

$$= \sum_x \mathcal{D} \left(\mathbf{F} \mathbf{N} \mathbf{e}_j \right)^{-1} x \mathbf{F} \mathbf{U}^x \mathbf{e}_j$$

$$= \mathcal{D} \left(\mathbf{F} \mathbf{N} \mathbf{e}_j \right)^{-1} \mathbf{F} \left(\sum_x x \mathbf{U}^x \right) \mathbf{e}_j. \tag{5.110}$$

The summation can be simplified

$$\sum_x x\mathbf{U}^x = \mathbf{0} + \mathbf{U} + 2\mathbf{U}^2 + 3\mathbf{U}^3 + \cdots$$

$$= \mathbf{U}\left[\mathbf{0} + \mathbf{U} + 2\mathbf{U}^2 + \cdots + \mathbf{I} + \mathbf{U} + \mathbf{U}^2 + \cdots\right]$$

$$= \mathbf{U}\left[\mathbf{N} + \sum_x x\mathbf{U}^x\right]. \tag{5.111}$$

Solving this gives

$$\sum_x x\mathbf{U}^x = \mathbf{NUN}. \tag{5.112}$$

Putting all the pieces together gives the generation time

$$\boldsymbol{\mu}^{(j)} = \mathcal{D}\left(\mathbf{FNe}_j\right)^{-1} \mathbf{FNUNe}_j. \tag{5.113}$$

A.5.1 Sensitivity of Generation Time

To differentiate (5.113) may seem complicated. To make life easier, define some notation,

$$\mathbf{X} = \mathcal{D}\left(\mathbf{FNe}_j\right) \tag{5.114}$$

$$\mathbf{r} = \mathbf{FNUNe}_j \tag{5.115}$$

in terms of which (5.113)

$$\boldsymbol{\mu}^{(j)} = \mathbf{X}^{-1}\mathbf{r}. \tag{5.116}$$

Differentiate,

$$d\boldsymbol{\mu}^{(j)} = d\left(\mathbf{X}^{-1}\right)\mathbf{r} + \mathbf{X}^{-1}d\mathbf{r} \tag{5.117}$$

and apply the vec operator

$$d\boldsymbol{\mu}^{(j)} = \left(\mathbf{r}^{\mathsf{T}} \otimes \mathbf{I}\right)d\mathrm{vec}\,\mathbf{X}^{-1} + \mathbf{X}^{-1}d\mathrm{vec}\,\mathbf{r}. \tag{5.118}$$

The same steps that led to Eq. (5.14) for $d\mathrm{vec}\,\mathbf{N}$, and noting that $\mathbf{X}$ is symmetric, leads to

$$d\mathrm{vec}\,\mathbf{X}^{-1} = -\left(\mathbf{X}^{-1} \otimes \mathbf{X}^{-1}\right)d\mathrm{vec}\,\mathbf{X}. \tag{5.119}$$

The differential of vec $\mathbf{X}$ is obtained by writing

$$\mathbf{X} = \mathbf{I} \circ \left(\mathbf{FNe}_j \mathbf{1}^\mathsf{T}\right). \tag{5.120}$$

Differentiating and using the rule (5.91) for the vec of a Hadamard product gives

$$d\,\mathrm{vec}\,\mathbf{X} = \mathcal{D}\,(\mathrm{vec}\,\mathbf{I})\left[\left(\mathbf{1e}_j^\mathsf{T}\mathbf{N}^\mathsf{T} \otimes \mathbf{I}\right) d\,\mathrm{vec}\,\mathbf{F} + \left(\mathbf{1e}_j \otimes \mathbf{F}\right) d\,\mathrm{vec}\,\mathbf{N}\right] \tag{5.121}$$

Differentiating $\mathbf{r}$ and applying the vec operator gives

$$d\,\mathrm{vec}\,\mathbf{r} = \left[\left(\mathbf{NUNe}_j\right)^\mathsf{T} \otimes \mathbf{I}\right] d\,\mathrm{vec}\,\mathbf{F} + \left[\left(\mathbf{UNe}_j\right)^\mathsf{T} \otimes \mathbf{F}\right] d\,\mathrm{vec}\,\mathbf{N}$$

$$+ \left[\left(\mathbf{Ne}_j\right)^\mathsf{T} \otimes \mathbf{FN}\right] d\,\mathrm{vec}\,\mathbf{U} + \left[\mathbf{e}_j^\mathsf{T} \otimes \mathbf{fNU}\right] d\,\mathrm{vec}\,\mathbf{N} \tag{5.122}$$

Whew!

Finally, substituting (5.119), (5.121) and (5.122) into (5.117), we obtain

$$\frac{d\boldsymbol{\mu}^{(j)}}{d\boldsymbol{\theta}^\mathsf{T}} = -\left(\mathbf{r}^\mathsf{T} \otimes \mathbf{I}\right)\left(\mathbf{X}^{-1} \otimes \mathbf{X}^{-1}\right)\mathcal{D}\,(\mathrm{vec}\,\mathbf{I})$$

$$\times \left[\left(\mathbf{1e}_j^\mathsf{T}\mathbf{N}^\mathsf{T} \otimes \mathbf{I}\right)\frac{d\,\mathrm{vec}\,\mathbf{F}}{d\boldsymbol{\theta}^\mathsf{T}} + \left(\mathbf{1e}_j \otimes \mathbf{F}\right)\frac{d\,\mathrm{vec}\,\mathbf{N}}{d\boldsymbol{\theta}^\mathsf{T}}\right]$$

$$+ \left\{\left[\left(\mathbf{NUNe}_j\right)^\mathsf{T} \otimes \mathbf{I}\right]\frac{d\,\mathrm{vec}\,\mathbf{F}}{d\boldsymbol{\theta}^\mathsf{T}} + \left[\left(\mathbf{UNe}_j\right)^\mathsf{T} \otimes \mathbf{F}\right]\frac{d\,\mathrm{vec}\,\mathbf{N}}{d\boldsymbol{\theta}^\mathsf{T}}\right.$$

$$\left. + \left[\left(\mathbf{Ne}_j\right)^\mathsf{T} \otimes \mathbf{FN}\right]\frac{d\,\mathrm{vec}\,\mathbf{U}}{d\boldsymbol{\theta}^\mathsf{T}} + \left[\mathbf{e}_j^\mathsf{T} \otimes \mathbf{FNU}\right]\frac{d\,\mathrm{vec}\,\mathbf{N}}{d\boldsymbol{\theta}^\mathsf{T}}\right\}. \tag{5.123}$$

This may be an impressive formula, but it is straightforward to compute, given the derivatives of $\mathbf{U}$, $\mathbf{F}$, and $\mathbf{N}$ with respect to $\boldsymbol{\theta}$.

Bibliography

Bacaër, N. 2009. Periodic matrix population models: growth rate, basic reproduction number, and entropy. Bulletin of Mathematical Biology **71**:1781–1792.

Carey, J. R. 2003. Longevity: The biology and demography of life span. Princeton University Press, Princeton, New Jersey, USA.

Caswell, H. 1978. A general formula for the sensitivity of population growth rate to changes in life history parameters. Theoretical Population Biology **14**:215–230.

Caswell, H. 1989. Matrix Population Models: Construction, Analysis, and Interpretation. 1st edition. Sinauer Associates, Sunderland, MA, USA.

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. 2009. Stage, age and individual stochasticity in demography. Oikos **118**:1763–1782.

Caswell, H. 2011. Beyond R_0: Demographic models for variability of lifetime reproductive output. PloS ONE **6**:e20809.

Caswell, H., A. M. de Roos, R. M. Nisbet, and S. Tuljapurkar, 1997. Structured population models: many methods, a few general principles. Pages 3–18 *in* Structured population models in marine, terrestrial and freshwater systems. Chapman and Hall, New York, New York, USA.

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 M. Fujiwara. 2004. Beyond survival estimation: mark-recapture, matrix population models, and population dynamics. Animal Biodiversity and Conservation **27**:471–488.

Caswell, H., M. Fujiwara, and S. Brault. 1999. Declining survival probability threatens the North Atlantic right whale. Proceedings of the National Academy of Sciences **96**:3308–3313.

Caswell, H., and T. N. Kaye. 2001. Stochastic demography and conservation of an endangered perennial plant (*Lomatium bradshawii*) in a dynamic fire regime. Advances in Ecological Research **32**:1–51.

Caswell, H., and F. A. Kluge. 2015. Demography and the statistics of lifetime economic transfers under individual stochasticity. Demographic Research **32**:563–588.

Caswell, H., and N. Ouellette, 2018. Cause-of-death analysis: matrix formulation and sensitivity analysis. In prep.

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

Caswell, H., and Y. Vindenes. 2018. Demographic variance in heterogeneous populations: Matrix models and sensitivity analysis. Oikos **127**:648–663.

Clutton-Brock, T. H. 1988. Reproductive success: studies of individual variation in contrasting breeding systems. University of Chicago Press, Chicago, Illinois.

Coale, A. J. 1972. The growth and structure of human populations: a mathematical approach. Princeton University Press, Princeton, New Jersey.

Cochran, M. E., and S. Ellner. 1992. Simple methods for calculating age-based life history parameters for stage-structured populations. Ecological Monographs **62**:345–364.

Crone, M. J., and S. D. Kraus. 1990. Right whale (*Eubalaena glacialis*) in the Western North Atlantic: a catalog of identied individuals. New England Aquarium, Boston, Massachusetts.

Crow, J. F. 1958. Some possibilities for measuring selection intensities in man. Human Biology **30**:1–13.

Cushing, J. M., and A. S. Ackleh. 2012. A net reproductive number for periodic matrix models. Journal of Biological Dynamics **6**:166–188.

Cushing, J. M., and Y. Zhou. 1994. The net reproductive value and stability in matrix population models. Natural Resources Modeling **8**:297–333.

de Camino-Beck, T., and M. A. Lewis. 2007. A new method for calculating net reproductive rate from graph reduction with applications to the control of invasive species. Bulletin of Mathematical Biology **69**:1341–1354.

de Camino-Beck, T., and M. A. Lewis. 2008. On net reproductive rate and the timing of reproductive output. American Naturalist **172**:128–139.

DeAngelis, D. L., and L. J. Gross. 1992. Individual based models and approaches in ecology. Chapman and Hall, New York, New York.

Diekmann, O., J. A. P. Heesterbeek, and J. A. J. Metz. 1990. On the definition and the computation of the basic reproduction number R_0 in models for infectious diseases in heterogeneous population. Journal of Mathematical Biology **28**:365–382.

Edwards, R. D. 2011. Changes in world inequality in length of life: 1970–2000. Population and Development Review **37**:499–528.

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.

Ellner, S. P., D. Z. Childs, and M. Rees. 2016. Data-driven modelling of structured populations. Springer, New York, New York.

Endler, J. A. 1986. Natural selection in the wild. Princeton University Press, Princeton, New Jersey.

Feichtinger, G. 1971. Stochastische Modelle demographischer Prozesse. Lecture Notes in Economics and Mathematical Systems, Springer-Verlag, Berlin.

Feichtinger, G. 1973. Markovian models for some demographic processes. Statistiche Hefte **14**:310–334.

Ferriére, R., and M. Gatto. 1993. Chaotic population dynamics can result from natural selection. Proceedings of the Royal Society of London B **251**:33–38.

Frühwirth-Schnatter, S. 2006. Finite mixture and Markov switching models. Springer-Verlag, New York, New York.

Fujiwara, M., and H. Caswell. 2001. Demography of the endangered North Atlantic right whale. Nature **414**:537–541.

Fujiwara, M., and H. Caswell. 2002. Estimating population projection matrices from multi-stage mark-recapture data. Ecology **83**:3257–3265.

Goldman, N. 1994. Social factors and health: the causation-selection issue revisited. Proceedings of the National Academy of Sciences USA **91**:1251–1255.

Hamilton, W. D. 1966. The moulding of senescence by natural selection. Journal of Theoretical Biology **12**:12–45.

Hartemink, N., and H. Caswell. 2018. Variance in animal longevity: contributions of heterogeneity and stochasticity. Population Ecology **60**:89–99.

Hartemink, N., T. I. Missov, and H. Caswell. 2017. Stochasticity, heterogeneity, and variance in longevity in human populations. Theoretical Population Biology **114**:107–117.

Henderson, H. V., and S. R. Searle. 1981. The vec-permutation matrix, the vec operator and Kronecker products: a review. Linear and Multilinear Algebra **9**:271–288.

Hernandez-Suarez, C., J. Rabinovich, and K. Hernandez. 2012. The long-run distribution of births across environments under environmental stochasticity and its use in the calculation of unconditional life-history parameters. Theoretical Population Biology **82**:264–274.

Hoem, J. M. 1969. Markov chain models in life insurance. Sonderdruck aus Blatter der Deutschen Gelellschaft fur Versicherungsmathematik **9**:91–107.

Horvitz, C. C., and D. W. Schemske, 1986. Seed dispersal and environmental heterogeneity in a neotropical herb: a model of population and patch dynamics. Pages 169–186 *in* A. Estrada and T. H. Fleming, editors. Frugivores and seed dispersal. Dr. W. Junk, Dordrecht, Netherlands.

Horvitz, C. C., and S. Tuljapurkar. 2008. Stage dynamics, period survival, and mortality plateaus. American Naturalist **172**:203–215.

Howard, R. A. 1960. Dynamic programming and Markov processes. Wiley, New York, New York.

Hunter, C. M., and H. Caswell. 2005*a*. Selective harvest of sooty shearwater chicks: effects on population dynamics and sustainability. Journal of Animal Ecology **74**:589–600.

Hunter, C. M., and H. Caswell. 2005*b*. The use of the vec-permutation matrix in spatial matrix population models. Ecological Modelling **188**:15–21.

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

IUCN Species Survival Commission. 2001. IUCN Red List categories and criteria: version 3.1. IUCN, Gland, Switzerland and Cambridge, UK.

Jenouvrier, S., L. M. Aubry, C. Barbraud, H. Weimerskirch, and H. Caswell. 2018. Interacting effects of unobserved heterogeneity and individual stochasticity in the life cycle of the Southern fulmar. Journal of Animal Ecology **87**:212–222.

Kaye, T. N., K. L. Pendergrass, K. Finley, and J. B. Kauffman. 2001. The effect of fire on the population viability of an endangered prairie plant. Ecological Applications **11**:1366–1380.

Kaye, T. N., and D. A. Pyke. 2003. The effect of stochastic technique on estimates of population viability from transition matrix models. Ecology pages 1464–1476.

Kemeny, J. G., and J. L. Snell. 1976. Finite Markov Chains. Second edition. Undergraduate Texts in Mathematics, Springer-Verlag, New York, New York, USA.

Keyfitz, N. 1967. Reconciliation of population models: matrix, integral equation and partial fraction. Journal of the Royal Statistical Society **130**:61–83.

Keyfitz, N. 1971. Linkages of intrinsic to age-specific rates. Journal of the American Statistical Association **66**:275–281.

Keyfitz, N., and H. Caswell. 2005. Applied mathematical demography. 3rd edition. Springer, New York, New York.

Kozlowski, J. 1999. Adaptation: a life history perspective. Oikos **86**:185–194.

Kraus, S. D., M. W. Brown, H. Caswell, C. W. Clark, M. Fujiwara, P. K. Hamilton, R. D. Kenney, A. R. Knowlton, S. Landry, C. A. Mayo, et al. 2005. North Atlantic right whales in crisis. Science **309**:561–562.

Kraus, S. D., and R. M. Rolland. 2007. The urban whale: North Atlantic right whales at the crossroads. Harvard University Press, Cambridge, Massachusetts, USA.

Land, K. C., and A. Rogers, editors. 1982. Multidimensional Mathematical Demography. Academic Press, New York, New York, USA.

Lande, R., and S. J. Arnold. 1983. The measurement of selection on correlated characters. Evolution pages 1210–1226.

Lebreton, J. 1996. Demographic models for subdivided populations: the renewal equation approach. Theoretical Population Biology **49**:291–313.

LeCorff, J., and C. C. Horvitz. 2005. The contrasting roles of chasmogamous and cleistogamous reproduction in population growth and population spread in an ant-dispersed neotropical herb *Calathea micans* (Marantaceae). Ecological Modelling **188**:41–51.

Lotka, A. J. 1939. Théorie analytique des associations biologiques. Part II. Analyse démographique avec application particulière à l'espèce humaine. Number 780 in Actualités Scientifiques et Industrielles, Hermann et Cie, Paris, France.

Magnus, J. R., and H. Neudecker. 1979. The commutation matrix: some properties and applications. Annals of Statistics **7**:381–394.

Metz, J. A. J., and O. Diekmann. 1986. The dynamics of physiologically structured populations. Springer-Verlag, Berlin, Germany.

Metz, J. A. J., R. M. Nisbet, and S. A. H. Geritz. 1992. How should we define 'fitness' for general ecological scenarios? Trends in Ecology and Evolution **7**:198–202.

Newton, I. 1989. Lifetime reproductive success in birds. Academic Press, San Diego, California.

Nisbet, R. M., and W. S. C. Gurney. 1982. Modelling fluctuating populations. Wiley, New York, New York, USA.

Otto, S. P., and T. Day. 2007. A biologist's guide to mathematical modeling in ecology and evolution. Princeton University Press, Princeton, New Jersey, USA.

Pascarella, J. B., and C. C. Horvitz. 1998. Hurricane disturbance and the population dynamics of a tropical understory shrub: megamatrix elasticity analysis. Ecology **79**:547–563.

Pollard, J. H. 1982. The expectation of life and its relationship to mortality. Journal of the Institute of Actuaries **109**:225–240.

Puterman, M. L. 1994. Markov decision processes: Discrete dynamic stochastic programming. John Wiley, New York, New York.

Reeves, R. R., T. D. Smith, and E. A. Josephson, 2007. Near-annihilation of a species: right whaling in the North Atlantic. Pages 39–74 *in* S. D. Kraus and R. M. Rolland, editors. The urban whale: North Atlantic right whales at the crossroads. Harvard University Press, Cambridge, Massachusetts, USA.

Rényi, A. 1970. Probability Theory. North-Holland, Amsterdam.

Rhodes, E. C. 1940. Population mathematics – I. Journal of the Royal Statistical Society **103**:61–89.

Robine, J.-M., C. Jagger, C. D. Mathers, E. M. Crimmins, and R. M. Suzman. 2003. Determining health expectancies. Wiley, Chichester, England.

Rogers, A. 1974. The multiregional net maternity function and multiregional stable growth. Demography **11**:473–481.

Roth, G., and H. Caswell. 2018. Occupancy time in sets of states for demographic models. Theoretical Population Biology **120**:62–77.

Sheskin, T. J. 2010. Markov chains and decision processes for engineers and managers. CRC Press.

Shkolnikov, V. M., E. E. Andreev, and A. Z. Begun. 2003. Gini coefficient as a life table function: computation from discrete data, decomposition of differences and empirical examples. Demographic Research **8**:305–358.

Steiner, U. K., and S. Tuljapurkar. 2012. Neutral theory for life histories and individual variability in fitness components. Proceedings of the National Academy of Sciences **109**:4684–4689.

Steiner, U. K., S. Tuljapurkar, and S. H. Orzack. 2010. Dynamic heterogeneity and life history variability in the kittiwake. Journal of Animal Ecology **79**:436–444.

Tuljapurkar, S. 1990. Population Dynamics in Variable Environments. Springer-Verlag, New York, New York, USA.

Tuljapurkar, S., and H. Caswell. 1997. Structured-Population Models in Marine, Terrestrial, and Freshwater Ecosystems. Chapman and Hall, New York, New York.

Tuljapurkar, S., and C. C. Horvitz. 2006. From stage to age in variable environments: life expectancy and survivorship. Ecology **87**:1497–1509.

Tuljapurkar, S., U. K. Steiner, and S. H. Orzack. 2009. Dynamic heterogeneity in life histories. Ecology Letters **12**:93–106.

van Daalen, S., and H. Caswell. 2015. Lifetime reproduction and the second demographic transition: Stochasticity and individual variation. Demographic Research **33**:561–588.

van Daalen, S. F., and H. Caswell. 2017. Lifetime reproductive output: individual stochasticity, variance, and sensitivity analysis. Theoretical Ecology **10**:355–374.

van den Driessche, P., and J. Watmough. 2002. Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. Mathematical Biosciences **180**:29–48.

Van Raalte, A. A., and H. Caswell. 2013. Perturbation analysis of indices of lifespan variability. Demography **50**:1615–1640.

Vaupel, J. W. 1986. How change in age-specific mortality affects life expectancy. Population Studies **40**:147–157.

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.

Willekens, F. 2014. Multistate Analysis of Life Histories with R. Springer, New York, New York.

Wilmoth, J. R., and S. Horiuchi. 1999. Rectangularization revisited: variability of age at death within human populations. Demography **36**:475–495.

Chapter 6
Age × Stage-Classified Models

6.1 Introduction

The first step in developing any kind of structured population model is choosing one or more variables in terms of which to describe the population structure. The job of these i-state variables is to encapsulate all the information about the past experience of an individual that is relevant to its future behavior (Metz and Diekmann 1986; Caswell 2001). Classical demography (for both humans and for non-human animals and plants) uses age as a i-state, but other, more biologically relevant criteria (e.g., size, developmental stage, parity, physiological condition, etc.) are now widely used in ecology, with age-classified models viewed as a special case.

However, it has long been recognized that cases exist where it is important to classify individuals by *both* age and stage.

1. Even in a stage-classified model, age still exists; every individual becomes older, by one unit of age, with the passage of each unit of time (e.g., Feichtinger 1971a; Caswell 2001, 2006, 2009; Tuljapurkar and Horvitz 2006; Horvitz and Tuljapurkar 2008). In these analyses, age dependence is implicit in the stage-classified model (see Chap. 5). Models that include both age and stage provide information that goes beyond this implicit dependence.
2. If the vital rates depend on both age and stage, only a model that includes both can reveal the joint action of age-and stage-specific processes (e.g., Goodman 1969). Such models, of course, require information on the joint age-dependence and stage-dependence of the vital rates, and thus are challenging to construct. A special case that has been extensively explored is the multi-regional case, in

---

H. Caswell, *Sensitivity Analysis: Matrix Methods in Demography and Ecology*, Demographic Research Monographs,
https://doi.org/10.1007/978-3-030-10534-1_6

which the stage variable describes spatial location (e.g., Rogers 1966; Lebreton 1996). Models that combine age and some measure of health or disability status are an important part of health demography (e.g., Willekens 2014; Peeters et al. 2002; Wu et al. 2006; Zhou et al. 2016).

This chapter presents a model framework in which individuals are classified by age and stage, using the vec-permutation matrix approach (so-called for the role that the vec-permutation matrix plays in rearranging age and stage categories in the population vector). This formalism was introduced by Hunter and Caswell (2005) for populations classified by stage and location, was used in Chap. 5 to classify individuals by stage and environmental state; it has also been applied to stage and infection status (Klepac and Caswell 2011), stage and age (Caswell 2012; Caswell et al. 2018), and age and frailty (Caswell 2014). Megamatrix models (e.g., Pascarella and Horvitz 1998; Horvitz and Tuljapurkar 2008) can be written using this approach, as can block-structured multiregional models (e.g., Rogers 1975; Lebreton 1996). Matrix models can describe both population dynamics and cohort dynamics. Population dynamics (population growth, age and stage structure, reproductive value) depend on both the transitions of extant individuals and the production of new individuals by reproduction. In contrast, cohort dynamics (survivorship, life expectancy, age at death, generation time) depend only on the fates of already existing individuals. This chapter describes both kinds of analysis. For a more complete review and treatment, see Caswell et al. (2018).

6.2   Model Construction

The construction and analysis of these models requires a number of different matrices and operators (some of the notation is collected in Table 6.1). Individuals are classified into stages $1, \ldots, s$ and age classes $1, \ldots, \omega$. The model treats the processes of moving among stages and moving among age classes as alternating. First, stage-specific demography operates to move individuals among stages and to produce new offspring, with rates appropriate to their ages. Then aging acts to move individuals to the next older age, and the process repeats.

Define a stage-classified projection matrix $\mathbf{A}_i$, of dimension $s \times s$, for each age class, $i = 1, \ldots, \omega$. Decompose $\mathbf{A}_i$ into

$$\mathbf{A}_i = \mathbf{U}_i + \mathbf{F}_i \tag{6.1}$$

where $\mathbf{U}_i$ contains the transition probabilities of extant individuals and $\mathbf{F}_i$ describes the generation of new individuals by reproduction.

Table 6.1 Mathematical notation used in this chapter. Dimensions are shown, where relevant, for matrices and vectors; s denotes the number of stages and ω the number of age classes

Quantity	Description	Dimension
$\mathbf{A}_i, \mathbf{F}_i, \mathbf{U}_i$	Stage-classified projection, fertility, and transition matrices for age class i	$s \times s$
$\mathbf{D}_U, \mathbf{D}_F$	Age transition matrices for individuals already present in the population and for new individuals produced by reproduction	$\omega \times \omega$
$\mathbb{A}, \mathbb{F}, \mathbb{U}, \mathbb{D}$	Block-diagonal matrices	$s\omega \times s\omega$
$\tilde{\mathbf{A}}, \tilde{\mathbf{U}}$, etc.	Age $\times$ stage matrices constructed from block-diagonal matrices using the vec-permutation matrix	$s\omega \times s\omega$
$\mathbf{K}_{s,\omega}, \mathbf{K}$	Vec-permutation matrix	$s\omega \times s\omega$
$\mathbf{I}_s$	Identity matrix	$s \times s$
$\mathbf{1}_s$	Vector of ones	$s \times 1$
$\mathbf{e}_i$	The ith unit vector, with a 1 in the ith entry and zeros elsewhere	Various
$\mathbf{E}_{ij}$	A matrix with a 1 in the (i, j) position, and zeros elsewhere	Various
$\otimes$	Kronecker product	
$\circ$	Hadamard, or element-by-element, product	
vec $\mathbf{X}$	The vec operator, which stacks the columns of a $m \times n$ matrix $\mathbf{X}$ into a $mn \times 1$ vector	
$\mathcal{D}(\mathbf{x})$	A diagonal matrix with $\mathbf{x}$ on the diagonal and zeros elsewhere	

Aging is described by two matrices, each of dimension $\omega \times \omega$ (shown here for 3×3, but easily generalized),

$$\mathbf{D}_U = \begin{pmatrix} 0\ 0\ 0 \\ 1\ 0\ 0 \\ 0\ 1\ 1 \end{pmatrix} \qquad \text{dimension } \omega \times \omega \tag{6.2}$$

$$\mathbf{D}_F = \begin{pmatrix} 1\ 1\ 1 \\ 0\ 0\ 0 \\ 0\ 0\ 0 \end{pmatrix} \qquad \omega \times \omega \tag{6.3}$$

The matrix $\mathbf{D}_U$ applies to extant individuals; such an individual advances to the next age class. I have set the (ω, ω) entry of $\mathbf{D}_U$ to 1, so that the last age class contains individuals of age ω and older. If this entry were set to 0, all individuals in the last age class would die. The matrix $\mathbf{D}_F$ applies to individuals newly created by reproduction; such newborn individuals are placed in the first age class, regardless of the age of their parents.

Using the matrices $\mathbf{A}_i$, $\mathbf{U}_i$, $\mathbf{F}_i$, $\mathbf{D}_U$, and $\mathbf{D}_F$, construct block-diagonal matrices, each of dimension $s\omega \times s\omega$. For example,

$$\mathbb{A} = \begin{pmatrix} \mathbf{A}_1 & & \\ & \ddots & \\ & & \mathbf{A}_\omega \end{pmatrix} \tag{6.4}$$

with similar structures for $\mathbb{U}$, $\mathbb{F}$, $\mathbb{D}_U$, and $\mathbb{D}_F$. These block-diagonal matrices can be written

$$\mathbb{A} = \sum_{i=1}^{\omega} (\mathbf{E}_{ii} \otimes \mathbf{A}_i) \tag{6.5}$$

$$\mathbb{U} = \sum_{i=1}^{\omega} (\mathbf{E}_{ii} \otimes \mathbf{U}_i) \tag{6.6}$$

$$\mathbb{F} = \sum_{i=1}^{\omega} (\mathbf{E}_{ii} \otimes \mathbf{F}_i) \tag{6.7}$$

$$\mathbb{D}_U = \mathbf{I}_s \otimes \mathbf{D}_U \tag{6.8}$$

$$\mathbb{D}_F = \mathbf{I}_s \otimes \mathbf{D}_F \tag{6.9}$$

where $\mathbf{E}_{ii}$ is of dimension $\omega \times \omega$.

If the demography is strictly stage-dependent, so that $\mathbf{A}_i = \mathbf{A}$, for $i = 1, \ldots, \omega$, then the block-diagonal matrices $\mathbb{A}$, $\mathbb{F}$, and $\mathbb{U}$ reduce to, e.g.,

$$\mathbb{A} = \mathbf{I}_\omega \otimes \mathbf{A} \tag{6.10}$$

with corresponding expressions for $\mathbb{F}$ and $\mathbb{U}$.

The state of the population at time t could be described by a 2-dimensional array

$$\boldsymbol{\mathcal{N}}(t) = \begin{pmatrix} n_{11} & \cdots & n_{1\omega} \\ \vdots & & \vdots \\ n_{s1} & \cdots & n_{s\omega} \end{pmatrix} (t) \qquad s \times \omega \tag{6.11}$$

where rows correspond to stages and columns to age classes. However, such a 2-dimensional array cannot be projected directly; instead, it is transformed to a vector,

$$\mathbf{n}(t) = \text{vec}\,\boldsymbol{\mathcal{N}}(t) = \begin{pmatrix} n_{11} \\ \vdots \\ n_{s1} \\ \hline \vdots \\ n_{1\omega} \\ \vdots \\ n_{s\omega} \end{pmatrix} (t) \qquad s\omega \times 1 \tag{6.12}$$

using the vec operator, which stacks the columns of the matrix one above the next. The vector $\mathbf{n}(t)$ created in this way contains the stages arranged within age classes. An alternative configuration, with ages arranged within stages, is obtained by applying the vec operator to $\mathcal{N}^\mathsf{T}$:

$$
\mathrm{vec}\,\mathcal{N}^\mathsf{T}(t) = \begin{pmatrix} n_{11} \\ \vdots \\ n_{1\omega} \\ \hline \vdots \\ \hline n_{s1} \\ \vdots \\ n_{s\omega} \end{pmatrix}(t) \qquad s\omega \times 1 \tag{6.13}
$$

The two vectors $\mathrm{vec}\,\mathcal{N}$ and $\mathrm{vec}\,\mathcal{N}^\mathsf{T}$ are related by the vec-permutation matrix, or commutation matrix, $\mathbf{K}$, (Henderson and Searle 1981),

$$
\mathrm{vec}\,\mathcal{N}^\mathsf{T} = \mathbf{K}_{s,\omega}\mathrm{vec}\,\mathcal{N} \tag{6.14}
$$

(see Sect. 2.2.3). Where no confusion seems likely to arise, we will suppress the subscripts and write $\mathbf{K}_{s,\omega}$ as $\mathbf{K}$. As with any permutation matrix, $\mathbf{K}^\mathsf{T} = \mathbf{K}^{-1}$.

The goal of the model is to project the age-stage vector $\mathbf{n} = \mathrm{vec}\,\mathcal{N}$ from t to $t+1$. The complete projection is given by

$$
\mathbf{n}(t+1) = \left(\mathbf{K}^\mathsf{T}\mathbb{D}_\mathrm{U}\mathbf{K}\mathbb{U} + \mathbf{K}^\mathsf{T}\mathbb{D}_\mathrm{F}\mathbf{K}\mathbb{F} \right)\mathbf{n}(t) \tag{6.15}
$$

This deserves some explanation. Consider the first term on the right hand side, $\mathbf{K}^\mathsf{T}\mathbb{D}_\mathrm{U}\mathbf{K}\mathbb{U}$. Reading from right to left, it first operates on the vector $\mathbf{n}(t)$ with the block diagonal matrix $\mathbb{U}$, which moves surviving extant individuals among stages without changing their age. Then the resulting vector is rearranged by the vec-permutation matrix $\mathbf{K}$ to group individuals by age classes within each stage. The block diagonal matrix $\mathbb{D}_\mathrm{U}$ then moves each surviving individual to the next older age class. Finally, $\mathbf{K}^\mathsf{T}$ rearranges the vector back to the stage-within-age arrangement of $\mathbf{n}(t)$.

The second term in (6.15), $\mathbf{K}^\mathsf{T}\mathbb{D}_\mathrm{F}\mathbf{K}\mathbb{F}$, carries out a similar sequence of transformations for the generation of new individuals. First, newborn individuals are produced according to the block-diagonal fertility matrix $\mathbb{F}$. The resulting vector is rearranged by the vec-permutation matrix, and then the matrix $\mathbb{D}_\mathrm{F}$ places all the newborn individuals into the first age class. Finally, $\mathbf{K}^\mathsf{T}$ rearranges the vector to the stage-within-age arrangement.

I will write the age × stage projection matrix in (6.15) as

$$\tilde{\mathbf{A}} = \left(\mathbf{K}^{\mathsf{T}} \mathbb{D}_{\mathrm{U}} \mathbf{K} \mathbb{U} + \mathbf{K}^{\mathsf{T}} \mathbb{D}_{\mathrm{F}} \mathbf{K} \mathbb{F} \right) \tag{6.16}$$

$$= \left(\tilde{\mathbf{U}} + \tilde{\mathbf{F}} \right) \tag{6.17}$$

The matrices $\tilde{\mathbf{A}}$, $\tilde{\mathbf{U}}$, and $\tilde{\mathbf{F}}$ that operate on the age-stage vector $\mathbf{n}$ are denoted with a tilde ($\tilde{\mathbf{A}}$, $\tilde{\mathbf{U}}$, $\tilde{\mathbf{F}}$); these matrices define the age × stage-classified model and can be subjected to all the usual demographic analyses.

6.3   Sensitivity Analysis

Age-stage models pose particular challenges for perturbation analysis, because interest naturally focuses on changes in the matrices $\mathbf{F}_i$ and $\mathbf{U}_i$ ($i = 1, \ldots, \omega$), which are deeply embedded within $\tilde{\mathbf{F}}$, $\tilde{\mathbf{U}}$, and $\tilde{\mathbf{A}}$.

Consider a generic dependent variable $\boldsymbol{\xi}$, which is a scalar- or vector-valued function of $\tilde{\mathbf{A}}$. In the examples to follow, $\boldsymbol{\xi}$ will be either the population growth rate λ or the joint distribution of age and stage at death in a cohort, but it could be any variable calculated from $\tilde{\mathbf{A}}$. Let $\boldsymbol{\theta}$ be a vector of parameters; these could be entries of the matrices, or lower-level parameters determining those entries. The goal of perturbation analysis is to obtain the derivative of $\boldsymbol{\xi}$ with respect to $\boldsymbol{\theta}$,

$$\frac{d\boldsymbol{\xi}}{d\boldsymbol{\theta}^{\mathsf{T}}} = \frac{d\boldsymbol{\xi}}{d\mathrm{vec}^{\mathsf{T}}\tilde{\mathbf{A}}} \, \frac{d\mathrm{vec}\,\tilde{\mathbf{A}}}{d\boldsymbol{\theta}^{\mathsf{T}}}. \tag{6.18}$$

The first term in (6.18) is the derivative of $\boldsymbol{\xi}$ with respect to the matrix $\tilde{\mathbf{A}}$. If, for example, $\boldsymbol{\xi}$ was the dominant eigenvalue λ, then this term would be the matrix calculus version of the well-known eigenvalue sensitivity equation.

The second term in (6.18) requires differentiating $\tilde{\mathbf{A}}$ with respect to the parameters that determine it. From (6.16), write

$$\tilde{\mathbf{A}} = \mathbf{Q}_{\mathrm{U}} \mathbb{U} + \mathbf{Q}_{\mathrm{F}} \mathbb{F} \tag{6.19}$$

where $\mathbf{Q}_{\mathrm{U}} = \mathbf{K}^{\mathsf{T}} \mathbb{D}_{\mathrm{U}} \mathbf{K}$ and $\mathbf{Q}_{\mathrm{F}} = \mathbf{K}^{\mathsf{T}} \mathbb{D}_{\mathrm{F}} \mathbf{K}$ are the (constant) matrix products appearing in the definition of $\tilde{\mathbf{U}}$ and $\tilde{\mathbf{F}}$ in (6.16).

Differentiating $\tilde{\mathbf{A}}$ in (6.19) gives

$$d\mathrm{vec}\,\tilde{\mathbf{A}} = (\mathbf{I}_{sw} \otimes \mathbf{Q}_{\mathrm{U}}) \, d\mathrm{vec}\,\mathbb{U} + (\mathbf{I}_{sw} \otimes \mathbf{Q}_{\mathrm{F}}) \, d\mathrm{vec}\,\mathbb{F} \tag{6.20}$$

This requires the differentials of $\mathbb{U}$ and $\mathbb{F}$. Differentiating $\mathbb{U}$ in (6.6) gives

$$d\mathbb{U} = \sum_{i=1}^{\omega} (\mathbf{E}_{ii} \otimes d\mathbf{U}_i) \tag{6.21}$$

Applying the vec operator to $d\mathbb{U}$ gives

$$d\text{vec}\,\mathbb{U} = \sum_{i=1}^{\omega} (\mathbf{E}_{ii} \otimes \mathbf{K} \otimes \mathbf{I}_s) \left(\text{vec}\,\mathbf{I}_\omega \otimes \mathbf{I}_{s^2}\right) d\text{vec}\,\mathbf{U}_i \tag{6.22}$$

using the results of Magnus and Neudecker (1985, Theorem 11); see also Klepac and Caswell (2011, Appendix B) on the derivative of the Kronecker product. Differentiation of $\mathbb{F}$ proceeds in the same fashion, yielding

$$d\text{vec}\,\mathbb{F} = \sum_{i=1}^{\omega} (\mathbf{E}_{ii} \otimes \mathbf{K} \otimes \mathbf{I}_s) \left(\text{vec}\,\mathbf{I}_\omega \otimes \mathbf{I}_{s^2}\right) d\text{vec}\,\mathbf{F}_i \tag{6.23}$$

In the special case where $\mathbb{U}$ and $\mathbb{F}$ are constructed from single stage-classified matrices $\mathbf{U}$ and $\mathbf{F}$, as in (6.10), Eqs. (6.22) and (6.23) simplify even further to

$$d\text{vec}\,\mathbb{U} = (\mathbf{I}_\omega \otimes \mathbf{K} \otimes \mathbf{I}_s) \left(\text{vec}\,\mathbf{I}_\omega \otimes \mathbf{I}_{s^2}\right) d\text{vec}\,\mathbf{U} \tag{6.24}$$

$$d\text{vec}\,\mathbb{F} = (\mathbf{I}_\omega \otimes \mathbf{K} \otimes \mathbf{I}_s) \left(\text{vec}\,\mathbf{I}_\omega \otimes \mathbf{I}_{s^2}\right) d\text{vec}\,\mathbf{F} \tag{6.25}$$

Substituting (6.22) and (6.23) into (6.20) and then substituting (6.20) into (6.18) yields the general result for the derivative

$$\frac{d\boldsymbol{\xi}}{d\boldsymbol{\theta}^{\mathsf{T}}} = \frac{d\boldsymbol{\xi}}{d\text{vec}^{\mathsf{T}}\tilde{\mathbf{A}}} \left[(\mathbf{I}_{sw} \otimes \mathbf{Q}_{\mathrm{U}}) \sum_{i=1}^{\omega} (\mathbf{E}_{ii} \otimes \mathbf{K} \otimes \mathbf{I}_s) \left(\text{vec}\,\mathbf{I}_\omega \otimes \mathbf{I}_{s^2}\right) \frac{d\text{vec}\,\mathbf{U}_i}{d\boldsymbol{\theta}^{\mathsf{T}}} \right]$$

$$+ \frac{d\boldsymbol{\xi}}{d\text{vec}^{\mathsf{T}}\tilde{\mathbf{A}}} \left[(\mathbf{I}_{sw} \otimes \mathbf{Q}_{\mathrm{F}}) \sum_{i=1}^{\omega} (\mathbf{E}_{ii} \otimes \mathbf{K} \otimes \mathbf{I}_s) \left(\text{vec}\,\mathbf{I}_\omega \otimes \mathbf{I}_{s^2}\right) \frac{d\text{vec}\,\mathbf{F}_i}{d\boldsymbol{\theta}^{\mathsf{T}}} \right] \tag{6.26}$$

Notice that (6.26) requires only three pieces of demographic information: the derivatives of $\mathbf{U}_i$ and $\mathbf{F}_i$ with respect to the parameters (whatever those may be in the case at hand) and the sensitivity of the dependent variable $\boldsymbol{\xi}$ (whatever that may be) to the elements of the matrix $\tilde{\mathbf{A}}$ from which it is calculated. All the other pieces of (6.26) are constants. Some of these constant matrices may be large, depending on s and ω, but they are very sparse; the sparse matrix technology available in MATLAB can be extremely useful in implementation. An alternative formulation of the differentials of the block matrices $\mathbb{U}$ and $\mathbb{F}$ is given in Caswell and van Daalen (2016).

6.4   Examples

Here we consider two examples of the sensitivity analysis of age-stage model to extract age-classified information from a stage-classified model. The first example will derive the sensitivity of the population growth rate λ, obtaining the sensitivity of λ to both age- and stage-specific survival, permitting examination of how selection pressures on senescence-inducing traits would vary from stage to stage. The second example is an analysis of the joint distribution of age and stage at death.

These examples are based on a stage-classified model (Parker 2000) for Scotch broom (*Cytisus scoparius*). Scotch broom is a large (up to 4 m tall) leguminous shrub, introduced into North America from Europe in the late nineteenth century. It is an invasive plant, considered a pest in the northwestern parts of North America. Stage-classified demographic models have been used to evaluate potential management policies for the plant (Parker 2000) and to investigate its potential for spatial spread (Neubert and Parker 2004).

The model contains seven stages (stage 1 = seeds, 2 = seedlings, 3 = juveniles, 4 = small adults, 5 = medium adults, 6 = large adults, 7 = extra-large adults), and parameters were estimated at a number of locations in Washington State. As is typical with many perennial plant species, survival is low for seeds and seedlings, but increases dramatically in larger stages. Parker's study presented estimated projection matrices for plants at the edge, at intermediate locations, and at the center of an invading stand. Plants near the center experience more crowding, with resulting reduced rates of survival, growth, and fertility.

6.4.1   Population Growth Rate and Selection Gradients

The population growth rate λ, the stable age or stage distribution **w**, and age or stage-specific reproductive value vector **v** are given by the dominant eigenvalue and corresponding right and left eigenvectors of the population projection matrix, respectively. In evolutionary demography, λ measures the fitness of a phenotype, in that it gives the eventual rate at which descendants of an individual with that phenotype will increase. The selection gradient on a vector of traits $\boldsymbol{\theta}$ is given by

$$\frac{d\lambda}{d\boldsymbol{\theta}^{\mathsf{T}}} \tag{6.27}$$

These gradients play a fundamental role in evolutionary biodemography, whether evolution is conceived of in terms of population genetics, quantitative genetics, adaptive dynamics, or mutation accumulation (e.g., Metz et al. 1992; Dercole and Rinaldi 2008; Rice 2004; Barfield et al. 2011). If the gradient is positive, selection favors an increase in the trait, and vice-versa.

In this application, $\boldsymbol{\xi}$ in (6.18) is the dominant eigenvalue λ. Let $\mathbf{w}$ and $\mathbf{v}$ be the right and left eigenvectors corresponding to λ, scaled so that $\mathbf{v}^\mathsf{T}\mathbf{w} = 1$. Then, in (6.26),

$$\frac{d\lambda}{d\operatorname{vec}^\mathsf{T}\tilde{\mathbf{A}}} = \mathbf{w}^\mathsf{T} \otimes \mathbf{v}^\mathsf{T}. \tag{6.28}$$

See Chap. 3 and Caswell (2010).

In this model, the vital rates are functions only of stage; the phenotype is blind to the age of the individual. However, the terms in the summations in (6.26) give the selection gradients on traits that would modify the phenotype at each age. That is,

$$\frac{d\lambda}{d\boldsymbol{\theta}^\mathsf{T}}\bigg|_{\text{age}=i} = \frac{d\lambda}{d\operatorname{vec}^\mathsf{T}\tilde{\mathbf{A}}} \left[(\mathbf{I}_{sw} \otimes \mathbf{Q}_\mathrm{U})\, (\mathbf{E}_{ii} \otimes \mathbf{K} \otimes \mathbf{I}_s)\, \big(\operatorname{vec}\mathbf{I}_\omega \otimes \mathbf{I}_{s^2}\big)\, \frac{d\operatorname{vec}\mathbf{U}_i}{d\boldsymbol{\theta}^\mathsf{T}} \right]$$
$$+ \frac{d\lambda}{d\operatorname{vec}^\mathsf{T}\tilde{\mathbf{A}}} \left[(\mathbf{I}_{sw} \otimes \mathbf{Q}_\mathrm{F})\, (\mathbf{E}_{ii} \otimes \mathbf{K} \otimes \mathbf{I}_s)\, \big(\operatorname{vec}\mathbf{I}_\omega \otimes \mathbf{I}_{s^2}\big)\, \frac{d\operatorname{vec}\mathbf{F}_i}{d\boldsymbol{\theta}^\mathsf{T}} \right] \tag{6.29}$$

Thus, these terms reveal the selection patterns that would operate on a mutation that was able to detect the age of an individual within a given stage, or that affected age differentially depending on the stage of the individual.

To examine the selection gradients on survival, it is necessary to separate survival from inter-stage transitions in $\mathbf{U}$. Let $\boldsymbol{\sigma}$ be the vector of stage-specific survival probabilities. The matrix $\mathbf{U}$ can be written as the product of a matrix $\boldsymbol{\Sigma} = \mathbf{1}_s\boldsymbol{\sigma}^\mathsf{T}$ containing the survival probabilities on the diagonal and a matrix $\mathbf{G}$ of transition probabilities, conditional on survival;

$$\mathbf{U} = \mathbf{G}\boldsymbol{\Sigma}. \tag{6.30}$$

(cf. Chap. 8). If $\mathbf{F}$ is independent[1] of $\boldsymbol{\sigma}$, then

$$d\mathbf{U} = \mathbf{G}\, d\boldsymbol{\Sigma}. \tag{6.31}$$

Applying the vec operator gives

$$d\operatorname{vec}\mathbf{U} = (\mathbf{I}_s \otimes \mathbf{G})\operatorname{vec}\mathcal{D}\,(\mathbf{1}_s d\boldsymbol{\sigma}^\mathsf{T})$$
$$= (\mathbf{I}_s \otimes \mathbf{G})\,\mathcal{D}\,(\operatorname{vec}\mathbf{I}_s)\,(\mathbf{I}_s \otimes \mathbf{1}_s)\, d\boldsymbol{\sigma} \tag{6.32}$$

---

[1]By assuming that $\mathbf{F}$ does not depend on $\boldsymbol{\sigma}$, I am in effect choosing a pre-breeding census and excluding neonatal mortality from $\boldsymbol{\sigma}$.

which implies that

$$\frac{d\text{vec}\,\mathbf{U}}{d\boldsymbol{\sigma}^{\mathsf{T}}} = (\mathbf{I}_s \otimes \mathbf{G})\,\mathcal{D}\,(\text{vec}\,\mathbf{I}_s)\,(\mathbf{I}_s \otimes \mathbf{1}_s) \tag{6.33}$$

Setting $\boldsymbol{\theta} = \boldsymbol{\sigma}$ and substituting (6.33) and (6.28) into (6.18) gives the selection gradient on $\boldsymbol{\sigma}$. Substituting (6.33) and (6.28) into (6.29), with $d\text{vec}\,\mathbf{F}/d\boldsymbol{\theta}^{\mathsf{T}} = \mathbf{0}$, gives the selection gradient on $\boldsymbol{\sigma}$ as a function of age and stage.

Results   The projection matrix $\mathbf{A}$ for Scotch broom[2] is

$$\mathbf{A} = \begin{pmatrix} 0.740 & 0 & 3.400 & 47.1 & 108.700 & 1120.0 & 3339.0 \\ 0.001 & 0.310 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.350 & 0.310 & 0 & 0 & 0 & 0 \\ 0 & 0.038 & 0.290 & 0.024 & 0 & 0 & 0 \\ 0 & 0 & 0.069 & 0.390 & 0.320 & 0 & 0.091 \\ 0 & 0 & 0 & 0.440 & 0.440 & 0.530 & 0.091 \\ 0 & 0 & 0 & 0 & 0.029 & 0.400 & 0.730 \end{pmatrix} \tag{6.34}$$

The matrix $\mathbf{U}$ is obtained from $\mathbf{A}$ by setting all elements in the first row, except for a_{11}, to zero. The matrix $\mathbf{F}$ is a 7×7 matrix with the elements of row 1, columns 2–7 of $\mathbf{A}$ in the corresponding positions, and zeros elsewhere. The maximum age was set to $\omega = 30$. The aging matrices $\mathbf{D}_{\mathrm{U}}$ and $\mathbf{D}_{\mathrm{F}}$ are given by (6.2) and (6.3) with $\omega = 30$. Because the vital rates do not depend on age, the dominant eigenvalues of $\mathbf{A}$ and $\tilde{\mathbf{A}}$ should be identical, and they are; $\lambda = 1.268$.

The selection gradients on stage-specific survival (i.e., sensitivities of λ to $\boldsymbol{\sigma}$) are shown in Fig. 6.1. There is a steady decline with increasing stage, from seeds to medium-sized adults, but then an increase for large and extra-large adults. A quite different pattern emerges when the selection gradients are calculated as functions of both age and stage, using (6.29). These results are shown in Fig. 6.2. The age-specific selection gradients on survival in stages 1–3 are strictly decreasing with age. But the age-specific selection gradients on survival in the adult stages 4–7 *increase* with age, level off, and then decline. The increase is longer and more pronounced in the larger adult stages.

It is now known that this pattern is widespread in plant populations. It appears in all eight of the Scotch broom populations studied by Parker (2000), and in almost all of 36 species of plants examined by Caswell and Salguero-Gómez (2013). It has important implications for the evolution of senescence. Hamilton (1966) showed that the selection gradient on age-specific mortality is always decreases with age, and argued that this implied that selection would always lead to senescence. Incorporating stage-dependence as well as age-dependence of the vital

---

[2]This is the matrix for the Discovery Park population, 1993–1994, edge conditions; taken from the Appendix of Parker (2000).

Fig. 6.1 Sensitivity of population growth rate λ to stage-specific survival probabilities. Calculated for the stage-classified model of Scotch broom (*Cytisus scoparius*) using data from Parker (2000). Stages: 1 = seeds, 2 = seedlings, 3 = juveniles, 4 = small adults, 5 = medium adults, 6 = large adults, 7 = extra-large adults
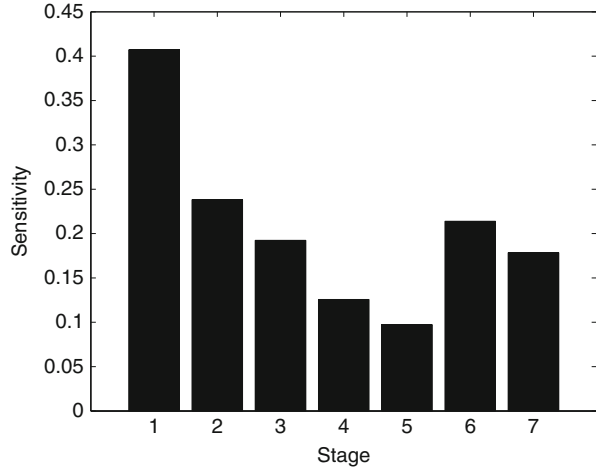
Fig. 6.2 Sensitivity of population growth rate λ to stage-specific survival as a function of age, for Scotch broom. Stages defined as in Fig. 6.1
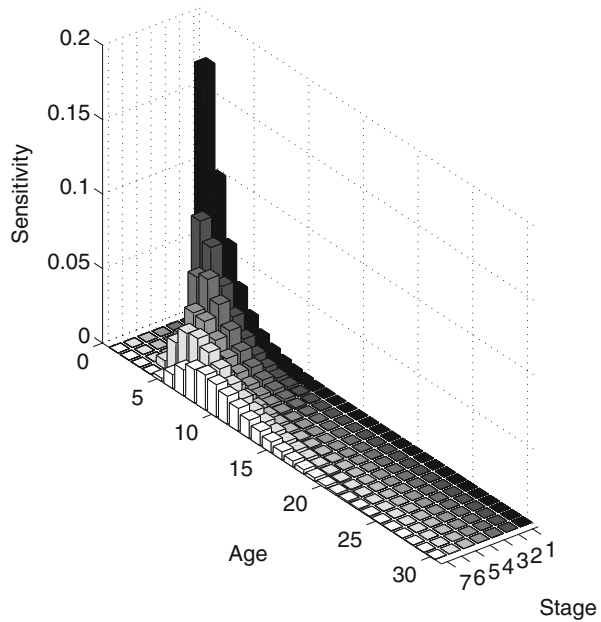
rates means that, over some range of ages, the selection gradient increases (*contra-senescent* selection in the terminology of Caswell and Salguero-Gómez 2013). Thus conclusions that follow from the general decline in selection gradients with age may not apply to traits that affect age-specific survival differentially depending on developmental stage. Traits that affect survival in adult stages should postpone senescence for at least some time.

6.4.2   Distributions of Age and Stage at Death

The pattern of longevity within a population is captured by the probability distribution of the age at death, one of the standard results of age-classified life table analysis. The moments of the age at death and their sensitivity can also be calculated directly from stage-classified models using Markov chain methods (Feichtinger 1971b; Caswell 2001, 2006, 2009; Tuljapurkar and Horvitz 2006; Horvitz and Tuljapurkar 2008); see Chaps. 4 and 5. Here we can go beyond that and get the full joint distribution of stage and age at death, along with the marginal distributions of age at death and stage at death, implied by an age × stage classified model.

   To do this, note that the cohort projection matrix $\tilde{\mathbf{U}}$ describes movement of individuals among transient states of an absorbing Markov chain, where the absorbing state is death, or death classified by stage or age at death. The transition matrix of the chain is

$$\tilde{\mathbf{P}} = \left(\begin{array}{c|c} \tilde{\mathbf{U}} & 0 \\ \hline \tilde{\mathbf{M}} & \mathbf{I} \end{array} \right) \tag{6.35}$$

By properly structuring $\mathbf{M}$, the model can give information about the age, stage, or the joint distribution of age and stage at death.[3] Each row of $\tilde{\mathbf{M}}$ corresponds to an absorbing state, and $\tilde{m}_{ij}$ is the probability of a transition from transient state j to absorbing state i. To compute the distribution of age and stage at death, we define the absorbing states to correspond to the age × stage combination at death. Thus $\tilde{\mathbf{M}}$ contains probabilities of death on the diagonal and zeros elsewhere,

$$\tilde{\mathbf{M}} = \mathbf{I}_{s\omega} - \mathcal{D} \left(\mathbf{1}_{s\omega}^{\mathsf{T}} \tilde{\mathbf{U}} \right). \tag{6.36}$$

   The fundamental matrix of the Markov chain in (6.35) is

$$\tilde{\mathbf{N}} = \left(\mathbf{I} - \tilde{\mathbf{U}} \right)^{-1} \tag{6.37}$$

The (i, j) element of $\tilde{\mathbf{N}}$ is the expected number of visits that an individual in state j will make to transient state i before death.

   Consider the eventual fate of an individual starting in transient state j. Let

$$\tilde{b}_{ij} = P \left[\text{eventual absorption in } i \mid \text{starting in } j \right] \tag{6.38}$$

---

[3]This also leads to a powerful approach, including sensitivity analysis, for cause of death calculations (Caswell and Ouellette 2016, 2018).

The $\tilde{b}_{ij}$ are the elements of the matrix $\tilde{\mathbf{B}}$ ($s\omega \times s\omega$) given by

$$\tilde{\mathbf{B}} = \tilde{\mathbf{M}}\tilde{\mathbf{N}} \tag{6.39}$$

(Iosifescu 1980, Theorem 3.3; see also Caswell 2001, Section 5.1). Since the absorbing states (the rows of $\tilde{\mathbf{M}}$) correspond to combinations of age and stage at death, column j of $\tilde{\mathbf{B}}$ gives the joint distribution of age and stage at death, starting from state (i.e., age × stage combination) j:

$$\tilde{\mathbf{B}}(:, j) = \tilde{\mathbf{B}}\mathbf{e}_j \tag{6.40}$$

using MATLAB notation in which $\mathbf{X}(:, j)$ is column j of $\mathbf{X}$, and where $\mathbf{e}_j$ is a vector of length $s\omega$ with a 1 in the jth entry and zeros elsewhere. The rows of $\tilde{\mathbf{B}}$ correspond to combinations of stage and age at death. Summing the rows over stages gives the marginal distribution of *age* at death, starting in column j of $\tilde{\mathbf{B}}$, as

$$\mathbf{g}_j = \left(\mathbf{I}_\omega \otimes \mathbf{1}_s^\mathsf{T}\right)\tilde{\mathbf{B}}(:, j) \qquad \text{marginal age distribution} \qquad \omega \times 1 \tag{6.41}$$

Similarly, summing over ages gives the marginal distribution of *stage* at death:

$$\mathbf{h}_j = \left(\mathbf{1}_\omega^\mathsf{T} \otimes \mathbf{I}_s\right)\tilde{\mathbf{B}}(:, j) \qquad \text{marginal stage distribution} \qquad s \times 1 \tag{6.42}$$

6.4.2.1 Perturbation Analysis

In the general sensitivity equation (6.18), the dependent variable $\boldsymbol{\xi} = \tilde{\mathbf{B}}(:, j)$. This depends only on $\tilde{\mathbf{U}}$, so the first term in (6.18) can be shown to be

$$\frac{d\boldsymbol{\xi}}{d\text{vec}\,\tilde{\mathbf{A}}} = \frac{d\tilde{\mathbf{B}}(:, j)}{d\text{vec}\,\tilde{\mathbf{U}}} \tag{6.43}$$

$$= -\left(\mathbf{e}_j^\mathsf{T}\tilde{\mathbf{N}}^\mathsf{T} \otimes \mathbf{I}_{s\omega}\right)\mathcal{D}\left(\text{vec}\,\mathbf{I}_{s\omega}\right)\left(\mathbf{I}_{s\omega} \otimes \mathbf{1}_{s\omega}\mathbf{1}_{s\omega}^\mathsf{T}\right) + \left(\mathbf{e}_j^\mathsf{T}\tilde{\mathbf{N}}^\mathsf{T} \otimes \tilde{\mathbf{B}}\right) \tag{6.44}$$

The desired derivative $d\tilde{\mathbf{B}}(:, j)/d\boldsymbol{\theta}^\mathsf{T}$ is obtained by substituting (6.44) for $d\boldsymbol{\xi}/d\text{vec}\,\tilde{\mathbf{A}}$ in (6.26), setting $d\text{vec}\,\mathbf{F}_i/d\boldsymbol{\theta}^\mathsf{T} = 0$.

The sensitivities of the marginal distributions of age and stage at death are then given by

$$\frac{d\mathbf{g}_j}{d\boldsymbol{\theta}^\mathsf{T}} = \left(\mathbf{I}_\omega \otimes \mathbf{1}_s^\mathsf{T}\right)\frac{d\tilde{\mathbf{B}}(:, j)}{d\boldsymbol{\theta}^\mathsf{T}} \tag{6.45}$$

$$\frac{d\mathbf{h}_j}{d\boldsymbol{\theta}^\mathsf{T}} = \left(\mathbf{1}_\omega^\mathsf{T} \otimes \mathbf{I}_s\right)\frac{d\tilde{\mathbf{B}}(:, j)}{d\boldsymbol{\theta}^\mathsf{T}} \tag{6.46}$$

Derivation To derive the sensitivity of the joint distribution of age and stage at death, conditional on some starting age × stage combination, we start by differentiating equation (6.40) for column j of $\tilde{\mathbf{B}}$ and applying the vec operator,

$$d\tilde{\mathbf{B}}(:, j) = \left(\mathbf{e}_j^\mathsf{T} \otimes \mathbf{I}_{s\omega}\right) d\text{vec}\,\tilde{\mathbf{B}}. \tag{6.47}$$

However, from (6.39), $\tilde{\mathbf{B}} = \tilde{\mathbf{M}}\tilde{\mathbf{N}}$, so

$$d\tilde{\mathbf{B}} = \left(d\tilde{\mathbf{M}}\right)\tilde{\mathbf{N}} + \tilde{\mathbf{M}}\left(d\tilde{\mathbf{N}}\right). \tag{6.48}$$

and

$$d\text{vec}\,\tilde{\mathbf{B}} = \left(\tilde{\mathbf{N}}^\mathsf{T} \otimes \mathbf{I}_{s\omega}\right) d\text{vec}\,\tilde{\mathbf{M}} + \left(\mathbf{I}_{s\omega} \otimes \tilde{\mathbf{M}}\right) d\text{vec}\,\tilde{\mathbf{N}}. \tag{6.49}$$

The differential of the fundamental matrix $\tilde{\mathbf{N}}$ is

$$d\text{vec}\,\tilde{\mathbf{N}} = \left(\tilde{\mathbf{N}}^\mathsf{T} \otimes \tilde{\mathbf{N}}\right) d\text{vec}\,\tilde{\mathbf{U}} \tag{6.50}$$

(Caswell 2006; see Chap. 5). The differential of $\tilde{\mathbf{M}}$ is obtained by rewriting (6.36) as

$$\tilde{\mathbf{M}} = \mathbf{I}_{s\omega} - \mathbf{I}_{s\omega} \circ \left(\mathbf{1}_{s\omega}\mathbf{1}_{s\omega}^\mathsf{T}\tilde{\mathbf{U}}\right), \tag{6.51}$$

differentiating,

$$d\tilde{\mathbf{M}} = -\mathbf{I}_{s\omega} \circ \left[\mathbf{1}_{s\omega}\mathbf{1}_{s\omega}^\mathsf{T}\left(d\tilde{\mathbf{U}}\right)\right], \tag{6.52}$$

and applying the vec operator to obtain

$$d\text{vec}\,\tilde{\mathbf{M}} = -\mathcal{D}\,(\text{vec}\,\mathbf{I}_{s\omega})\left(\mathbf{I}_{s\omega} \otimes \mathbf{1}_{s\omega}\mathbf{1}_{s\omega}^\mathsf{T}\right)d\text{vec}\,\tilde{\mathbf{U}} \tag{6.53}$$

Substituting (6.50) and (6.53) into (6.49) gives

$$d\text{vec}\,\tilde{\mathbf{B}} = \Big[-\left(\tilde{\mathbf{N}}^\mathsf{T} \otimes \mathbf{I}_{s\omega}\right)\mathcal{D}\,(\text{vec}\,\mathbf{I}_{s\omega})\left(\mathbf{I}_{s\omega} \otimes \mathbf{1}_{s\omega}\mathbf{1}_{s\omega}^\mathsf{T}\right)$$
$$+ \left(\mathbf{I}_{s\omega} \otimes \tilde{\mathbf{M}}\right)\left(\tilde{\mathbf{N}}^\mathsf{T} \otimes \tilde{\mathbf{N}}\right)\Big]d\text{vec}\,\tilde{\mathbf{U}} \tag{6.54}$$

Substituting this into (6.47) gives

$$d\tilde{\mathbf{B}}(:, j) = \Big[-\left(\mathbf{e}_j^\mathsf{T} \otimes \mathbf{I}_{s\omega}\right)\left(\tilde{\mathbf{N}}^\mathsf{T} \otimes \mathbf{I}_{s\omega}\right)\mathcal{D}\,(\text{vec}\,\mathbf{I}_{s\omega})\left(\mathbf{I}_{s\omega} \otimes \mathbf{1}_{s\omega}\mathbf{1}_{s\omega}^\mathsf{T}\right)$$
$$+ \left(\mathbf{e}_j^\mathsf{T} \otimes \mathbf{I}_{s\omega}\right)\left(\mathbf{I}_{s\omega} \otimes \tilde{\mathbf{M}}\right)\left(\tilde{\mathbf{N}}^\mathsf{T} \otimes \tilde{\mathbf{N}}\right)\Big]d\text{vec}\,\tilde{\mathbf{U}} \tag{6.55}$$

Equation (6.55) can be simplified to obtain (6.44), using the fact that

$$(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = (\mathbf{AC} \otimes \mathbf{BD}),$$

provided the products exist.

Results Figure 6.3 shows the joint distribution of age and stage at death for a seed of age 1 (one definition of "newborn" in this life cycle), with $\omega = 40$. Almost all seeds will die as seeds, because the germination probability is low, $a_{21} = 0.001$; see (6.34). The fates of seedlings (another possible choice for newborn status) are more diverse, and those of juveniles and small adults even moreso; the distributions show what proportion will die as seedlings, juveniles, etc., and at what ages (Fig. 6.3).

The marginal distribution of age at death, for individuals in each initial stage, is given in Fig. 6.4. Not surprisingly, larger stages have an age distribution of death shifted to later ages, including some probability of survival to age class ω (≥ 40 years in this calculation).

The sensitivity of $\mathbf{g}_2$ (the marginal distribution of age at death for a seedling) is shown in Fig. 6.5. Changes in the survival of seeds (σ_1) have no effect on this
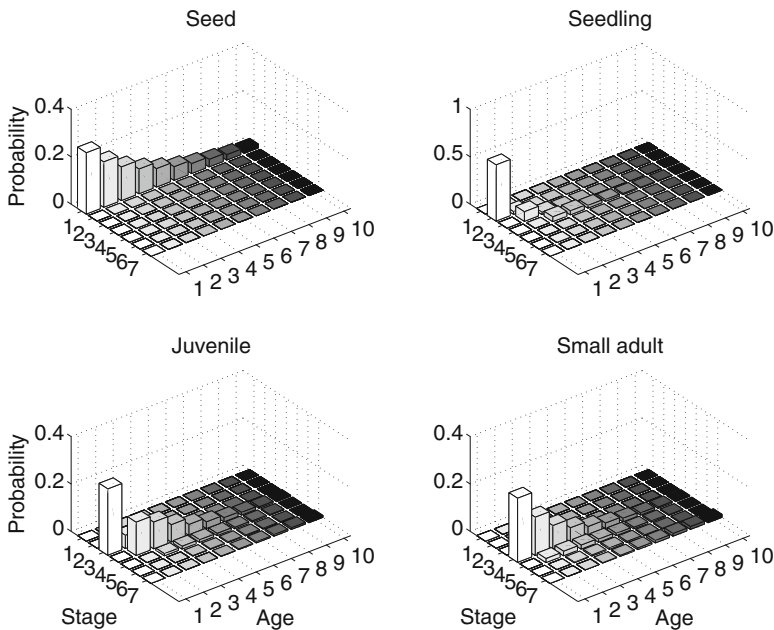
Fig. 6.3 The joint probability distribution of age $(1, \ldots, 10)$ and stage $(1, \ldots, 7)$ at death for an individual seed, seedling, juvenile, or small adult of Scotch broom. Stages as in Fig. 6.1

Fig. 6.4 The marginal
distributions of age at death
for individuals of Scotch
broom in each stage.
Maximum age is $\omega = 40$.
Stages as in Fig. 6.1
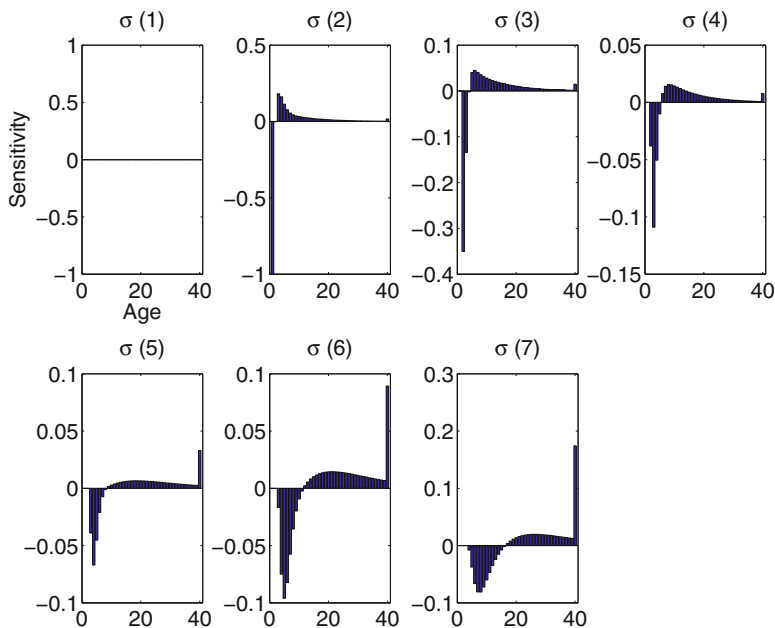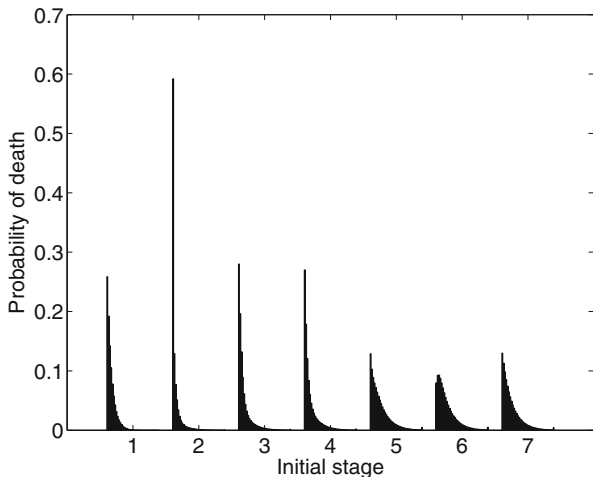
Fig. 6.5 Sensitivity of the marginal distribution of age at death, $\mathbf{g}_2$, to the survival probabilities of
each stage, for an individual starting in stage 2 (seedlings). Stages as in Fig. 6.1

distribution, because seedlings have already left the seed stage. Changes in σ_2–σ_7
shift the distribution to progressively older ages, by reducing the probability of death
at young ages and increasing it at older ages.

6.5 Discussion

Models in which individuals are classified by both age and stage extend demographic analyses in several directions. They permit biodemographic analyses of aging to take advantage of the many stage-classified demographic analyses accumulated by ecologists (Salguero-Gómez et al. 2015, 2016). They also permit human demographers to take account of factors other than age in determining mortality, longevity, fertility, and population dynamics.

Age- and stage-specific demographic processes are often combined in demography using multistate life table methods (e.g., Rogers 1975; Willekens 2002, 2014). These are usually focused on cohort dynamics and associated survival statistics (but see Rogers 1975, Chap. 5 for an explicit consideration of population projection). Multistate life table models are written as continuous-parameter, discrete-state Markov chains, where the parameter represents age and the states represent stages. In order to solve the resulting equations, the dynamics must be approximated over a (usually short) finite age interval; this would correspond to the sequence of matrices $\mathbf{A}_i$ in the model here. The age × stage-classified model described by $\tilde{\mathbf{A}}$ is a way to solve the discretized equations in a single step, and makes possible a variety of analyses that are difficult or impossible in the usual life table formulation. Further investigation of the relation between continuous multistate life table methods and age × stage-classified models will be interesting.

These analyses blur the distinction (Chap. 5) between implicit and explicit age dependence. If the $\mathbf{A}_i$ are truly identical, by definition only implicit age dependence is revealed. But the structure of the age × stage model separates all of the age-dependent $\mathbf{A}_i$, and thus is ready to include any degree of explicit joint dependence of the vital rates on age and stage.

Given sufficient longitudinal data on both age and stage, it is possible to estimate the stage-specific matrices $\mathbf{A}_i$ as explicit functions of age; see Peeters et al. (2002) for an example of a study of human heart disease, and Lebreton et al. (2009) for a review of methods used in multistate capture-mark-recapture analysis in ecology. Needless to say, the data requirements for a full age × stage parameterization are challenging. I suspect that the development of estimation methods at intermediate levels of detail will be an important step.

6.5.1 Reducibility and Ergodicity

The properties of $\tilde{\mathbf{A}}$ raise an important theoretical and technical issue regarding population growth, fitness, and selection gradients. The use of λ as a measure of fitness is usually justified by the strong ergodic theorem (Cohen 1979, Caswell 2001, Section 4.5.2), which guarantees the eventual convergence to the stable population structure and growth at a rate given by the dominant eigenvalue λ. A sufficient condition for this convergence is that the projection matrix be irreducible; i.e.,

that there exist a pathway connecting any two stages. Stott et al. (2010) surveyed published population projection matrices and found that reducible matrices were not uncommon, and explored the implications for ergodicity. Reducible matrices are not as bad as some people think, but it is important to understand their implications, especially for age × stage models.

General results about the irreducibility of block-structured matrices are difficult; see Csetenyi and Logofet (1989), Logofet (1993, Chap. 3), and Logofet and Belova (2007) for some important graph-theoretical results. However, the age × stage matrices developed here are unusual among population models in that they are (almost) always reducible, because they contain categories to which there are no possible pathways. This arises because age 1 individuals are produced only by reproduction. Hence there can never be age 1 individuals in any stage that is not produced by reproduction. For example, Scotch broom reproduces only by seeds, so age 1 seeds appear in the model. However, the matrix $\tilde{\mathbf{A}}$ also contains entries corresponding to age 1 seedlings, age 1 juveniles, age 1 adults, etc. These do not exist, and because there are no pathways to these stages from any other stages, the matrix $\tilde{\mathbf{A}}$ is reducible.

The Perron-Frobenius theorem guarantees that a reducible non-negative matrix will have a real, non-negative, dominant eigenvalue that is at least as large as any of the others. However, the asymptotic population growth rate and structure may depend on initial conditions (Caswell 2001, Section 4.5.4) This means that one must ascertain that the eigenvalues and eigenvectors under analysis correspond to initial conditions of interest.

Appendix A shows that a necessary and sufficient condition for population growth to be described by the dominant eigenvalue λ of $\tilde{\mathbf{A}}$, regardless of the (non-negative and non-zero) initial population vector, is that the left eigenvector $\mathbf{v}$ be strictly positive, and that this corresponds to a particular block-triangular form of $\tilde{\mathbf{A}}$. This provides a simple check for the ergodicity of population growth, and justifies the use of λ as a population growth rate and measure of fitness.

Primitivity may be difficult to evaluate for an age × stage matrix (but see Logofet 1993) but as with any projection matrix model, the long-term average growth rate of a primitive matrix is still given by the dominant real eigenvalue.

The matrix $\tilde{\mathbf{A}}$ for Scotch broom in (6.34) is reducible, as shown by calculating $\left(\mathbf{I}_{s\omega} + \tilde{\mathbf{A}}\right)^{s\omega}$ and finding that this matrix contains zeros (Caswell 2001). However, the left eigenvector $\mathbf{v}$ is strictly positive, so we know that the population eventually grows at the rate λ regardless of initial conditions.

6.5.2   A Protocol for Age × Stage-Classified Models

The approach outlined here gives a step-by-step procedure for constructing and analyzing age × stage-classified matrix population models.

1. Choose a question, and a corresponding demographic outcome. Are you interested in population dynamics (growth, structure, transients)? Or in cohort dynamics (survival, longevity)? Or in some combination of the two?
2. Obtain the stage-classified projection matrices $\mathbf{A}_i$ for ages $i = 1, \ldots, \omega$.
3. Decompose $\mathbf{A}_i = \mathbf{U}_i + \mathbf{F}_i$.
4. Construct the block-diagonal matrices $\mathbb{A}$, $\mathbb{F}$, $\mathbb{U}$, and $\mathbb{D}$, according to Eqs. (6.5)–(6.10).
5. Construct the age × stage matrices $\tilde{\mathbf{A}}$, $\tilde{\mathbf{F}}$, $\tilde{\mathbf{U}}$ using (6.16) and, if appropriate for the question at hand, also $\tilde{\mathbf{M}}$ and $\tilde{\mathbf{P}}$ using (6.35) and (6.36).
6. Analyze the model, e.g., by computing eigenvalues, eigenvectors, the fundamental matrix, etc., as appropriate. If necessary, check for reducibility and ergodicity using the methods in Sect. 6.5.1.
7. For sensitivity analysis,

   (a) choose a dependent variable $\boldsymbol{\xi}$ and a vector of parameters $\boldsymbol{\theta}$,
   (b) compute the sensitivity matrix $d\boldsymbol{\xi}/d\mathrm{vec}^{\mathsf{T}}\tilde{\mathbf{A}}$,
   (c) compute the matrices:

$$\frac{d\mathrm{vec}\,\mathbf{A}_i}{d\boldsymbol{\theta}^{\mathsf{T}}}, \quad \frac{d\mathrm{vec}\,\mathbf{U}_i}{d\boldsymbol{\theta}^{\mathsf{T}}}, \quad \text{and} \quad \frac{d\mathrm{vec}\,\mathbf{F}_i}{d\boldsymbol{\theta}^{\mathsf{T}}}$$

   (d) compute $d\boldsymbol{\xi}/d\boldsymbol{\theta}^{\mathsf{T}}$ according to (6.18).

The explicit connection between matrix population models and absorbing Markov chain theory makes it possible to analyze both population dynamics and cohort dynamics in a unified framework (cf. Feichtinger 1971a; Caswell 2001, 2006, 2009). Cohort dynamics are, in essence, the demography of individuals. It may seem paradoxical to speak of the demography of individuals, but that is what it is, because the *statistical* properties of a cohort (e.g., average lifespan) are *probabilistic* properties of an individual (e.g., life expectancy). Demography in general, and matrix population models in particular, provides the link between the individual and the population.

A   Appendix: Population Growth and Reducible Matrices

Some ergodic properties of population growth under the action of reducible matrices are described by Caswell (2001, Section4.5.4). Here we can extend the analysis.

Let $\mathbf{A}$ be a reducible non-negative projection matrix. By permutation of its rows and columns (i.e., renumbering the stages in the life cycle), $\mathbf{A}$ can be transformed to a block lower-triangular form. Here is an example:

$$\mathbf{A} = \begin{pmatrix} \mathbf{B}_{11} & 0 & 0 & 0 \\ \mathbf{B}_{21} & \mathbf{B}_{22} & 0 & 0 \\ \mathbf{B}_{31} & \mathbf{B}_{32} & \mathbf{B}_{33} & 0 \\ \mathbf{B}_{41} & \mathbf{B}_{42} & \mathbf{B}_{43} & \mathbf{B}_{44} \end{pmatrix}. \tag{6.56}$$

In this form, all the diagonal blocks $\mathbf{B}_{ii}$ are either irreducible matrices or 1×1 (i.e. scalar) zero matrices. The block triangular form is unique, up to a renumbering of the blocks and permutation of indices within blocks (Gantmacher 1959). It corresponds to a decomposition of the state space into a set of subspaces; let R_i be the subspace corresponding to the block $\mathbf{B}_{ii}$.

Some or all of the subdiagonal blocks in (6.56) may be zero. For reasons that will become apparent, consider an example where $\mathbf{B}_{21} = \mathbf{B}_{43} = \mathbf{0}$; i.e.,

$$\mathbf{A} = \begin{pmatrix} \mathbf{B}_{11} & 0 & 0 & 0 \\ 0 & \mathbf{B}_{22} & 0 & 0 \\ \mathbf{B}_{31} & \mathbf{B}_{32} & \mathbf{B}_{33} & 0 \\ \mathbf{B}_{41} & \mathbf{B}_{42} & 0 & \mathbf{B}_{44} \end{pmatrix} \tag{6.57}$$

Gantmacher (1959, Section 13.4) calls a block $\mathbf{B}_{ii}$ *isolated* if there are no other non-zero blocks on its row, that is, if $\mathbf{B}_{ij} = 0$ for $j < i$. I will call such a block *row-isolated*, and introduce the term *column-isolated* to describe any block $\mathbf{B}_{ii}$ with no other non-zero blocks in its column, that is, $\mathbf{B}_{ji} = 0$ for $j > i$. In the matrix in (6.57), the blocks $\mathbf{B}_{11}$ and $\mathbf{B}_{22}$ are row-isolated and the blocks $\mathbf{B}_{33}$ and $\mathbf{B}_{44}$ are column-isolated.

If $\mathbf{B}_{ii}$ is row-isolated, then the life cycle graph contains no pathways from any state outside of the subspace R_i to any state inside R_i, and R_i is a source. If $\mathbf{B}_{ii}$ is column-isolated, then the life cycle graph contains no pathways from any state in R_i to any state outside R_i, and R_i is a sink.

The eigenvalues of $\mathbf{A}$ are the eigenvalues of the diagonal blocks $\mathbf{B}_{ii}$. Let λ_1 be the dominant eigenvalue of $\mathbf{A}$, with right and left eigenvectors $\mathbf{w}_1$ and $\mathbf{v}_1$. The Perron-Frobenius theorem guarantees that λ_1, $\mathbf{w}_1$, and $\mathbf{v}_1$ are real and non-negative. Gantmacher (1959, Chap. 13, Theorem 6) proves that the eigenvector $\mathbf{w}_1$ is strictly positive if and only if λ_1 is an eigenvalue of every row-isolated block, and is not an eigenvalue of any of the non-row-isolated blocks. This makes it easy to demonstrate the following corollary.

Corollary: Positivity of $\mathbf{v}_1$ Let $\mathbf{v}_1$ be the left eigenvector corresponding to $\lambda_1[\mathbf{A}]$. Then $\mathbf{v}_1$ is strictly positive if and only if $\lambda_1[\mathbf{A}]$ is an eigenvalue of every column-isolated block, and is not an eigenvalue of any non-column-isolated block.

To see this, note that $\mathbf{v}_1$ is the right eigenvector of $\mathbf{A}^\mathsf{T}$. The column-isolated blocks of $\mathbf{A}$ become row-isolated blocks of the block lower-triangular form of $\mathbf{A}^\mathsf{T}$, and application of Gantmacher's Theorem 6 proves the Corollary.

For example, transposing (6.57) gives

$$\mathbf{A}^\mathsf{T} = \begin{pmatrix} \mathbf{B}_{11}^\mathsf{T} & 0 & \mathbf{B}_{31}^\mathsf{T} & \mathbf{B}_{41}^\mathsf{T} \\ 0 & \mathbf{B}_{22}^\mathsf{T} & \mathbf{B}_{32}^\mathsf{T} & \mathbf{B}_{42}^\mathsf{T} \\ 0 & 0 & \mathbf{B}_{33}^\mathsf{T} & 0 \\ 0 & 0 & 0 & \mathbf{B}_{44}^\mathsf{T} \end{pmatrix} \tag{6.58}$$

Reversing the order of the rows and columns gives the block lower-triangular form

$$
\begin{pmatrix}
\mathbf{B}_{44}^{\mathsf{T}} & 0 & 0 & 0 \\
0 & \mathbf{B}_{33}^{\mathsf{T}} & 0 & 0 \\
\mathbf{B}_{42}^{\mathsf{T}} & \mathbf{B}_{32}^{\mathsf{T}} & \mathbf{B}_{22}^{\mathsf{T}} & 0 \\
\mathbf{B}_{41}^{\mathsf{T}} & \mathbf{B}_{31}^{\mathsf{T}} & 0 & \mathbf{B}_{11}^{\mathsf{T}}
\end{pmatrix}
\tag{6.59}
$$

The column-isolated blocks in $\mathbf{A}$ ($\mathbf{B}_{33}$ and $\mathbf{B}_{44}$) now appear as row-isolated blocks in $\mathbf{A}^{\mathsf{T}}$. Gantmacher's result shows that the eigenvector $\mathbf{v}_1$ will be positive if and only if λ_1 is an eigenvalue of each of those blocks.

The usefulness of the Corollary follows from the population projection model

$$
\mathbf{n}(t+1) = \mathbf{A}\mathbf{n}(t) \qquad \mathbf{n}(0) = \mathbf{n}_0
\tag{6.60}
$$

and its solution[4]

$$
\mathbf{n}(t) = \sum_{i=1}^{s} c_i \lambda_i^t \mathbf{w}_i
\tag{6.61}
$$

$$
= \sum_{i=1}^{s} \left(\mathbf{v}_i^{\mathsf{T}} \mathbf{n}_0 \right) \lambda_i^t \mathbf{w}_i
\tag{6.62}
$$

Caswell (2001). If $\mathbf{n}_0$ is such that $c_1 = \mathbf{v}_1^{\mathsf{T}} \mathbf{n}_0$ is positive, then λ_1^t will eventually dominate all other terms in the solution and the population will grow at the rate λ_1 with stable structure $\mathbf{w}_1$. We know the following about c_1:

1. If $\mathbf{A}$ is irreducible, then by the Perron-Frobenius theorem $\mathbf{v}_1$ is strictly positive, so any non-negative, non-zero initial population $\mathbf{n}_0$ leads to a positive value of c_1 and eventual growth at the rate λ_1.
2. If $\mathbf{A}$ is reducible and $\mathbf{v}_1$ is strictly positive, any non-negative, non-zero $\mathbf{n}_0$ leads to a positive value of c_1 and growth at the rate λ_1.
3. If $\mathbf{A}$ is reducible and $\mathbf{v}_1$ contains zero entries corresponding to a subspace R_i, then initial conditions with positive support only in R_i will lead to $c_1 = 0$, and λ_1 will make no contribution to population growth from those initial vectors.

In the first two cases, population growth is ergodic from any non-zero initial population. In the third case, there exists a basin of attraction leading to growth according to λ_1, and a basin (or basins) of attraction for growth according to the dominant eigenvalues of the diagonal blocks $\mathbf{B}_{ii}$ corresponding to the zero entries of $\mathbf{v}_1$.

---

[4]This holds provided that $\mathbf{A}$ is diagonalizable, which is a generic property for linear operators (Hirsch and Smale 1974, p. 157).

Bibliography

Barfield, M., R. D. Holt, and R. Gomulkiewicz. 2011. Evolution in stage-structured populations. American Naturalist **177**:397–409.

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. 2009. Stage, age and individual stochasticity in demography. Oikos **118**:1763–1782.

Caswell, H. 2010. Reproductive value, the stable stage distribution, and the sensitivity of the population growth rate to changes in vital rates. Demographic Research **23**:531–548.

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. 2014. A matrix approach to the statistics of longevity in heterogeneous frailty models. Demographic Research **31**:553–592.

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 N. Ouellette, 2016. Mortality and causes of death: matrix formulation and sensitivity analysis. European Population Conference 2016, Mainz, Germany. URL http://epc2016.princeton.edu/papers/160437.

Caswell, H., and N. Ouellette, 2018. Cause-of-death analysis: matrix formulation and sensitivity analysis. In prep.

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

Caswell, H., and S. F. van Daalen. 2016. A note on the vec operator applied to unbalanced block-structured matrices. Journal of Applied Mathematics **2016**:1–3.

Cohen, J. E. 1979. Ergodic theorems in demography. Bulletin of the American Mathematical Society **1**:275–295.

Csetenyi, A. I., and D. O. Logofet. 1989. Leslie model revisited: some generalizations to block structures. Ecological Modelling **48**:277–290.

Dercole, F., and S. Rinaldi. 2008. Analysis of evolutionary processes: the adaptive dynamics approach and its applications. Princeton University Press, Princeton, NJ, USA.

Feichtinger, G. 1971a. Stochastische Modelle Demographischer Prozesse. Lecture notes in operations research and mathematical systems, Springer-Verlag, Berlin, Germany.

Feichtinger, G. 1971b. Stochastische Modelle demographischer Prozesse. Lecture Notes in Economics and Mathematical Systems, Springer-Verlag, Berlin.

Gantmacher, F. R. 1959. Matrix Theory. Chelsea, New York, New York.

Goodman, L. A. 1969. The analysis of population growth when the birth and death rates depend upon several factors. Biometrics **25**:659–681.

Hamilton, W. D. 1966. The moulding of senescence by natural selection. Journal of Theoretical Biology **12**:12–45.

Henderson, H. V., and S. R. Searle. 1981. The vec-permutation matrix, the vec operator and Kronecker products: a review. Linear and Multilinear Algebra **9**:271–288.

Hirsch, M. W., and S. Smale. 1974. Differential equations, dynamical systems, and linear algebra. Academic Press, New York, NY, USA.

Horvitz, C. C., and S. Tuljapurkar. 2008. Stage dynamics, period survival, and mortality plateaus. American Naturalist **172**:203–215.

Hunter, C. M., and H. Caswell. 2005. The use of the vec-permutation matrix in spatial matrix population models. Ecological Modelling **188**:15–21.

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

Klepac, P., and H. Caswell. 2011. The stage-structured epidemic: linking disease and demography with a multi-state matrix approach model. Theoretical Ecology **4**:301–319.

Lebreton, J. 1996. Demographic models for subdivided populations: the renewal equation approach. Theoretical Population Biology **49**:291–313.

Lebreton, J.-D., J. D. Nichols, R. J. Barker, R. Pradel, and J. A. Spendelow. 2009. Modeling individual animal histories with multistate capture-recapture models. Advances in Ecological Research **41**:87–173.

Logofet, D. O. 1993. Matrices and graphs: stability problems in mathematical ecology. CRC Press, Boca Raton, FL, USA.

Logofet, D. O., and I. N. Belova. 2007. Nonnegative matrices as a tool to model population dynamics: classical models and contemporary expansions (in Russian with English summary). Journal of Mathematical Sciences **155**:894–907.

Magnus, J. R., and H. Neudecker. 1985. Matrix differential calculus with applications to simple, Hadamard, and Kronecker products. Journal of Mathematical Psychology **29**:474–492.

Metz, J. A. J., and O. Diekmann. 1986. The dynamics of physiologically structured populations. Springer-Verlag, Berlin, Germany.

Metz, J. A. J., R. M. Nisbet, and S. A. H. Geritz. 1992. How should we define 'fitness' for general ecological scenarios? Trends in Ecology and Evolution **7**:198–202.

Neubert, M. G., and I. M. Parker. 2004. Projecting rates of spread for invasive species. Risk Analysis **24**:817–831.

Parker, I. M. 2000. Invasion dynamics of *Cytisus scoparius*: a matrix model approach. Ecological Applications **10**:726–743.

Pascarella, J. B., and C. C. Horvitz. 1998. Hurricane disturbance and the population dynamics of a tropical understory shrub: megamatrix elasticity analysis. Ecology **79**:547–563.

Peeters, A., A. A. Mamun, F. Willekens, and L. Bonneux. 2002. A cardiovascular life history: a life course analysis of the original Framingham Heart Study cohort. European Heart Journal **23**:458–466.

Rice, S. H. 2004. Evolutionary theory: mathematical and conceptual foundations. Sinauer, Sunderland, MA, USA.

Rogers, A. 1966. The multiregional matrix growth operator and the stable interregional age structure. Demography **3**:537–544.

Rogers, A. 1975. Introduction to multiregional mathematical demography. Wiley, New York, NY, USA.

Salguero-Gómez, R., O. R. Jones, C. R. Archer, C. Bein, H. Buhr, C. Farack, F. Gottschalk, A. Hartmann, A. Henning, G. Hoppe, G. Romer, T. Ruoff, V. Sommer, J. Wille, J. Voigt, S. Zeh, D. Vieregg, Y. M. Buckley, J. Che-Castaldo, D. Hodgson, A. Scheuerlein, H. Caswell, and J. W. Vaupel. 2016. COMADRE: a global data base of animal demography. Journal of Animal Ecology **85**:371–384.

Salguero-Gómez, R., O. R. Jones, C. R. Archer, Y. M. Buckley, J. Che-Castaldo, H. Caswell, D. Hodgson, A. Scheuerlein, D. A. Conde, E. Brinks, et al. 2015. The COMPADRE Plant Matrix Database: an open online repository for plant demography. Journal of Ecology **103**:202–218.

Stott, I., S. Townley, D. Carslake, and D. J. Hodgson. 2010. On reducibility and ergodicity of population projection matrix models. Methods in Ecology and Evolution **1**:242–252.

Tuljapurkar, S., and C. C. Horvitz. 2006. From stage to age in variable environments: life expectancy and survivorship. Ecology **87**:1497–1509.

Willekens, F. 2014. Multistate Analysis of Life Histories with R. Springer, New York, New York.

Willekens, F. J., 2002. Multistate demography. *in* Encyclopaedia of Population. Macmillan Reference, New York, NY, USA.

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.

Zhou, Y., H. Putter, and G. Doblhammer. 2016. Years of life lost due to lower extremity injury in association with dementia, and care need: A 6-year follow-up population-based study using a multi-state approach among German elderly. BMC Geriatrics **16**:9.

Part III
Time-Varying and Stochastic Models

Chapter 7
Transient Population Dynamics

7.1 Introduction

Short-term, transient population dynamics can differ in important ways from long-term asymptotic dynamics. Just as perturbation analysis (sensitivity and elasticity) of the asymptotic growth rate reveals the effects of the vital rates on long-term growth (Chap. 3), the perturbation analysis of transient dynamics can reveal the determinants of short-term patterns. This chapter presents a comprehensive approach to transient sensitivity analysis that applies to linear time-invariant, time-varying, subsidized, stochastic, nonlinear, and spatial models.

In a constant environment, once a population converges to its stable stage structure, it grows exponentially at a constant rate λ. However, depending on initial conditions, short-term transient dynamics can differ from the asymptotic dynamics. It has long been recognized that a focus on λ alone can obscure these important transient effects (e.g., Lotka 1939; Coale 1972). There have been attempts to develop transient sensitivity analyses using all the eigenvalues of the projection matrix (Fox and Gurevitch 2000), but these are complicated to calculate and limited in application. Matrix calculus allows us to do better (Caswell 2007).

7.2 Time-Invariant Models

Armed with matrix calculus, consider the linear time-invariant model,

$$\mathbf{n}(t+1) = \mathbf{A}\mathbf{n}(t) \qquad \mathbf{n}(0) = \mathbf{n}_0, \tag{7.1}$$

where $\mathbf{n}$ is $s \times 1$ and $\mathbf{A}$ is $s \times s$; s the number of stages. Assume that $\mathbf{A} = \mathbf{A}[\boldsymbol{\theta}]$ depends on a $p \times 1$ vector of parameters $\boldsymbol{\theta}$, which could be entries of $\mathbf{A}$, lower-level parameters, or elements of the initial vector.

The sequence of matrices

$$\frac{d\mathbf{n}(t)}{d\boldsymbol{\theta}^{\mathsf{T}}} \qquad t = 1, 2, \ldots \tag{7.2}$$

gives the effect of all the parameters on all the entries of $\mathbf{n}(t)$. From it we can calculate the sensitivities and elasticities of other dependent variables (Sect. 7.3).

We differentiate the model (7.1), obtaining

$$d\mathbf{n}(t+1) = \mathbf{A}\, d\mathbf{n}(t) + (d\mathbf{A})\, \mathbf{n}(t), \tag{7.3}$$

and then apply the vec operator to both sides, remembering that since $\mathbf{n}$ is a vector, $\mathrm{vec}\,\mathbf{n} = \mathbf{n}$,

$$d\mathbf{n}(t+1) = \mathbf{A}d\mathbf{n}(t) + \left(\mathbf{n}^{\mathsf{T}}(t) \otimes \mathbf{I}_s\right) d\mathrm{vec}\,\mathbf{A}. \tag{7.4}$$

Then the first identification theorem and the chain rule, from Eqs. (2.47) and (2.18), give the sensitivity of $\mathbf{n}(t+1)$ to the elements of $\mathbf{A}$,

$$\frac{d\mathbf{n}(t+1)}{d\mathrm{vec}^{\mathsf{T}}\mathbf{A}} = \mathbf{A}\frac{d\mathbf{n}(t)}{d\mathrm{vec}^{\mathsf{T}}\mathbf{A}} + \left(\mathbf{n}^{\mathsf{T}}(t) \otimes \mathbf{I}_s\right). \tag{7.5}$$

The chain rule extends (7.5) to give the sensitivity to lower-level parameters,

$$\frac{d\mathbf{n}(t+1)}{d\boldsymbol{\theta}^{\mathsf{T}}} = \frac{d\mathbf{n}(t+1)}{d\mathrm{vec}^{\mathsf{T}}\mathbf{A}}\frac{d\mathrm{vec}\,\mathbf{A}}{d\boldsymbol{\theta}^{\mathsf{T}}}$$

$$= \mathbf{A}\frac{d\mathbf{n}(t)}{d\boldsymbol{\theta}^{\mathsf{T}}} + \left(\mathbf{n}^{\mathsf{T}}(t) \otimes \mathbf{I}_s\right)\frac{d\mathrm{vec}\,\mathbf{A}}{d\boldsymbol{\theta}^{\mathsf{T}}}. \tag{7.6}$$

Equations (7.5) and (7.6) are matrix difference equations in the sensitivities of $\mathbf{n}(t)$ to the elements of $\mathrm{vec}\,\mathbf{A}$ or of $\boldsymbol{\theta}$. If we know $d\mathbf{n}(t)/d\boldsymbol{\theta}^{\mathsf{T}}$ and $\mathbf{n}(t)$, we can calculate $d\mathbf{n}(t+1)/d\boldsymbol{\theta}^{\mathsf{T}}$ and $\mathbf{n}(t+1)$ and continue this iteration to obtain the transient sensitivities at any time. If the parameters in $\boldsymbol{\theta}$ affect the vital rates but not the initial population, the appropriate initial condition for this iteration is

$$\frac{d\mathbf{n}(0)}{d\boldsymbol{\theta}^{\mathsf{T}}} = \mathbf{0}_{s \times p}. \tag{7.7}$$

If $\boldsymbol{\theta}$ affects only the initial population, then

$$\frac{d\mathbf{n}(0)}{d\boldsymbol{\theta}^{\mathsf{T}}} = \mathbf{I}_s \tag{7.8}$$

gives the sensitivity of transient dynamics to a change in initial conditions.

7.3 Sensitivity of What? Choosing Dependent Variables

The sensitivity of other dependent variables may be more interesting than that of $\mathbf{n}(t)$. In an early (and relatively crude) transient analysis, Caswell and Werner (1978) analyzed the transient dynamics of the plant teasel (*Dipsacus sylvestris*) in terms of rosette area at time t (which might affect resistance to invasion by later successional species) and cumulative seed production up to time t (which might affect colonization of new sites). For a weedy species like teasel, either of these dependent variables might be more relevant than the asymptotic growth rate.

Here are some other biologically interesting dependent variables. They are easy to calculate from $d\mathbf{n}(t)/d\boldsymbol{\theta}^{\mathsf{T}}$.

1. Population density, as measured by a weighted sum of stage densities. Let $\mathbf{c} \geq 0$ be a weight vector. Then population density is $N(t) = \mathbf{c}^{\mathsf{T}}\mathbf{n}(t)$. This includes total density ($\mathbf{c} = \mathbf{1}_s$, a vector of ones), the density of a subset of stages ($c_i = 1$ for stages to be counted; $c_i = 0$ otherwise), biomass (c_i is the biomass of stage i), basal area, metabolic rate, etc. The sensitivity of $N(t)$ is

$$\frac{dN(t)}{d\boldsymbol{\theta}^{\mathsf{T}}} = \mathbf{c}^{\mathsf{T}}\frac{d\mathbf{n}(t)}{d\boldsymbol{\theta}^{\mathsf{T}}}. \tag{7.9}$$

2. Ratios measuring the relative abundances of different stages:

$$R(t) = \frac{\mathbf{a}^{\mathsf{T}}\mathbf{n}(t)}{\mathbf{b}^{\mathsf{T}}\mathbf{n}(t)}. \tag{7.10}$$

where $\mathbf{a}$ and $\mathbf{b}$ are weight vectors. Examples include the dependency ratio (in human demography, the ratio of the individuals below 15 or above 65 to those between 15 and 65), the sex ratio in a two-sex model, and the ratio of juveniles to adults, which is important in wildlife management (Williams et al. 2002; Skalski et al. 2005). The sensitivity of $R(t)$ is

$$\frac{dR(t)}{d\boldsymbol{\theta}^{\mathsf{T}}} = \left(\frac{\mathbf{b}^{\mathsf{T}}\mathbf{n}(t)\mathbf{a}^{\mathsf{T}} - \mathbf{a}^{\mathsf{T}}\mathbf{n}(t)\mathbf{b}^{\mathsf{T}}}{(\mathbf{b}^{\mathsf{T}}\mathbf{n}(t))^2}\right)\frac{d\mathbf{n}(t)}{d\boldsymbol{\theta}^{\mathsf{T}}}. \tag{7.11}$$

3. Cumulative density up to a specified time,

$$C(t) = \sum_{i=0}^{t}\mathbf{c}^{\mathsf{T}}\mathbf{n}(i), \tag{7.12}$$

the sensitivity of which is

$$\frac{dC(t)}{d\boldsymbol{\theta}^{\mathsf{T}}} = \mathbf{c}^{\mathsf{T}}\sum_{i=0}^{t}\frac{d\mathbf{n}(i)}{d\boldsymbol{\theta}^{\mathsf{T}}}. \tag{7.13}$$

4. Average density over an interval,

$$\bar{N}(t_1, t_2) = \frac{1}{t_2 - t_1} \sum_{i=t_1}^{t_2} N(i), \tag{7.14}$$

the sensitivity of which is

$$\frac{d\bar{N}(t_1, t_2)}{d\boldsymbol{\theta}^{\mathsf{T}}} = \frac{1}{t_2 - t_1} \sum_{i=t_1}^{t_2} \mathbf{c}^{\mathsf{T}} \frac{d\mathbf{n}(i)}{d\boldsymbol{\theta}^{\mathsf{T}}}. \tag{7.15}$$

5. Maximum (or minimum) density over an interval,

$$M(t_1, t_2) = \max_{t_1 \le i \le t_2} N(i). \tag{7.16}$$

Let $\tilde{t}$ be the time such that $M(t_1, t_2) = N(\tilde{t})$. Then, except in the unlikely event of ties,

$$\frac{dM(t_1, t_2)}{d\boldsymbol{\theta}^{\mathsf{T}}} = \mathbf{c}^{\mathsf{T}} \frac{d\mathbf{n}(\tilde{t})}{d\boldsymbol{\theta}^{\mathsf{T}}} \tag{7.17}$$

with a similar expression for the minimum.
6. Variance in density over an interval $t_1 \le t \le t_2$,

$$V(t_1, t_2) = \frac{1}{t_2 - t_1} \sum_{i=t_1}^{t_2} N^2(i) - \left[\bar{N}(t_1, t_2)\right]^2. \tag{7.18}$$

The sensitivity of V is

$$\frac{dV(t_1, t_2)}{d\boldsymbol{\theta}^{\mathsf{T}}} = \frac{2}{t_2 - t_1} \left[\sum_{i=t_1}^{t_2} N(i) \frac{dN(i)}{d\boldsymbol{\theta}^{\mathsf{T}}} - \bar{N}(t_1, t_2) \sum_{i=t_1}^{t_2} \frac{dN(i)}{d\boldsymbol{\theta}^{\mathsf{T}}} \right] \tag{7.19}$$

$$= \frac{2}{t_2 - t_1} \left[\sum_{i=t_1}^{t_2} \left(N(i) - \bar{N}(t_1, t_2) \right) \frac{dN(i)}{d\boldsymbol{\theta}^{\mathsf{T}}} \right]. \tag{7.20}$$

7. The transient population growth rate at time t,

$$r(t) = \log \frac{N(t + 1)}{N(t)}. \tag{7.21}$$

The sensitivity of r is

$$\frac{dr(t)}{d\boldsymbol{\theta}^{\mathsf{T}}} = \frac{\mathbf{c}^{\mathsf{T}}}{N(t+1)} \frac{d\mathbf{n}(t+1)}{d\boldsymbol{\theta}^{\mathsf{T}}} - \frac{\mathbf{c}^{\mathsf{T}}}{N(t)} \frac{d\mathbf{n}(t)}{d\boldsymbol{\theta}^{\mathsf{T}}}. \tag{7.22}$$

8. Average growth rate over an interval $t_1 \le t \le t_2$,

$$\bar{r}(t_1, t_2) = \frac{1}{t_2 - t_1} \log \frac{N(t_2)}{N(t_1)}, \tag{7.23}$$

the sensitivity of which is

$$\frac{d\bar{r}(t_1, t_2)}{d\boldsymbol{\theta}^{\mathsf{T}}} = \frac{1}{t_2 - t_1} \left(\frac{\mathbf{c}^{\mathsf{T}}}{N(t_2)} \frac{d\mathbf{n}(t_2)}{d\boldsymbol{\theta}^{\mathsf{T}}} - \frac{\mathbf{c}^{\mathsf{T}}}{N(t_1)} \frac{d\mathbf{n}(t_1)}{d\boldsymbol{\theta}^{\mathsf{T}}} \right). \tag{7.24}$$

7.4 Elasticity Analysis

Transient elasticities are easily calculated from the sensitivities. The elasticity of $n_i(t)$ to θ_j is

$$\frac{\epsilon n_i}{\epsilon \theta_j} = \frac{\theta_j}{n_i(t)} \frac{dn_i(t)}{d\theta_j}. \tag{7.25}$$

Creating a matrix of these elasticities requires multiplying column j of $d\mathbf{n}/d\boldsymbol{\theta}^{\mathsf{T}}$ by θ_j and dividing row i by n_i. This is just

$$\mathcal{D}\left[\mathbf{n}(t)\right]^{-1} \frac{d\mathbf{n}(t)}{d\boldsymbol{\theta}^{\mathsf{T}}} \mathcal{D}[\boldsymbol{\theta}], \tag{7.26}$$

where $\mathcal{D}[\mathbf{x}]$ is a matrix with $\mathbf{x}$ on the diagonal and zeros elsewhere. The elasticity of any other (scalar- or vector-valued) dependent variable $f(\mathbf{n}(t))$ is given by

$$\mathcal{D}\left[f(\mathbf{n}(t))\right]^{-1} \frac{df(\mathbf{n}(t))}{d\boldsymbol{\theta}^{\mathsf{T}}} \mathcal{D}[\boldsymbol{\theta}]. \tag{7.27}$$

Example: A transient outbreak: elasticity to lower-level parameters Consider a hypothetical size-classified population with

$$\mathbf{A} = \begin{pmatrix} 0.3763 & 0 & 0.8431 & 8.4312 \\ 0.1939 & 0.5421 & 0 & 0 \\ 0 & 0.1177 & 0.5240 & 0 \\ 0 & 0 & 0.1291 & 0.5254 \end{pmatrix}. \tag{7.28}$$

Fig. 7.1 Dynamics of a
transient population outbreak.
The projection matrix (7.28)
has $\lambda = 0.92$, but an initial
condition of a single adult
leads to a rapid outbreak that
lasts for over 25 years
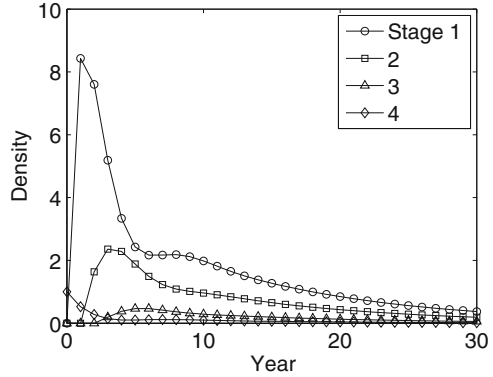
The asymptotic growth rate calculated as the dominant eigenvalue of **A** is
$\lambda = 0.92$, so the population is headed for eventual decline. However, the initial
condition

$$\mathbf{n}_0 = \begin{pmatrix} 0 & 0 & 0 & 1 \end{pmatrix}^{\mathsf{T}} \tag{7.29}$$

(introduction of a large adult) produces a dramatic transient outbreak (Fig. 7.1),
during which total population increases by over 900% and remains above its initial
value for about 25 years.[1]

   If this was a pest its asymptotic fate (extinction) would be reassuring, but
λ would reveal nothing about the transient outbreak. A manager might want to
know how changes in the lower-level survival probabilities σ_i, growth probabil-
ities γ_i, and fertilities f_i would affect the outbreak, where the elements of **A**
are

$$\begin{aligned}
a_{ii} &= \sigma_i(1 - \gamma_i) & i &= 1, \dots, 4 \\
a_{i+1,i} &= \sigma_i \gamma_i & i &= 1, \dots, 3 \\
a_{1i} &= f_i & i &= 3, 4.
\end{aligned} \tag{7.30}$$

If the impact of the pest was related to size, the manager might measure population
density with weights, say $\mathbf{c}^{\mathsf{T}} = \begin{pmatrix} 1 & 2 & 3 & 4 \end{pmatrix}$. Two measures of damage might be
the maximum of the outbreak and the cumulative population size over the entire
outbreak. Finally, to put everything on a proportional basis, the manager might want
to use elasticities.

---

[1]The curious reader may wish to know that **A** was obtained by a random search for size-classified
matrices with high reactivity (Neubert and Caswell 1997; Caswell and Neubert 2005; Verdy and
Caswell 2008).

Define $\boldsymbol{\theta}$ as the 9×1 vector whose entries are $\sigma_1-\sigma_4$, $\gamma_1-\gamma_3$, and f_3-f_4. The derivatives $d\text{vec}\,\mathbf{A}/d\boldsymbol{\theta}^{\mathsf{T}}$ are obtained from (7.30). The sensitivity of $\mathbf{n}(t)$ to changes in $\boldsymbol{\theta}$ is given by (7.6). Using (7.9) and (7.27) we obtain the elasticity of $N(t)$ to $\boldsymbol{\theta}$ as

$$\frac{\epsilon N(t)}{\epsilon \boldsymbol{\theta}^{\mathsf{T}}} = \frac{1}{N(t)} \, \mathbf{c}^{\mathsf{T}} \frac{d\mathbf{n}(t)}{d\boldsymbol{\theta}^{\mathsf{T}}} \, \mathcal{D}(\boldsymbol{\theta}). \tag{7.31}$$

The peak of the outbreak occurs at $t = 2$; thus (7.17) gives the elasticity of the peak density to $\boldsymbol{\theta}$ as

$$\frac{\epsilon N(2)}{\epsilon \boldsymbol{\theta}^{\mathsf{T}}} = \frac{1}{N(2)} \, c^{\mathsf{T}} \frac{d\mathbf{n}(2)}{d\boldsymbol{\theta}^{\mathsf{T}}} \, \mathcal{D}(\boldsymbol{\theta}). \tag{7.32}$$

The cumulative density up to time t is given by (7.12) and the sensitivity by (7.13), so the elasticity is

$$\frac{1}{\mathbf{c}^{\mathsf{T}} \sum_0^t \mathbf{n}(t)} \, \mathbf{c}^{\mathsf{T}} \sum_{i=0}^{t} \frac{d\mathbf{n}(i)}{d\boldsymbol{\theta}^{\mathsf{T}}} \, \mathcal{D}(\boldsymbol{\theta}). \tag{7.33}$$

Results are shown in Fig. 7.2. The elasticities of the maximum outbreak density are very different from those of λ. The elasticity of the cumulative density over the first 5 years has a similar pattern, also very different from that of λ. However, by the end of the outbreak (25 years) the elasticity of cumulative density is quite similar to that
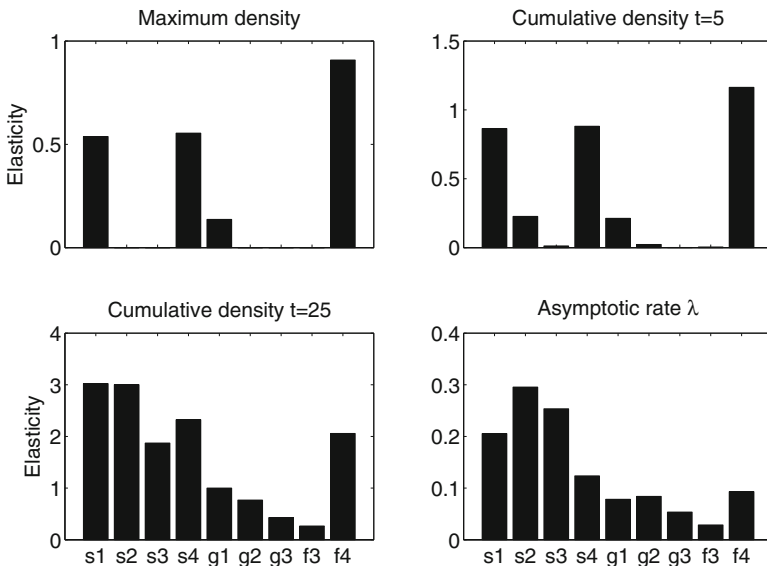
Fig. 7.2 The elasticities of the maximum population density, of the cumulative densities up to $t = 5$ and $t = 25$, and of λ to the lower-level demographic parameters, for the outbreak shown in Fig. 7.1

of λ, so management over this time scale could reasonably rely on the elasticity of λ to compare control tactics. Intermediate steps and MATLAB code are found in an appendix to Caswell (2007). ∎

7.5 Sensitivity of Time-Varying Models

Now consider the time-varying model

$$\mathbf{n}(t+1) = \mathbf{A}_t \mathbf{n}(t) \qquad \mathbf{n}(0) = \mathbf{n}_0, \tag{7.34}$$

where $\mathbf{A}_t$, $t = 1, \ldots, T$ is a specified sequence of matrices.

Take the differential of both sides of (7.34)

$$d\mathbf{n}(t+1) = \mathbf{A}_t d\mathbf{n}(t) + (d\mathbf{A}_t) \mathbf{n}(t), \tag{7.35}$$

and apply the vec operator to obtain

$$d\mathbf{n}(t+1) = \mathbf{A}_t d\mathbf{n}(t) + \left(\mathbf{n}^\mathsf{T}(t) \otimes \mathbf{I}_s\right) (d\mathrm{vec}\,\mathbf{A}_t). \tag{7.36}$$

Not only the transient behavior of the population, but also the parameter vector $\boldsymbol{\theta}$, the matrix $\mathbf{A}_t$, and the perturbation applied to $\boldsymbol{\theta}$ may change over time. The sensitivity analysis must reflect both types of variation. So, let us treat $\mathbf{A}_t$ as a function of $\boldsymbol{\theta}(t)$, and consider a perturbation of $\boldsymbol{\theta}$ at some time u. Applying the chain rule to (7.36), we obtain

$$\frac{d\mathbf{n}(t+1)}{d\boldsymbol{\theta}^\mathsf{T}(u)} = \mathbf{A}_t \frac{d\mathbf{n}(t)}{d\boldsymbol{\theta}^\mathsf{T}(u)} + \left(\mathbf{n}^\mathsf{T}(t) \otimes \mathbf{I}_s\right) \frac{d\mathrm{vec}\,\mathbf{A}_t}{d\boldsymbol{\theta}^\mathsf{T}(u)} \tag{7.37}$$

which has the same form as (7.6) except that the matrix and the matrix derivative vary over time.

Some useful simplifications follow from this formulation.

1. Perturbation of matrix elements. If $\boldsymbol{\theta}(t)$ consists of the elements of vec $\mathbf{A}_t$, then

$$\frac{d\mathrm{vec}\,\mathbf{A}_t}{d\boldsymbol{\theta}^\mathsf{T}(t)} = \mathbf{I}_{s^2} \tag{7.38}$$

and can be eliminated from the expressions where it appears.

2. No time travel. Suppose that $\boldsymbol{\theta}(t)$ is perturbed at some time $t = u$. Then

$$\frac{d\mathrm{vec}\,\mathbf{A}_t}{d\boldsymbol{\theta}^\mathsf{T}(u)} = \mathbf{0}_{s^2 \times p} \qquad \text{for } t < u \tag{7.39}$$

However, the effects of the perturbation continue after $t = u$, so that $d\mathbf{n}(t)/d\boldsymbol{\theta}^\mathsf{T}(u)$ will generally be non-zero for $t > u$.

3. Perturbations at every time. A permanent modification of the parameters can be considered a perturbation of $\boldsymbol{\theta}(t)$ for every time $t = 0, 1, \ldots$, so that

$$\boldsymbol{\theta}(t) \longrightarrow \boldsymbol{\theta}(t) + d\boldsymbol{\theta}. \tag{7.40}$$

The sensitivity of the population vector is then

$$\frac{d\mathbf{n}(t+1)}{d\boldsymbol{\theta}^{\mathsf{T}}} = \mathbf{A}_t \frac{d\mathbf{n}(t)}{d\boldsymbol{\theta}^{\mathsf{T}}} + \left(\mathbf{n}^{\mathsf{T}}(t) \otimes \mathbf{I}_s\right) \frac{d\operatorname{vec}\mathbf{A}_t}{d\boldsymbol{\theta}^{\mathsf{T}}} \tag{7.41}$$

4. Perturbation over a range of times. One might be interested in perturbation over some time period $T_1 \leq t \leq T_2$. The effect of such a perturbation on transient dynamics is

$$\frac{d\mathbf{n}(t+1)}{d\boldsymbol{\theta}^{\mathsf{T}}(u)} = \mathbf{A}_t \frac{d\mathbf{n}(t)}{d\boldsymbol{\theta}^{\mathsf{T}}(u)} + \left(\mathbf{n}^{\mathsf{T}}(t) \otimes \mathbf{I}_s\right) J(t) \frac{d\operatorname{vec}\mathbf{A}_t}{d\boldsymbol{\theta}^{\mathsf{T}}(u)} \tag{7.42}$$

where $J(t)$ is an indicator variable

$$J(t) = \begin{cases} 1 & T_1 \leq t \leq T_2 \\ 0 & \text{otherwise} \end{cases} \tag{7.43}$$

These calculations have been extended to apply to population projections (Caswell and Sanchez Gassen 2015; Sanchez Gassen and Caswell 2018); see Sect. 7.8 below.

7.6   Sensitivity of Subsidized Populations

An interesting special case of time-varying models is that of subsidized populations (e.g., Pascual and Caswell 1991), which receive an input of individuals[2]

$$\mathbf{n}(t+1) = \mathbf{A}_t \mathbf{n}(t) + \mathbf{b}(t). \tag{7.44}$$

The subsidy vector $\mathbf{b}(t)$ might represent immigration, or the introduction of individual animals from a captive release program, or dispersal of the larvae of marine invertebrates or the seeds of plants. If $\mathbf{b}(t) < 0$, then it could represent the removal or harvest of individuals from the population (e.g., Hauser et al. 2006).[3]
Differentiating gives:

$$\frac{d\mathbf{n}(t+1)}{d\boldsymbol{\theta}^{\mathsf{T}}} = \mathbf{A}_t \frac{d\mathbf{n}(t)}{d\boldsymbol{\theta}^{\mathsf{T}}} + \left(\mathbf{n}^{\mathsf{T}}(t) \otimes \mathbf{I}_s\right) \frac{d\operatorname{vec}\mathbf{A}_t}{d\boldsymbol{\theta}^{\mathsf{T}}} + \frac{d\mathbf{b}(t)}{d\boldsymbol{\theta}^{\mathsf{T}}}. \tag{7.45}$$

---

[2]See Chap. 10 and Caswell (2008) for analysis of the equilibria of both linear and nonlinear versions of this equation, with applications to organizational dynamics and marine invertebrates.

[3]This type of harvest is unstable in the long run, but we are dealing here with transient dynamics.

If $\boldsymbol{\theta}$ affects only the vital rates and not the subsidy process, then $d\mathbf{n}(t)/\boldsymbol{\theta}^{\mathsf{T}}$ reduces to (7.37), and subsidy affects the sensitivity only through its effect on $(\mathbf{n}^{\mathsf{T}}(t) \otimes \mathbf{I}_s)$. On the other hand, setting $\boldsymbol{\theta} = \mathbf{b}$ gives the effect of changes in the subsidy process:

$$\frac{d\mathbf{n}(t+1)}{d\mathbf{b}^{\mathsf{T}}} = \mathbf{A}_t \frac{d\mathbf{n}(t)}{d\mathbf{b}^{\mathsf{T}}} + \mathbf{I}_s. \tag{7.46}$$

Example: A subsidized model for the reintroduction of the Griffon vulture
The griffon vulture (*Gyps fulvus*) was once widely distributed in Europe, but has been eliminated from many areas, due primarily to poisoning and shooting. A reintroduction program has re-established a population in the Massif Central of southern France; Sarrazin and Legendre (2000) have analyzed this program. Reintroduction programs are increasingly important in conservation biology (Sarrazin and Barbault 1996; Snyder and Snyder 2000), and will become an important application of subsidized models. Transient dynamics are naturally critical for evaluating reintroduction programs, because the programs are of finite duration and are evaluated by short-term measures of success at, or shortly after, their conclusion.

In the case of the griffon vulture, birds can be introduced as juveniles or adults. Adults introduced from captivity have lower fertility and lower survival than wild adults. Here I use a simplification of the Sarrazin-Legendre model to show how transient sensitivity analysis could be used. The life cycle contains four age classes and a stage representing captive-reared adults (Fig. 7.3a). The survival of released adults is a fraction p of that of wild adults, and their fertility a fraction q of that of the wild adults. I assume these costs persist indefinitely; Sarrazin and Legendre (2000) explore both short- and long-term costs. Suppose that a manager is interested in the effects of the annual number b_1 of juveniles released, the number b_5 of adults released, and the relative survival p and relative fertility q of captive-reared adults.

One measure of success will be the population size at the end of the introduction program. The best such population, in terms of future population size, would be one with the highest total reproductive value, $N = \mathbf{v}^{\mathsf{T}}\mathbf{n}$ (also called the stable equivalent population; see Chapters 8–9 of Keyfitz and Caswell 2005). The elasticity of stable equivalent population size[4] is

$$\frac{\epsilon N}{\epsilon \boldsymbol{\theta}^{\mathsf{T}}} = \frac{1}{\mathbf{v}^{\mathsf{T}}\mathbf{n}(t)} \mathbf{v}^{\mathsf{T}} \frac{d\mathbf{n}(t)}{d\boldsymbol{\theta}^{\mathsf{T}}} \mathcal{D}(\boldsymbol{\theta}) \qquad t = 1, \ldots, T \tag{7.47}$$

where $\mathbf{v}$ is the reproductive value vector from $\mathbf{A}$ and $\boldsymbol{\theta}^{\mathsf{T}} = \begin{pmatrix} b_1 & b_5 & p & q \end{pmatrix}$.

---

[4]The parameters under investigation here do not affect the reproductive value vector $\mathbf{v}$. To analyze the sensitivity of stable equivalent population to, say, σ_i, would require the derivative of $\mathbf{v}$ as well; this is presented in Chap. 10.

(a) Life cycle graph (b) Elasticity
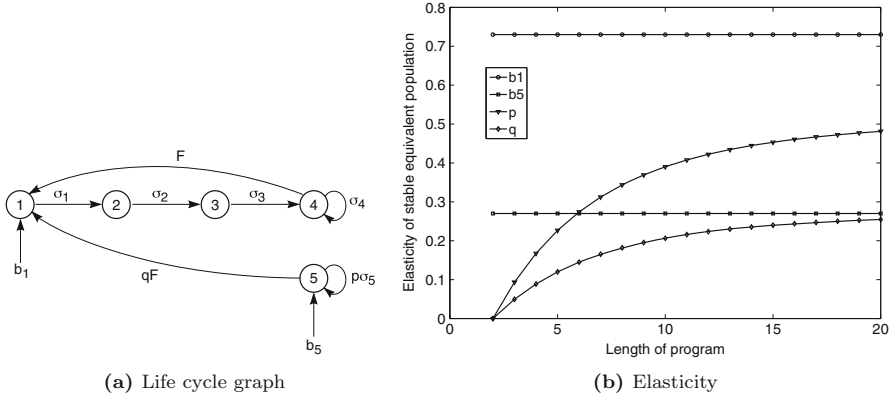
Fig. 7.3 (**a**) The life cycle graph and (**b**) the transient elasticity of stable equivalent population size $N(t) = \mathbf{v}^\mathsf{T}\mathbf{n}(t)$ to changes in juvenile introductions (b_1), adult introductions (b_5), adult survival costs (p), and adult fertility costs (q) for the Griffon vulture. Parameter values from Sarrazin and Legendre (2000); $\sigma_j = 0.86$, $\sigma_a = 0.98$, $f = 0.33$, $p = 0.75$, $q = 0.51$

Using parameter values in Sarrazin and Legendre (2000) and setting $b_1 = b_5$ (i.e., evaluating the value of juveniles and adults from a situation where they are introduced in equal numbers) gives the result in Fig. 7.3b, for an introduction program duration of up to 20 years.

It is always better to increase the number of juveniles relative to the number of adults introduced. The benefits of reducing survival and fertility costs (i.e., increasing p or q) increases with the duration of the program, as they have longer times available to operate. Reductions in the survival cost would have more impact than reductions in the fertility costs. These results are strongly influenced by the fact that the reproductive value of captive-reared adults is lower than that of newly fledged or released juveniles, which is reflected in the high elasticity of $N(t)$ to juvenile releases. ∎

7.7 Sensitivity of Nonlinear Models

In density- or frequency-dependent models, the vital rates depend on the parameters $\boldsymbol{\theta}$ and current population density $\mathbf{n}(t)$:

$$\mathbf{n}(t+1) = \mathbf{A}[\boldsymbol{\theta}, \mathbf{n}(t)] \, \mathbf{n}(t). \tag{7.48}$$

Changes in $\boldsymbol{\theta}$ affect dynamics directly, through $\mathbf{A}$, and indirectly, through $\mathbf{n}(t)$. The transient sensitivity of $\mathbf{n}(t)$ to parameter changes must include both effects.

Differentiating both sides of (7.48) and applying the vec operator gives the familiar differential expression

$$d\mathbf{n}(t+1) = \mathbf{A}[\boldsymbol{\theta}, \mathbf{n}(t)]d\mathbf{n}(t) + \left(\mathbf{n}^\mathsf{T}(t) \otimes \mathbf{I}_s\right) d\mathrm{vec}\, \mathbf{A}[\boldsymbol{\theta}, \mathbf{n}(t)]. \tag{7.49}$$

But now, unlike in the linear case, $d\text{vec}\,\mathbf{A}$ includes both direct effects through $\boldsymbol{\theta}$ and indirect effects through $\mathbf{n}$, so the total differential is

$$d\text{vec}\,\mathbf{A} = \frac{\partial\text{vec}\,\mathbf{A}}{\partial\boldsymbol{\theta}^{\mathsf{T}}}d\boldsymbol{\theta} + \frac{\partial\text{vec}\,\mathbf{A}}{\partial\mathbf{n}^{\mathsf{T}}}\frac{\partial\mathbf{n}(t)}{\partial\boldsymbol{\theta}^{\mathsf{T}}}d\boldsymbol{\theta}. \tag{7.50}$$

Substituting (7.50) into (7.49) gives

$$\frac{d\mathbf{n}(t+1)}{d\boldsymbol{\theta}^{\mathsf{T}}} = \mathbf{A}[\boldsymbol{\theta}, \mathbf{n}(t)]\,\frac{d\mathbf{n}(t)}{d\boldsymbol{\theta}^{\mathsf{T}}}$$

$$+ \left(\mathbf{n}^{\mathsf{T}}(t)\otimes\mathbf{I}_s\right)\frac{\partial\text{vec}\,\mathbf{A}[\boldsymbol{\theta}, \mathbf{n}(t)]}{\partial\boldsymbol{\theta}^{\mathsf{T}}}$$

$$+ \left(\mathbf{n}^{\mathsf{T}}(t)\otimes\mathbf{I}_s\right)\frac{\partial\text{vec}\,\mathbf{A}[\boldsymbol{\theta}, \mathbf{n}(t)]}{\partial\mathbf{n}^{\mathsf{T}}(t)}\frac{d\mathbf{n}(t)}{d\boldsymbol{\theta}^{\mathsf{T}}}. \tag{7.51}$$

The first two terms are familiar from the density-independent case; the third term accounts for the effects of $\boldsymbol{\theta}$ on $\mathbf{A}$ through its effects on $\mathbf{n}(t)$. Rearranging terms gives the transient sensitivity,

$$\frac{d\mathbf{n}(t+1)}{d\boldsymbol{\theta}^{\mathsf{T}}} = \left\{\mathbf{A}[\boldsymbol{\theta}, \mathbf{n}(t)] + \left(\mathbf{n}^{\mathsf{T}}(t)\otimes\mathbf{I}_s\right)\frac{\partial\text{vec}\,\mathbf{A}[\boldsymbol{\theta}, \mathbf{n}(t)]}{\partial\mathbf{n}^{\mathsf{T}}(t)}\right\}\frac{d\mathbf{n}(t)}{d\boldsymbol{\theta}^{\mathsf{T}}}$$

$$+ \left(\mathbf{n}^{\mathsf{T}}(t)\otimes\mathbf{I}_s\right)\frac{\partial\text{vec}\,\mathbf{A}[\boldsymbol{\theta}, \mathbf{n}(t)]}{\partial\boldsymbol{\theta}^{\mathsf{T}}}. \tag{7.52}$$

Example: Transient sensitivity of *Tribolium* Flour beetles of the genus *Tribolium* have been used for a series of models of, and experiments on, nonlinear dynamics, reviewed by Cushing et al. (2003). *Tribolium* lives in stored flour. Adults and larvae cannibalize eggs, and adults cannibalize pupae; these interactions provide the density-dependence, and are captured in a three-stage (larvae, pupae, and adults) model, with

$$\mathbf{A}[\boldsymbol{\theta}, \mathbf{n}] = \begin{pmatrix} 0 & 0 & b\exp(-c_{el}n_1 - c_{ea}n_3) \\ 1 - \mu_l & 0 & 0 \\ 0 & \exp(-c_{pa})n_3 & 1 - \mu_a \end{pmatrix} \tag{7.53}$$

where b is the clutch size, c_{ea}, c_{el}, and c_{pa} are cannibalism rates (of eggs by adults, eggs by larvae, and pupae by adults), and μ_l and μ_a are larval and adult mortalities. Parameter values from experiments reported by Costantino et al. (1997) give the transient dynamics in Fig. 7.4, following introduction of a single adult.

The sensitivity of this transient behavior requires the derivatives of $\mathbf{A}[\boldsymbol{\theta}, \mathbf{n}]$ to the parameters and to the densities. Substituting these derivatives into (7.52) gives the transient sensitivities by a simple iteration. The derivative matrices are given in an appendix to Caswell (2007).

(a) Dynamics   **(b)** Elasticity
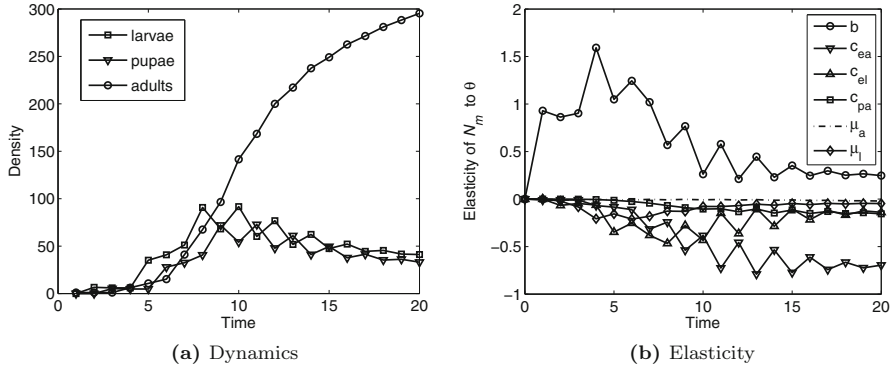
Fig. 7.4 (**a**) The transient dynamics of the *Tribolium* model following introduction of a single adult. Parameters from Costantino et al. (1997). (**b**) The transient elasticity of the metabolic population size $N_m(t)$ to each of the parameters of the *Tribolium* model, for the first 20 time steps following the introduction of a single adult

Tribolium is a pest. The damage it causes might, I suppose, be related to its consumption, which might be measured by the metabolic rate. Emekci et al. (2001) estimated the per capita metabolic rate of larvae, pupae, and adults. Using their results, we define the metabolic population size as $N_m(t) = \mathbf{c}^\mathsf{T}\mathbf{n}(t)$ where $\mathbf{c}^\mathsf{T} = \begin{pmatrix} 9 & 1 & 4.5 \end{pmatrix}$ μl CO_2 h^{-1}. The elasticities of $N_m(t)$ to the parameters are

$$\frac{\epsilon N_m}{\epsilon \boldsymbol{\theta}^\mathsf{T}} = \frac{1}{N_m(t)}\,\mathbf{c}^\mathsf{T}\frac{d\mathbf{n}(t)}{d\boldsymbol{\theta}^\mathsf{T}}\,\mathcal{D}(\boldsymbol{\theta}). \tag{7.54}$$

for $t = 1, \ldots, 20$.

The results are shown in Fig. 7.4. For the first 5 or so iterations, N_m is more elastic to the clutch size than to the cannibalism or mortality rates. After that, the impact of b declines and the impact (negative) of the cannibalism coefficients increases. Beyond 10 time steps, N_m is affected primarily by b (positively) and c_{ea} (*negatively*). Changes in mortality (μ_a and μ_l) have only small effects. Such changes in the relative impact of the parameters over short periods of time are typical of transient sensitivities. Interestingly, the elasticities of total population size $N_{tot} = \sum n_i$ (not shown) show a similar pattern, but lack the period-2 fluctuation evident in Fig. 7.4. This reflects the interaction of the weighting pattern (much more uneven in the calculation of N_m than N_{tot}) and transient fluctuations in the stage distribution. Asymptotic sensitivity calculations are unaffected by such differences.

The parameter values used here lead to a stable equilibrium, but the transient calculations apply equally to other types of dynamics.   ∎

7.8 Sensitivity of Population Projections

The most common transient analyses of populations appear in the population projections provided by local, national, and international offices. These projections are usually carried out by the cohort component method, which uses mortality, fertility, and migration to describe the dynamics of each age×sex combination. The calculations are transient because the begin with the current, rather than an asymptotic, age-sex distribution and are carried out over a short time horizon (usually a few decades). In the first issue of the first volume of the then-new journal *Demography*, Nathan Keyfitz described the "population projection as a matrix operator" (Keyfitz 1964). He showed that population projections using the cohort component method could be written as matrix population models, and emphasized the value in doing so to focus attention on the mathematical structure of the projection, inviting deeper analyses of its properties with more powerful mathematical tools. Considering projections as matrix operators allows the use of matrix calculus methods to develop a thorough perturbation analysis of population projections (Caswell and Sanchez Gassen 2015; Sanchez Gassen and Caswell 2018).

To present the basics of projection sensitivity analysis, we begin with a simple one-sex model, but we focus most of our attention on a two-sex model that includes separate rates for males and females.

The single-sex projection can be written as

$$\mathbf{n}(t+1) = \mathbf{A}(t)\mathbf{n}(t) + \mathbf{b}(t) \qquad \mathbf{n}(0) = \mathbf{n}_0 \tag{7.55}$$

where $\mathbf{n}(t)$ is a vector whose entries are the numbers of individuals in each age class or stage at time t, $\mathbf{A}(t)$ is a projection matrix incorporating the vital rates at time t, and $\mathbf{b}(t)$ is a vector giving the number of immigrants in each age class or stage at time t. The projection begins with a specified initial condition, denoted $\mathbf{n}_0$, and is carried out until some target time T.

To develop a two-sex projection, we define population vectors $\mathbf{n}_f$ and $\mathbf{n}_m$, and projection matrices $\mathbf{A}_f$ and $\mathbf{A}_m$, for females and males, respectively. We assume that reproduction is female dominant,[5] so all fertility is attributed to females. We decompose the projection matrices for females and males into

$$\mathbf{A}_f(t) = \mathbf{U}_f(t) + \phi \mathbf{F}(t) \tag{7.56}$$

$$\mathbf{A}_m(t) = \mathbf{U}_m(t) \tag{7.57}$$

where $\mathbf{U}$ describes transitions and survival of extant individuals and $\mathbf{F}$ describes the production of new individuals by reproduction.

---

[5]Two-sex models that do not assume dominance by one sex have been used to project animal populations, but not, as far as I know, human populations (e.g., Jenouvrier et al. 2009, 2010, 2012).

In an age-classified model, $\mathbf{F}$ will have effective fertilities (including infant and maternal survival as appropriate) on the first row and zeros elsewhere. A proportion ϕ of the offspring are female. This model attributes reproduction to females; hence there is no need to create separate fertility matrices for reproduction by males and females.

The male component of the population is projected by the survival matrix $\mathbf{U}_m$; the input of new individuals comes from the female population. The projection model becomes

$$\mathbf{n}_f(t+1) = \left[\mathbf{U}_f(t) + \phi \mathbf{F}(t) \right] \mathbf{n}_f(t) + \mathbf{b}_f(t) \tag{7.58}$$

$$\mathbf{n}_m(t+1) = \mathbf{U}_m(t)\mathbf{n}_m(t) + (1-\phi)\mathbf{F}(t)\mathbf{n}_f(t) + \mathbf{b}_m(t) \tag{7.59}$$

The sensitivity of the two-sex projection is given by the two derivatives,

$$\frac{d\mathbf{n}_f(t)}{d\boldsymbol{\theta}^{\mathsf{T}}(u)} \quad \text{and} \quad \frac{d\mathbf{n}_m(t)}{d\boldsymbol{\theta}^{\mathsf{T}}(u)} \qquad t, u = 0, \dots, T.$$

These sensitivities are obtained from dynamic expressions, for the female population

$$\underbrace{\frac{d\mathbf{n}_f(t+1)}{d\boldsymbol{\theta}^{\mathsf{T}}(u)}}_{\text{sensitivity at } t+1} = \underbrace{\left(\mathbf{U}_f(t) + \phi \mathbf{F}(t)\right) \frac{d\mathbf{n}_f(t)}{d\boldsymbol{\theta}^{\mathsf{T}}(u)}}_{\text{sensitivity at } t} + \underbrace{\left(\mathbf{n}_f^{\mathsf{T}}(t) \otimes \mathbf{I}_\omega\right) \left(\frac{d\,\mathrm{vec}\,\mathbf{U}_f(t)}{d\boldsymbol{\theta}^{\mathsf{T}}(u)} + \phi \frac{d\,\mathrm{vec}\,\mathbf{F}(t)}{d\boldsymbol{\theta}^{\mathsf{T}}(u)} \right)}_{\text{effects via female transitions and fertility}}$$

$$+ \underbrace{\frac{d\mathbf{b}_f(t)}{d\boldsymbol{\theta}^{\mathsf{T}}(u)}}_{\text{effects via immigration}} \tag{7.60}$$

and the male population

$$\underbrace{\frac{d\mathbf{n}_m(t+1)}{d\boldsymbol{\theta}^{\mathsf{T}}(u)}}_{\text{sensitivity at } t+1} = \underbrace{\mathbf{U}_m(t)\frac{d\mathbf{n}_m(t)}{d\boldsymbol{\theta}^{\mathsf{T}}(u)} + (1-\phi)\mathbf{F}(t)\frac{d\mathbf{n}_f(t)}{d\boldsymbol{\theta}^{\mathsf{T}}(u)}}_{\text{sensitivities at } t} + \underbrace{\left(\mathbf{n}_m^{\mathsf{T}}(t) \otimes \mathbf{I}_\omega\right) \frac{d\,\mathrm{vec}\,\mathbf{U}_m(t)}{d\boldsymbol{\theta}^{\mathsf{T}}(u)}}_{\text{effects via male transitions}}$$

$$+ \underbrace{(1-\phi)\left(\mathbf{n}_f^{\mathsf{T}}(t) \otimes \mathbf{I}_\omega\right) \frac{d\,\mathrm{vec}\,\mathbf{F}(t)}{d\boldsymbol{\theta}^{\mathsf{T}}(u)}}_{\text{effects via female fertility}} + \underbrace{\frac{d\mathbf{b}_m(t)}{d\boldsymbol{\theta}^{\mathsf{T}}(u)}}_{\text{effects via immigration}} \tag{7.61}$$

Equations (7.60) and (7.61) are iterated from initial conditions

$$\frac{d\mathbf{n}_f(0)}{d\boldsymbol{\theta}^{\mathsf{T}}(u)} = \frac{d\mathbf{n}_m(0)}{d\boldsymbol{\theta}^{\mathsf{T}}(u)} = \mathbf{0}_{\omega \times p} \tag{7.62}$$

along with the iteration of equations (7.58) and (7.59) for the population vectors $\mathbf{n}_f(t)$ and $\mathbf{n}_m(t)$. For complete details, see Caswell and Sanchez Gassen (2015).

The terms in (7.61) are labelled to show how the processes of transitions, fertility and migration, for males and females, combine to produce sensitivity of a transient population. As before, the sensitivity at $t + 1$ depends on the sensitivity at time t and on the effects of the parameter vector on the transition and fertility matrices and on the immigration vector. In the next section we turn to the calculation of these derivatives.

The elasticities of $\mathbf{n}_f(t)$ are given by

$$\frac{\epsilon \mathbf{n}_f(t)}{\epsilon \boldsymbol{\theta}^{\mathsf{T}}(u)} = \mathcal{D}\left[\mathbf{n}_f(t)\right]^{-1} \frac{d\mathbf{n}_f(t)}{d\boldsymbol{\theta}^{\mathsf{T}}(u)} \mathcal{D}\left[\boldsymbol{\theta}(u)\right] \tag{7.63}$$

with a similar expression for $\mathbf{n}_m$.

Caswell and Sanchez Gassen (2015) present a detailed analysis of a projection for the population of Spain, published by the Instituto Nacional de Estadística (INE), for the years 2012–2052. They calculated the sensitivity and elasticity of total population, male and female population, the school age population (6–16 years), the part of the population expected to suffer from dementia, and the dependency and support ratios. All these outcomes are calculated from the basic projection using the methods in Sect. 7.3. In a more extensive comparison, Sanchez Gassen and Caswell (2018) have applied the approach to the Europop2013 projections for the 28 member states of the European Union, plus Iceland, Norway, and Sweden, for the years 2013–2080.

7.9 Discussion

In addition to their obvious role in population projections, transient effects are critically important in studies of climate change and other short term management issues (Ezard et al. 2010). A recent study found that simulations of invasive species were strongly influenced by transient effects (Muthukrishnan et al. 2018). Matrix calculus makes transient sensitivity analysis straightforward and applicable to a wide range of models and perturbations. The approach calculates sensitivities and elasticities as a dynamic system, iterated in parallel with the dynamics of the transient solution itself.

This dynamic approach reveals the fundamental structure underlying the sensitivity calculation. The results bear a striking family resemblance, from the linear, time-invariant case (7.6), to the time-varying case (7.41), the case of subsidized populations (7.45), the nonlinear case (7.52), and the time-varying, two-sex, subsidized model that forms the basis for the cohort component method of population projection in equations (7.61) and (7.60).

The examples here sound like stories—*suppose that* someone (e.g., a manager) is interested in some aspect of the population (e.g., its total size, or variance, or

average growth,...) over some time interval. Or *suppose* that mortality, fertility, and immigration develop in the following way. This emphasizes the flexibility of this approach, and also the importance about thinking clearly about the dependent variables and time scales of interest. The list of dependent variables in Sect. 7.3 can no doubt be extended. It may be repeating the obvious, but transient sensitivity analysis depends on initial conditions. Each of the examples had to choose an initial condition and argue for its relevance.

Section 10.2.6 in Chap. 10 briefly considers the sensitivity analysis of equilibria to continuous-time systems. Richard et al. (2015) have developed a very general sensitivity analysis of transient dynamics in continuous systems (both linear and nonlinear). They point out and nicely demonstrate the parallels between continuous-time models and the discrete-time models considered here, the link being the creation of a dynamic model for the sensitivities that is solved along with the dynamics of the system itself.

Bibliography

Caswell, H. 2007. Sensitivity analysis of transient population dynamics. Ecology Letters **10**:1–15.

Caswell, H. 2008. Perturbation analysis of nonlinear matrix population models. Demographic Research **18**:59–116.

Caswell, H., and M. G. Neubert. 2005. Reactivity and transient dynamics of discrete-time ecological systems. Journal of Difference Equations and Applications **11**:295–310.

Caswell, H., and N. Sanchez Gassen. 2015. The sensitivity analysis of population projections. Demographic Research **33**:801–840.

Caswell, H., and P. A. Werner. 1978. Transient behavior and life history analysis of teasel (*Dipsacus sylvestris* Huds.). Ecology **59**:53–66.

Coale, A. J. 1972. The growth and structure of human populations: a mathematical approach. Princeton University Press, Princeton, New Jersey.

Costantino, R. F., R. A. Desharnais, J. M. Cushing, and B. Dennis. 1997. Chaotic dynamics in an insect population. Science **275**:389–391.

Cushing, J. M., R. F. Costantino, B. Dennis, R. A. Desharnais, and S. M. Henson. 2003. Chaos in ecology: experimental nonlinear dynamics. Academic Press, San Diego, California, USA.

Emekci, M., S. Navarro, J. E. Donahaye, M. Rindner, and A. Azrieli, 2001. Respiration of stored product pests in hermetic conditions. Pages 26–35 *in* J. Donahaye, S. Navarro, and J. G. Leesch, editors. Proceedings of the International Conference on Controlled Atmosphere and Fumigation in Stored Products. Executive Printing Services, Clovis, California, USA.

Ezard, T. H. G., J. M. Bullock, H. J. Dalgleish, A. Millon, F. Pelletier, A. Ozgul, and D. N. Koons. 2010. Matrix models for a changeable world: the importance of transient dynamics in population management. Journal of Applied Ecology **47**:515–523.

Fox, G. A., and J. Gurevitch. 2000. Population numbers count: tools for near-term demographic analysis. American Naturalist **156**:242–256.

Hauser, C. E., E. G. Cooch, and J.-D. Lebreton. 2006. Control of structured populations by harvest. Ecological Modelling **196**:462–470.

Jenouvrier, S., H. Caswell, C. Barbraud, M. Holland, J. Strœve, and H. Weimerskirch. 2009. Demographic models and IPCC climate projections predict the decline of an emperor penguin population. Proceedings of the National Academy of Sciences **106**:1844–1847.

Jenouvrier, S., H. Caswell, C. Barbraud, and H. Weimerskirch. 2010. Mating behavior, population growth, and the operational sex ratio: a periodic two-sex model approach. The American Naturalist **175**:739–752.

Jenouvrier, S., M. Holland, J. Stroeve, C. Barbraud, H. Weimerskirch, M. Serreze, and H. Caswell. 2012. Effects of climate change on an emperor penguin population: analysis of coupled demographic and climate models. Global Change Biology **18**:2756–2770.

Keyfitz, N. 1964. The population projection as a matrix operator. Demography **1**:56–73.

Keyfitz, N., and H. Caswell. 2005. Applied mathematical demography. 3rd edition. Springer, New York, New York.

Lotka, A. J. 1939. Théorie analytique des associations biologiques. Part II. Analyse démographique avec application particulière à l'espèce humaine. Number 780 in Actualités Scientifiques et Industrielles, Hermann et Cie, Paris, France.

Muthukrishnan, R., A. S. Davis, N. R. Jordan, and J. D. Forester. 2018. Invasion complexity at large spatial scales is an emergent property of interactions among landscape characteristics and invader traits. PLoS ONE **13**:e0195892.

Neubert, M. G., and H. Caswell. 1997. Alternatives to resilience for measuring the responses of ecological systems to perturbations. Ecology **78**:653–665.

Pascual, M., and H. Caswell. 1991. The dynamics of a size-classified benthic population with reproductive subsidy. Theoretical Population Biology **39**:129–147.

Richard, R., J. Casas, and E. McCauley. 2015. Sensitivity analysis of continuous-time models for ecological and evolutionary theories. Theoretical Ecology **8**:481–490.

Sanchez Gassen, N., and H. Caswell, 2018. The determinants of population change in Europe 2013–2080: a sensitivity analysis. In prep.

Sarrazin, F., and R. Barbault. 1996. Reintroduction: challenges and lessons for basic ecology. Trends in Ecology and Evolution **11**:474–478.

Sarrazin, F., and S. Legendre. 2000. Demographic approach to releasing adults versus young in reintroductions. Conservation Biology **14**:488–500.

Skalski, J. R., K. E. Ryding, and J. Millspaugh. 2005. Wildlife demography: analysis of sex, age, and count data. Academic Press, San Diego, California, USA.

Snyder, N. F. R., and H. Snyder. 2000. The California condor: a saga of natural history and conservation. Academic Press, San Diego, California, USA.

Verdy, A., and H. Caswell. 2008. Sensitivity analysis of reactive ecological dynamics. Bulletin of Mathematical Biology **70**:1634–1659.

Williams, B. K., J. D. Nichols, and M. J. Conroy. 2002. Analysis and management of animal populations. Academic Press, San Diego, California, USA.

Chapter 8
Periodic Models

8.1 Introduction

Periodic matrix models are often used to study cyclical temporal variation (seasonal or interannual), sometimes as a (perhaps crude) approximation to stochastic models. However, formally periodic models also appear when multiple processes (e.g., demography and dispersal) operate within a single projection interval. The models take the form of periodic matrix products. A familiar example is when population projection over an annual interval is described as a product of seasonal operators. The perturbation analysis of periodic models (Caswell and Trevisan 1994; Lesnoff et al. 2003; Caswell and Shyu 2012) must specify both the vital rates affected by the perturbation and the timing of the perturbation within the cycle. This chapter presents a general approach to the perturbation analysis of both linear and nonlinear periodic models. The results consist of a series of analyses of some of the most commonly encountered periodic models.

If the environment is time-invariant on the scale of a chosen projection interval (e.g., from year to year), the result is a periodic matrix population model in which the seasonal product repeats itself. Such a model can be written as

$$\mathbf{n}(t + 1) = \mathbf{B}_p \cdots \mathbf{B}_2 \mathbf{B}_1 \mathbf{n}(t) \tag{8.1}$$

Here, $\mathbf{B}_i$ is the matrix at phase i of the cycle and p is the period. The period is the number of phases in the cycle; i.e., the number of matrices in the periodic matrix product in (8.1). Neither the identities nor the number of stages need be the same from one phase to the next, so the matrices $\mathbf{B}_i$ may be rectangular rather than square.

The phases need not be the same length, so the period may or may not be measured in units of time. For example, in the model of Pico et al. (2002), each season is of 2 months duration, and the period ($p = 6$) corresponds directly to a time scale. In contrast, the model of Hunter and Caswell (2005a) has three phases, with durations of 3 weeks, 5 weeks, and 10 months, respectively. The period ($p = 3$) of that model does not correspond to a time scale, but it identifies the number of matrices in the periodic product and appears in calculations in the same role as $p = 6$ in the model of Pico et al. (2002).

The projection matrix over the entire periodic cycle is[1]

$$\mathbf{A} = \mathbf{B}_p \cdots \mathbf{B}_2 \mathbf{B}_1 \tag{8.2}$$

The earliest studies of periodic matrix models were due to Darwin and Williams (1964), Skellam (1966), and MacArthur (1968). In recent years, with little fanfare, periodic models have emerged as an important tool for incorporating multiple processes within a single projection interval. Uses of periodic models include the following.

1. Seasonal variation. Plants and animals experience obvious and dramatic seasonal variation in their demographic rates. Periodic models have been used to describe this variation, with seasons variously defined in terms of monthly periods, calendar seasons, or in terms of environmental events such as rainfall or flood patterns (e.g., Smith et al. 2005).

   Although annual or near-annual species are obvious candidates for periodic models, within-year time scales may also be important for long-lived species. For example, Hunter and Caswell (2005a) incorporated chick development events on a time scale of weeks into a periodic model for the sooty shearwater, which has a lifetime of decades. Similarly, Jenouvrier et al. (2010, 2014) have used periodic models to capture the timing of events in the breeding cycle within a portion of the year in the long-lived emperor penguin.

---

[1]Although we will not address it in this chapter, the model (8.1) can be written in a way that explicitly defines the starting phase in the cycle. As written, $\mathbf{A}$ in (8.2) projects from phase 1 to phase 1; if desired we could write this as $\mathbf{A}_1$ and define matrices

$$\mathbf{A}_2 = \mathbf{B}_1 \mathbf{B}_p \cdots \mathbf{B}_2$$

$$\vdots$$

$$\mathbf{A}_p = \mathbf{B}_{p-1} \cdots \mathbf{B}_1 \mathbf{B}_p$$

The $\mathbf{A}_i$ are obtained by cyclic permutations of the sequence $\{\mathbf{B}_p, \ldots, \mathbf{B}_1\}$; each of these projects from a different phase in the cycle. Some demographic properties (e.g., the population growth rate λ) are invariant with respect to such permutations; others (e.g., the eigenvectors) are not (Caswell 2001). In this chapter, we will start with phase 1 and refer to $\mathbf{A}$ rather than $\mathbf{A}_1$.

2. Periodic interannual variability. Periodic models based on sequences of annual observations have been used to study effects of inter-event intervals, where events include fires, floods, ENSO events, etc.
3. Harvest and management. These activities often take place at specified points within an annual or interannual cycle. Periodic models have been used to study the effects of their timing; one of the earliest periodic models being devoted to seasonal harvesting (Darwin and Williams 1964).
4. Conditional probabilities. Periodic matrix products appear when models are written as products of conditional probabilities. In stage-classified models, for example, a transition matrix $\mathbf{U}$, is written as the product of a diagonal matrix $\boldsymbol{\Sigma}$ (with survival probabilities $\boldsymbol{\sigma}$ on the diagonal) and a matrix $\mathbf{G}$ of transition probabilities conditional on survival:

$$\mathbf{U} = \mathbf{G}\boldsymbol{\Sigma} \tag{8.3}$$

   which creates a period-2 periodic matrix product within the model.
5. Multistate vec-permutation models. When individuals are classified by two or more criteria (e.g., stage and location), the dynamics over the projection interval can be described in terms of the processes affecting each criterion (e.g., transitions and movement). The result is a periodic model that uses the vec-permutation matrix to generate a block-structured projection matrix over the entire interval. See Chap. 6 for analysis of such models.
6. Nonlinear models. Henson and Cushing (1997) developed a model for *Tribolium* in an experimental system in which container size was varied periodically. Shyu et al. (2013) developed a nonlinear seasonal model of an invasive plant to account for the timing of both density effects and management actions within the year. In such models, cyclic dynamics can be produced both by the environmental periodicity and the nonlinearities (e.g., Cushing 2006).

8.1.1 Perturbation Analysis

As in Fig. 8.1, we suppose that in phase i of the cycle, the parameter vector takes on the value $\boldsymbol{\theta}_i$ and determines the matrix $\mathbf{B}_i$. The projection matrix $\mathbf{A}$ is the product, in the specified order, of the $\mathbf{B}_i$. Although the output $\boldsymbol{\xi}$ is calculated from $\mathbf{A}$, the parameter dependence operates through the $\mathbf{B}_i$ (Fig. 8.1). The sensitivity of $\boldsymbol{\xi}$ to the elements of $\mathbf{A}$ is in general not of interest, because those elements are complicated expressions involving the elements of all the $\mathbf{B}_i$, and thus mix disparate biological processes. Here we calculate the sensitivity of demographic outcomes to the entries of the $\mathbf{B}_i$.

In this chapter we analyze linear periodic models of the form (8.1) and the cyclic dynamics of nonlinear seasonal models with delayed density effects. We will briefly discuss the generalization of the multistate age×stage-classified models

Table 8.1 Table of symbols in this chapter

Symbol	Meaning
s_i	Number of stages at phase i of the cycle
p	Period of the cycle
q	Dimension of parameter vector $\boldsymbol{\theta}$
r	Number of locations in spatial model
$\boldsymbol{\theta}_i$	Parameter vector evaluated at phase i
$\mathbf{B}_i$	Projection matrix from phase i to phase $i + 1$, or in location i
$\mathbf{C}_i^j$	Ordered product $\mathbf{B}_j \cdots \mathbf{B}_i$ of matrices from i to j
$\mathbf{M}_i$	Dispersal matrix for stage i
$\mathbf{A}$	Projection matrix over entire cycle
$\mathbf{A}_i$	Projection matrix over cycle, starting at phase i
$\mathbf{R}_i$	Matrix of LTRE contributions from phase i
$\mathbf{E}_{s,i}$	$s \times s$ matrix with 1 in (i, i) position and 0 elsewhere
$\mathbf{I}_s$	Identity matrix of dimension s
$\mathcal{D}(\mathbf{x})$	Diagonal matrix with $\mathbf{x}$ on the diagonal
$\mathbf{1}$	Vector of ones
$\mathbb{B}, \mathbb{M}$, etc.	Block-structured matrices
$\circ$	Hadamard, or element-by-element product
$\otimes$	Kronecker product
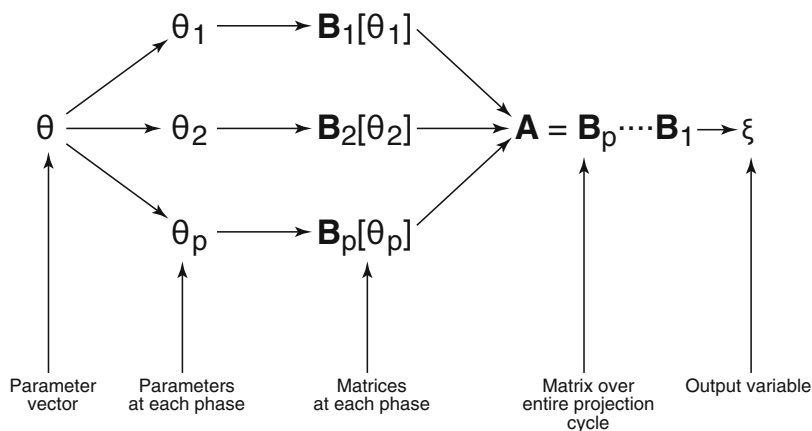
Fig. 8.1 A vector $\boldsymbol{\theta}$ of parameters determines an output variable $\boldsymbol{\xi}$, which may be a scalar, vector, or matrix. The parameter vector will generally take on different values at each phase in the cycle, and determine the phase-specific matrix $\mathbf{B}_i$. These matrices determine the projection matrix $\mathbf{A}$ as a periodic matrix product; the output variable is computed from $\mathbf{A}$. The perturbation problem is to compute the sensitivity or elasticity of $\boldsymbol{\xi}$ to $\boldsymbol{\theta}$

explored in Chap. 6 to an arbitrary number of classifications. We extend the LTRE decomposition analysis to the periodic case, making it possible to analyze effects of parameter changes at any point in a periodic environment.

8.2 Linear Models

Consider the basic model (8.1) with projection matrix (8.2). The period of the cycle is p. To allow for differences in the state vector at different phases within the cycle, define the number of stages at phase i as s_i. Thus the matrix $\mathbf{B}_i$ is of dimension $s_{i+1} \times s_i$, with the subscript i interpreted $\mathrm{mod}(p)$ (that is, $(p + 1)\,\mathrm{mod}(p) = 1$).

Let $\boldsymbol{\xi}$ (dimension $m \times 1$) denote an output variable calculated from $\mathbf{A}$, where $\boldsymbol{\xi}$ might be a scalar, a vector, or a vectorized matrix. Let $\boldsymbol{\theta}$ be a parameter vector (dimension $q \times 1$). The derivative of $\boldsymbol{\xi}$ with respect to $\boldsymbol{\theta}$ is the $m \times q$ matrix

$$\frac{d\boldsymbol{\xi}}{d\boldsymbol{\theta}^{\mathsf{T}}} = \left(\frac{d\xi_i}{d\theta_j} \right) \qquad i = 1, \ldots, m; \; j = 1, \ldots, q \tag{8.4}$$

By the chain rule, the effects of the parameters on $\boldsymbol{\xi}$ are captured in the matrix product

$$\frac{d\boldsymbol{\xi}}{d\boldsymbol{\theta}^{\mathsf{T}}} = \frac{d\boldsymbol{\xi}}{d\mathrm{vec}^{\mathsf{T}}\mathbf{A}} \frac{d\mathrm{vec}\,\mathbf{A}}{d\boldsymbol{\theta}^{\mathsf{T}}}. \tag{8.5}$$

The first term in (8.5) is the derivative of the output variable $\boldsymbol{\xi}$ with respect to the matrix $\mathbf{A}$ from which it is calculated. The second term in (8.5) is the derivative of the periodic product matrix $\mathbf{A}$ with respect to the parameter vector $\boldsymbol{\theta}$. To obtain this, differentiate (8.2), to obtain

$$d\mathbf{A} = \mathbf{B}_p \cdots \mathbf{B}_2 \, (d\mathbf{B}_1)$$
$$+ \mathbf{B}_p \cdots (d\mathbf{B}_2)\, \mathbf{B}_1$$
$$\vdots$$
$$+ (d\mathbf{B}_p)\, \mathbf{B}_{p-1} \cdots \mathbf{B}_1 \tag{8.6}$$

It is convenient to define the matrix $\mathbf{C}_i^j$ as the ordered product (from right to left) of the $\mathbf{B}$ matrices from i up to j:

$$\mathbf{C}_i^j = \mathbf{B}_j \cdots \mathbf{B}_i \qquad i \le j \tag{8.7}$$

and set $\mathbf{C}_1^0 = \mathbf{C}_{p+1}^p = \mathbf{I}_{s_1}$. Then (8.6) becomes

$$d\mathbf{A} = \mathbf{C}_2^p \, (d\mathbf{B}_1) + \mathbf{C}_3^p \, (d\mathbf{B}_2)\, \mathbf{C}_1^1 \cdots + (d\mathbf{B}_p)\, \mathbf{C}_1^{p-1} \tag{8.8}$$

Applying the vec operator to both sides gives

$$d\mathrm{vec}\,\mathbf{A} = \sum_{i=1}^{p} \left[\left(\mathbf{C}_1^{i-1} \right)^{\mathsf{T}} \otimes \mathbf{C}_{i+1}^p \right] \, d\mathrm{vec}\,\mathbf{B}_i \tag{8.9}$$

Equation (8.9) accounts automatically for the possibly different dimensions of the $\mathbf{B}_i$. The resulting derivative with respect to the parameter vector $\boldsymbol{\theta}$ is

$$\frac{d\mathrm{vec}\,\mathbf{A}}{d\boldsymbol{\theta}^{\mathsf{T}}} = \sum_{i=1}^{p} \left[\left(\mathbf{C}_1^{i-1} \right)^{\mathsf{T}} \otimes \mathbf{C}_{i+1}^{p} \right] \frac{d\mathrm{vec}\,\mathbf{B}_i}{d\boldsymbol{\theta}^{\mathsf{T}}}. \tag{8.10}$$

where $d\mathrm{vec}\,\mathbf{B}_i/d\boldsymbol{\theta}^{\mathsf{T}}$ is the derivative of the matrix $\mathbf{B}_i$ with respect to the parameter vector $\boldsymbol{\theta}$, evaluated at $\boldsymbol{\theta}_i$. Equation (8.10) sums the contributions of the derivatives of all of the phase-specific matrices $\mathbf{B}_i$ with respect to $\boldsymbol{\theta}$, thus accounting for all the ways in which $\boldsymbol{\theta}$ may affect the demographic rates at each point in the cycle. As written, (8.10) gives the result of perturbing $\boldsymbol{\theta}$ at each point in the cycle. The effect of a phase-specific perturbation is easily obtained by summing only over phases in which $\boldsymbol{\theta}_i$ is modified.

Substituting (8.10) into the formula (8.5) gives the general expression for the sensitivity of $\boldsymbol{\xi}$ to changes affecting any or all of the $\mathbf{B}_i$:

$$\frac{d\boldsymbol{\xi}}{d\boldsymbol{\theta}^{\mathsf{T}}} = \frac{d\boldsymbol{\xi}}{d\mathrm{vec}^{\mathsf{T}}\mathbf{A}} \left(\sum_{i=1}^{p} \left[\left(\mathbf{C}_1^{i-1} \right)^{\mathsf{T}} \otimes \mathbf{C}_{i+1}^{p} \right] \frac{d\mathrm{vec}\,\mathbf{B}_i}{d\boldsymbol{\theta}^{\mathsf{T}}} \right). \tag{8.11}$$

The elasticity of $\boldsymbol{\xi}$ to $\boldsymbol{\theta}$ is the matrix

$$\frac{\epsilon\boldsymbol{\xi}}{\epsilon\boldsymbol{\theta}^{\mathsf{T}}} = \left(\frac{\theta_j}{\xi_i} \frac{d\xi_i}{d\theta_j} \right) \tag{8.12}$$

$$= \mathcal{D}\left(\boldsymbol{\xi}\right)^{-1} \frac{d\boldsymbol{\xi}}{d\mathrm{vec}^{\mathsf{T}}\mathbf{A}} \left(\sum_{i=1}^{p} \left[\left(\mathbf{C}_1^{i-1} \right)^{\mathsf{T}} \otimes \mathbf{C}_{i+1}^{p} \right] \frac{d\mathrm{vec}\,\mathbf{B}_i}{d\boldsymbol{\theta}^{\mathsf{T}}} \right) \mathcal{D}\left(\boldsymbol{\theta}\right) \tag{8.13}$$

where $\mathcal{D}\left(\mathbf{x}\right)$ is a diagonal matrix with $\mathbf{x}$ on the diagonal and zeros elsewhere. Because elasticities are logarithmic derivatives, they apply only when $\boldsymbol{\xi} > 0$ and $\boldsymbol{\theta} \geq 0$.

8.2.1 A Simple Harvest Model

The projection matrix for a simple harvest model (e.g., Hauser et al. 2006) can be written

$$\mathbf{A} = \mathbf{B}\left(\mathbf{I} - \mathbf{H}\right). \tag{8.14}$$

The matrix $\mathbf{B}$ describes demography in the absence of harvest. The matrix $\mathbf{H} = \mathcal{D}\left(\mathbf{h}\right)$ is a harvest matrix, where h_i is the probability that an individual of stage i

is harvested.[2] Either $\mathbf{B}$, $\mathbf{H}$, or both may be functions of a vector $\boldsymbol{\theta}$ of parameters. Differentiating (8.14) and applying the vec operator gives

$$d\text{vec }\mathbf{A} = -\left(\mathbf{I}_s \otimes \mathbf{B}\right) d\text{vec }\mathbf{H} + \left[(\mathbf{I} - \mathbf{H})^\top \otimes \mathbf{I}_s\right] d\text{vec }\mathbf{B}. \tag{8.15}$$

The diagonal matrix $\mathbf{H}$ can be written

$$\mathbf{H} = \mathbf{I}_s \circ \left(\mathbf{1}_s \mathbf{h}^\top\right) \tag{8.16}$$

where $\mathbf{1}_s$ is a $s \times 1$ vector of ones and $\circ$ denotes the Hadamard product. The differential of $\mathbf{H}$ in (8.16) is

$$d\text{vec }\mathbf{H} = \mathcal{D}\,(\text{vec }\mathbf{I}_s)\,(\mathbf{I}_s \otimes \mathbf{1}_s)\,d\mathbf{h}. \tag{8.17}$$

Combining (8.16) and (8.15) and applying the chain rule gives the derivative with respect to $\boldsymbol{\theta}$:

$$\frac{d\text{vec }\mathbf{A}}{d\boldsymbol{\theta}^\top} = \underbrace{-\left(\mathbf{I}_s \otimes \mathbf{B}\right)\mathcal{D}\,(\text{vec }\mathbf{I}_s)\,(\mathbf{I}_s \otimes \mathbf{1}_s)\,\frac{d\mathbf{h}}{d\boldsymbol{\theta}^\top}}_{\text{perturbations of }\mathbf{h}} + \underbrace{\left[(\mathbf{I} - \mathbf{H})^\top \otimes \mathbf{I}_s\right]\frac{d\text{vec }\mathbf{B}}{d\boldsymbol{\theta}^\top}}_{\text{perturbations of }\mathbf{B}}.$$
$$\tag{8.18}$$

The conditional probability model (8.3) has the same form as the harvest model (8.14), so a similar analysis applies to it as well:

$$d\text{vec }\mathbf{U} = (\mathbf{I} \otimes \mathbf{G})\,\mathcal{D}\,(\text{vec }\mathbf{I})\,(\mathbf{I} \otimes \mathbf{1})\,d\boldsymbol{\sigma} + (\boldsymbol{\Sigma} \otimes \mathbf{I})\,d\text{vec }\mathbf{G}. \tag{8.19}$$

However, the conditional transition matrix $\mathbf{G}$ is column-stochastic (all columns sum to 1), because all loss of individuals is accounted for by $\boldsymbol{\Sigma}$. Thus relevant perturbations must be parameterized so that the stochasticity is preserved. For example, if $\mathbf{G}$ describes growth in the standard size-classified model (Caswell 2001, Section 4.2), e.g.,

$$\mathbf{G} = \begin{pmatrix} 1 - \gamma_1 & 0 & 0 \\ \gamma_1 & 1 - \gamma_2 & 0 \\ 0 & \gamma_2 & 1 - \gamma_3 \end{pmatrix} \tag{8.20}$$

then perturbations of the γ_i will preserve stochasticity of $\mathbf{G}$. If $\mathbf{G}$ has no such convenient parameterization, then changes in the entries of $\mathbf{G}$ must be compensated for by changes elsewhere in the same column (see Caswell 2001; Hill et al. 2004;

---

[2]Alternatively, let μ_i be the mortality due to harvest experienced by an individual in stage i. Then $\mathbf{H} = \exp\left[-\mathcal{D}\,(\boldsymbol{\mu})\right]$. Harvest imposes an additional, additive hazard on top of the natural mortality contained in $\mathbf{B}$.

Theorem 4.5 of Caswell 2013). For explicit formulas for compensation, see Chap. 11 of this volume; for an application, see van Daalen and Caswell (2017).

The harvest model (8.14) can be extended to describe harvest imposed at a specified phase within a p-cycle. Suppose that harvest takes place between phase m and phase $m + 1$, so that

$$\mathbf{A} = \mathbf{B}_p \cdots \mathbf{B}_{m+1} \ (\mathbf{I} - \mathbf{H}) \ \mathbf{B}_m \cdots \mathbf{B}_1 \tag{8.21}$$

(see Darwin and Williams (1964) for an early example of this kind of seasonal harvest model). Using the same approach, it can be shown that

$$
\begin{aligned}
\frac{d\text{vec}\,\mathbf{A}}{d\boldsymbol{\theta}^{\mathsf{T}}} = {} & -\left[\left(\mathbf{C}_1^m\right)^{\mathsf{T}} \otimes \mathbf{C}_{m+1}^p\right] \frac{d\text{vec}\,\mathbf{H}}{d\boldsymbol{\theta}^{\mathsf{T}}} \\
& + \left[\mathbf{I}_{s_1} \otimes \mathbf{C}_{m+1}^p\,(\mathbf{I} - \mathbf{H})\right] \sum_{i=1}^{m} \left[\left(\mathbf{C}_1^{i-1}\right)^{\mathsf{T}} \otimes \mathbf{C}_{i+1}^m\right] \frac{d\text{vec}\,\mathbf{B}_i}{d\boldsymbol{\theta}^{\mathsf{T}}} \\
& + \left[\left((\mathbf{I} - \mathbf{H})\,\mathbf{C}_1^m\right)^{\mathsf{T}} \otimes \mathbf{I}_{s_1}\right] \sum_{i=m+1}^{p} \left[\left(\mathbf{C}_{m+1}^{i-1}\right)^{\mathsf{T}} \otimes \mathbf{C}_{i+1}^p\right] \frac{d\text{vec}\,\mathbf{B}_i}{d\boldsymbol{\theta}^{\mathsf{T}}}
\end{aligned}
$$

$$\tag{8.22}$$

The expression (8.17) can be substituted for $d\text{vec}\,\mathbf{H}$ in (8.22), and the resulting expression for $d\text{vec}\,\mathbf{A}/d\boldsymbol{\theta}^{\mathsf{T}}$ substituted into (8.5).

8.3 Multistate Models

We have encountered several examples of models in which individuals are classified by two criteria (age and stage, stage and environmental state, stage and location, etc.). These multistate models can be constructed by the vec-permutation matrix approach; see Chaps. 5 and 6 or Hunter and Caswell (2005b) and Caswell et al. (2018).

Suppose individuals classified by two criteria; e.g., stages $(1, \ldots, s)$ and locations $(1, \ldots, r)$. One might describe population dynamics in terms of stage transitions within locations, and spatial movement within stages, with the two processes acting sequentially. Thus individuals first survive and reproduce according to their stage-specific demography, and then disperse among locations, and then repeat. Let $\mathbf{B}_i$ be the $s \times s$ matrix describing transitions and reproduction within location i, and $\mathbf{M}_j$ the $r \times r$ matrix describing movement probabilities for stage j. Let $\mathbb{B}$ and $\mathbb{M}$ be the $sr \times sr$ block diagonal matrices with the $\mathbf{B}_i$ and the $\mathbf{M}_j$, respectively, on the diagonal.

The population is projected by

$$\mathbf{n}(t + 1) = \mathbf{K}^{\mathsf{T}}\mathbb{M}\mathbf{K}\mathbb{B}\ \mathbf{n}(t).. \tag{8.23}$$

The matrix $\mathbf{K}$ is the vec-permutation matrix, or commutation matrix, (Henderson and Searle 1981; Magnus and Neudecker 1979), which satisfies

$$\operatorname{vec} \mathcal{N}^{\mathsf{T}} = \mathbf{K} \operatorname{vec} \mathcal{N} \tag{8.24}$$

For the calculation of $\mathbf{K}$, see Sect. 2.2.3.

The model (8.23) is formally periodic, with the operation of $\mathbb{B}$ and $\mathbb{M}$ alternating; thus the projection matrix is

$$\mathbf{A} = \mathbf{K}^{\mathsf{T}} \mathbb{M} \mathbf{K} \mathbb{B}. \tag{8.25}$$

The dependence of $\mathbf{A}$ on the parameters $\boldsymbol{\theta}$ can take place through $\mathbf{B}_i\,[\boldsymbol{\theta}]$, $\mathbf{M}_i\,[\boldsymbol{\theta}]$, or both.

The general sensitivity formula (8.5) requires the derivative $d\mathbf{A}/d\boldsymbol{\theta}^{\mathsf{T}}$. Differentiating (8.25) gives

$$d\mathbf{A} = \mathbf{K}^{\mathsf{T}}\,(d\mathbb{M})\,\mathbf{K}\mathbb{B} + \mathbf{K}^{\mathsf{T}}\mathbb{M}\mathbf{K}\,(d\mathbb{B})\,. \tag{8.26}$$

Applying the vec operator gives

$$d\operatorname{vec} \mathbf{A} = \left(\mathbb{B}^{\mathsf{T}}\mathbf{K}^{\mathsf{T}} \otimes \mathbf{K}^{\mathsf{T}}\right) d\operatorname{vec} \mathbb{M} + \left(\mathbf{I}_{sp} \otimes \mathbf{K}^{\mathsf{T}}\mathbb{M}\mathbf{K}\right) d\operatorname{vec} \mathbb{B} \tag{8.27}$$

We want to express $d\operatorname{vec} \mathbb{B}$ and $d\mathbb{M}$ in terms of the derivatives of their diagonal entries $\mathbf{B}_i$ and $\mathbf{M}_j$. This can be done using equations (14) and (15) of Caswell and van Daalen (2016). Define the matrices $\mathbf{P}_i$ and $\mathbf{Q}_i$, of dimension $rs \times s$ and $s \times rs$, respectively,

$$\mathbf{P}_i = \begin{pmatrix} \mathbf{0}_{s(i-1)\times s} \\ \mathbf{I}_s \\ \mathbf{0}_{s(r-i)\times s} \end{pmatrix} \qquad \mathbf{Q}_i = \begin{pmatrix} \mathbf{0}_{s\times(i-1)s} \ \mathbf{I}_s \ \mathbf{0}_{s\times(r-i)s} \end{pmatrix}. \tag{8.28}$$

Then

$$d\operatorname{vec} \mathbb{B} = \sum_{i=1}^{r} \left(\mathbf{Q}_i^{\mathsf{T}} \otimes \mathbf{P}_i\right) d\operatorname{vec} \mathbf{B}_i. \tag{8.29}$$

Similarly, for $\mathbb{M}$, define matrices $\mathbf{R}_i$ and $\mathbf{S}_i$

$$\mathbf{R}_i = \begin{pmatrix} \mathbf{0}_{r(i-1)\times r} \\ \mathbf{I}_r \\ \mathbf{0}_{r(s-i)\times r} \end{pmatrix} \qquad \mathbf{S}_i = \begin{pmatrix} \mathbf{0}_{r\times(i-1)r} \ \mathbf{I}_r \ \mathbf{0}_{r\times(s-i)r} \end{pmatrix}. \tag{8.30}$$

Then

$$d\text{vec}\,\mathbb{M} = \sum_{i=1}^{s} \left(\mathbf{S}_i^\mathsf{T} \otimes \mathbf{R}_i\right) \text{vec}\,\mathbf{M}_i. \tag{8.31}$$

Substituting (8.29) and (8.31) into the expression (8.27) for $d\text{vec}\,\mathbf{A}$ gives the final result

$$\frac{d\text{vec}\,\mathbf{A}}{d\boldsymbol{\theta}^\mathsf{T}} = \underbrace{\mathbf{X}_1 \sum_{j=1}^{s} \left(\mathbf{S}_j^\mathsf{T} \otimes \mathbf{R}_j\right) \frac{d\text{vec}\,\mathbf{M}_j}{d\boldsymbol{\theta}^\mathsf{T}}}_{\text{perturbations of the } \mathbf{M}_j} + \underbrace{\mathbf{X}_2 \sum_{i=1}^{r} \left(\mathbf{Q}_i^\mathsf{T} \otimes P_i\right) \frac{d\text{vec}\,\mathbf{B}_i}{d\boldsymbol{\theta}^\mathsf{T}}}_{\text{perturbations of the } \mathbf{B}_i} \tag{8.32}$$

where $\mathbf{X}_1$ and $\mathbf{X}_2$ are constant matrices,

$$\mathbf{X}_1 = \left(\mathbb{B}^\mathsf{T}\mathbf{K}^\mathsf{T} \otimes \mathbf{K}^\mathsf{T}\right) \tag{8.33}$$

$$\mathbf{X}_2 = \left(\mathbf{I}_{sr} \otimes \mathbf{K}^\mathsf{T}\mathbb{M}\mathbf{K}\right) \tag{8.34}$$

that need be calculated only once. Although $\mathbf{X}_1$, $\mathbf{X}_2$, and the Kronecker products appearing in the summations are large, they are also extremely sparse. The sparse matrix capabilities in MATLAB can take advantage of this fact. Substituting (8.32) into (8.5) gives the sensitivity of an output variable $\boldsymbol{\xi}$ to changes in parameters that perturb any or all of the $\mathbf{M}_j$ and $\mathbf{B}_i$.

8.4   Nonlinear Models and Delayed Density Dependence

Anticipating the more extensive treatment in Chap. 10, we consider the effects of nonlinearity in periodic models. You may want to return to this section after Sect. 10.7, which analyzes periodic oscillations arising from time-invariant non-linearities. When periodic environmental changes interact with such oscillations, the results can be complicated, and such interactions are the focus of the present section.

In a periodic nonlinear model, each of the $\mathbf{B}_i$ in (8.2) may depend on density. Especially in seasonal models, the vital rates in the matrix $\mathbf{B}_i$ may depend on densities not only at phase i, but at previous phases within the cycle as well. For example, in a study of the invasive plant garlic mustard (*Alliaria petiolata*) Shyu et al. (2013) found that seed production of fruiting plants in the fall reflected the density experienced by vegetative rosettes in the early spring.

To develop a model including such delayed density dependence, define

$$\mathbf{n}_i(t) = \text{population at season } i \text{ in year } t \tag{8.35}$$

Starting at season 1, the dynamics are given by

$$\mathbf{n}_1(t+1) = \mathbf{B}_p \mathbf{n}_p(t)$$
$$\mathbf{n}_2(t) = \mathbf{B}_1 \mathbf{n}_1(t)$$
$$\vdots \qquad\qquad\qquad (8.36)$$
$$\mathbf{n}_p(t) = \mathbf{B}_{p-1} \mathbf{n}_{p-1}(t)$$

Density-dependence, in a general form, means that the matrices $\mathbf{B}_i$ may be functions of densities over one cycle prior to season i:

$$\mathbf{B}_1 = \mathbf{B}_1 \left[\mathbf{n}_1(t), \mathbf{n}_p(t-1), \ldots, \mathbf{n}_2(t-1) \right]$$

$$\mathbf{B}_2 = \mathbf{B}_2 \left[\mathbf{n}_2(t), \mathbf{n}_1(t), \mathbf{n}_p(t-1), \ldots, \mathbf{n}_3(t-1) \right]$$

$$\vdots \qquad\qquad\qquad (8.37)$$

$$\mathbf{B}_p = \mathbf{B}_p \left[\mathbf{n}_p(t), \mathbf{n}_{p-1}(t), \ldots, \mathbf{n}_1(t) \right]$$

A fixed point on the interannual time scale, from t to $t+1$, is a p-cycle on the seasonal scale, satisfying

$$\hat{\mathbf{n}}_1 = \mathbf{B}_p \left[\hat{\mathbf{n}}_1, \ldots, \hat{\mathbf{n}}_p \right] \hat{\mathbf{n}}_p$$
$$\hat{\mathbf{n}}_2 = \mathbf{B}_1 \left[\hat{\mathbf{n}}_1, \ldots, \hat{\mathbf{n}}_p \right] \hat{\mathbf{n}}_1$$
$$\vdots \qquad\qquad\qquad (8.38)$$
$$\hat{\mathbf{n}}_p = \mathbf{B}_{p-1} \left[\hat{\mathbf{n}}_1, \ldots, \hat{\mathbf{n}}_p \right] \hat{\mathbf{n}}_{p-1}$$

A k-cycle on the interannual time scale is a kp-cycle on the seasonal time scale, the points of which are numbered $\hat{\mathbf{n}}_1, \ldots, \hat{\mathbf{n}}_{kp}$. The corresponding sequence of matrices, in which the annual cycle $\mathbf{B}_1, \ldots, \mathbf{B}_p$ is repeated k times, is defined as $\mathbf{B}_1, \ldots, \mathbf{B}_{kp}$. With this notation, (8.38) still holds, with kp instead of p entries.

Differentiating (8.38) yields

$$d\hat{\mathbf{n}}_i = (d\mathbf{B}_{i-1}) \hat{\mathbf{n}}_{i-1} + \mathbf{B}_{i-1} \left(d\hat{\mathbf{n}}_{i-1} \right) \qquad i = 1, \ldots, p \qquad (8.39)$$

where the subscripts on $\hat{\mathbf{n}}$ and $\mathbf{B}$ are interpreted modulo p. Applying the vec operator to (8.39) yields

$$d\hat{\mathbf{n}}_i = \left(\hat{\mathbf{n}}_{i-1}^{\mathsf{T}} \otimes \mathbf{I}_s \right) d\text{vec}\, \mathbf{B}_{i-1} + \mathbf{B}_{i-1} d\hat{\mathbf{n}}_{i-1}. \qquad (8.40)$$

The sensitivity analysis of the cycle involves a set of block-structured matrices, the form of which is easily generalized from the special case with $p = 3$. Assuming $p = 3$ and noting that $\mathbf{B}$ depends on all the $\hat{\mathbf{n}}_i$ as well as on the parameter vector $\boldsymbol{\theta}$, the total differential of $\mathbf{B}_{i-1}$ in (8.40) is

$$d\operatorname{vec}\mathbf{B}_{i-1} = \frac{\partial\operatorname{vec}\mathbf{B}_{i-1}}{\partial\mathbf{n}_1^\mathsf{T}}d\hat{\mathbf{n}}_1 + \frac{\partial\operatorname{vec}\mathbf{B}_{i-1}}{\partial\mathbf{n}_2^\mathsf{T}}d\hat{\mathbf{n}}_2 + \frac{\partial\operatorname{vec}\mathbf{B}_{i-1}}{\partial\mathbf{n}_3^\mathsf{T}}d\hat{\mathbf{n}}_3 + \frac{\partial\operatorname{vec}\mathbf{B}_{i-1}}{\partial\boldsymbol{\theta}^\mathsf{T}}d\boldsymbol{\theta}$$
(8.41)

For notational convenience, define the matrices

$$\mathbf{H}_i = \left(\hat{\mathbf{n}}_i^\mathsf{T} \otimes \mathbf{I}_s\right) \qquad i = 1, \ldots, p$$
(8.42)

Substituting (8.41) into (8.40) produces the set of equations

$$d\hat{\mathbf{n}}_1 = \mathbf{H}_3\frac{\partial\operatorname{vec}\mathbf{B}_3}{\partial\boldsymbol{\theta}^\mathsf{T}}d\boldsymbol{\theta} + \mathbf{H}_3\sum_{j=1}^{p}\frac{\partial\operatorname{vec}\mathbf{B}_3}{\partial\mathbf{n}_j^\mathsf{T}}d\hat{\mathbf{n}}_j + \mathbf{B}_3 d\hat{\mathbf{n}}_3$$

$$d\hat{\mathbf{n}}_2 = \mathbf{H}_1\frac{\partial\operatorname{vec}\mathbf{B}_1}{\partial\boldsymbol{\theta}^\mathsf{T}}d\boldsymbol{\theta} + \mathbf{H}_1\sum_{j=1}^{p}\frac{\partial\operatorname{vec}\mathbf{B}_1}{\partial\mathbf{n}_j^\mathsf{T}}d\hat{\mathbf{n}}_j + \mathbf{B}_1 d\hat{\mathbf{n}}_1 \qquad (8.43)$$

$$d\hat{\mathbf{n}}_1 = \mathbf{H}_2\frac{\partial\operatorname{vec}\mathbf{B}_2}{\partial\boldsymbol{\theta}^\mathsf{T}}d\boldsymbol{\theta} + \mathbf{H}_2\sum_{j=1}^{p}\frac{\partial\operatorname{vec}\mathbf{B}_2}{\partial\mathbf{n}_j^\mathsf{T}}d\hat{\mathbf{n}}_j + \mathbf{B}_2 d\hat{\mathbf{n}}_2$$

This set of equations can be reduced to a single equation by collecting all the points on the kp-cycle into a single vector. Write an array (of dimension $sp \times k$)

$$\mathcal{N} = \begin{array}{c|ccc} & \text{yr. 1} & & \text{yr. } k \\ \hline \text{season 1} & \hat{\mathbf{n}}_1 & \cdots & \hat{\mathbf{n}}_1 \\ \vdots & \vdots & & \vdots \\ \text{season } p & \hat{\mathbf{n}}_p & \cdots & \hat{\mathbf{n}}_p \end{array}$$
(8.44)

Then write the vector (of dimension $spk \times 1$)

$$\mathbb{N} = \operatorname{vec}\mathcal{N}$$
(8.45)

In terms of this vector, the set of equations (8.43) can be rewritten

$$\frac{d\mathbb{N}}{d\boldsymbol{\theta}^\mathsf{T}} = \left[\mathbf{I}_{skp} - \mathbb{B} - \mathbb{H}\mathbb{C}\right]^{-1}\mathbb{H}\mathbb{D}.$$
(8.46)

where $\mathbb{H}$ and $\mathbb{B}$ are the block-circulant matrices

$$\mathbb{H} = \begin{pmatrix} 0 & 0 & \mathbf{H}_3 \\ \mathbf{H}_1 & 0 & 0 \\ 0 & \mathbf{H}_2 & 0 \end{pmatrix} \tag{8.47}$$

$$\mathbb{B} = \begin{pmatrix} 0 & 0 & \mathbf{B}_3 \\ \mathbf{B}_1 & 0 & 0 \\ 0 & \mathbf{B}_2 & 0 \end{pmatrix}, \tag{8.48}$$

and $\mathbb{C}$ and $\mathbb{D}$ are the block matrices

$$\mathbb{C} = \begin{pmatrix} \dfrac{\partial \operatorname{vec} \mathbf{B}_1}{\partial \mathbf{n}_1^{\mathsf{T}}} & \cdots & \dfrac{\partial \operatorname{vec} \mathbf{B}_1}{\partial \mathbf{n}_3^{\mathsf{T}}} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial \operatorname{vec} \mathbf{B}_3}{\partial \mathbf{n}_1^{\mathsf{T}}} & \cdots & \dfrac{\partial \operatorname{vec} \mathbf{B}_3}{\partial \mathbf{n}_3^{\mathsf{T}}} \end{pmatrix} \tag{8.49}$$

$$\mathbb{D} = \begin{pmatrix} \dfrac{\partial \operatorname{vec} \mathbf{B}_1}{\partial \boldsymbol{\theta}^{\mathsf{T}}} \\ \dfrac{\partial \operatorname{vec} \mathbf{B}_2}{\partial \boldsymbol{\theta}^{\mathsf{T}}} \\ \dfrac{\partial \operatorname{vec} \mathbf{B}_3}{\partial \boldsymbol{\theta}^{\mathsf{T}}} \end{pmatrix}. \tag{8.50}$$

All the derivatives are evaluated at $\hat{\mathbf{n}}_1, \ldots, \hat{\mathbf{n}}_3$.

8.4.1 Averages

The vector $d\mathbb{N}/d\boldsymbol{\theta}^{\mathsf{T}}$ created by (8.46) contains the sensitivities of all s stages, at each of p seasons within the year, for each of the k years within the inter-annual k-cycle. If this is too much information, one can calculate the sensitivity of averages, or other linear combinations, taken in various ways.

To write these averages, let $\mathbf{b}_m$ be a $m \times 1$ vector of weights. For a simple average of m quantities, each entry of $\mathbf{b}_m$ is $1/m$; for a weighted average, the entries of $\mathbf{b}_m$ would be non-negative numbers summing to 1. More generally, $\mathbf{b}$ may contain arbitrary weights, such as biomass, metabolic rate, economic value, etc. See Chaps. 7 and 10. To calculate averages from $\mathbb{N}$, first apply these vectors to average over rows or columns of $\mathcal{N}$ and then apply the vec operator to express the results as averages over $\mathbb{N}$.

Annual fixed point If the dynamics are a fixed point on the annual time scale, averages can be calculated over stages (using a vector $\mathbf{b}_s$), over seasons (using a vector $\mathbf{b}_p$), or both. The $p \times 1$ vector of averages over stages is

$$\text{avg. over stages} = \text{vec}\left(\mathbf{b}_s^{\mathsf{T}}\mathcal{N}\right) \tag{8.51}$$

$$= \left(\mathbf{I}_p \otimes \mathbf{b}_s^{\mathsf{T}}\right)\mathbb{N} \tag{8.52}$$

The $s \times 1$ vector of averages over seasons is

$$\text{avg. over seasons} = \left(\mathbf{b}_p^{\mathsf{T}} \otimes \mathbf{I}_s\right)\mathbb{N} \tag{8.53}$$

The average over both seasons and stages (a scalar) is

$$\text{avg. over stages and seasons} = \left(\mathbf{b}_p^{\mathsf{T}} \otimes \mathbf{b}_s^{\mathsf{T}}\right)\mathbb{N} \tag{8.54}$$

Because the average is a linear operator, the sensitivities of these averages are obtained by applying the same weights to the derivative $d\mathbb{N}/d\boldsymbol{\theta}^{\mathsf{T}}$ in (8.46):

$$\text{sensitivity of avg. over stages} = \left(\mathbf{I}_p \otimes \mathbf{b}_s^{\mathsf{T}}\right)\frac{d\mathbb{N}}{d\boldsymbol{\theta}^{\mathsf{T}}} \tag{8.55}$$

$$\text{sensitivity of avg. over seasons} = \left(\mathbf{b}_p^{\mathsf{T}} \otimes \mathbf{I}_s\right)\frac{d\mathbb{N}}{d\boldsymbol{\theta}^{\mathsf{T}}} \tag{8.56}$$

$$\text{sensitivity of avg. over both} = \left(\mathbf{b}_p^{\mathsf{T}} \otimes \mathbf{b}_s^{\mathsf{T}}\right)\frac{d\mathbb{N}}{d\boldsymbol{\theta}^{\mathsf{T}}} \tag{8.57}$$

Annual k-cycle When the dynamics produce a k-cycle on the annual time scale, averages can be calculated over any desired combination of stages, seasons, and years. Table 8.2 gives the resulting expressions for the averages. As in the case of equations (8.55) and (8.56), the sensitivities of these averages to parameters are obtained by applying the same weights to $d\mathbb{N}/d\boldsymbol{\theta}^{\mathsf{T}}$.

8.4.2   A Nonlinear Example

As an example of the calculations for nonlinear systems, imagine an organism with two stages: immature juveniles and reproducing adults. Suppose that the year contains two seasons: a benign, reproduction-heavy Season 1 and a harsh, mortality-heavy Season 2. The life cycle graph is shown in Fig. 8.2. Adults in Season 1 all survive to Season 2 and give birth to new juveniles with per-capita fertility f, which depends on adult density in Season 1 according to $f[\mathbf{n}_1] = ae^{-bn_2}$, where a and b are the maximum fertility and the strength of density-dependence, respectively, and

Table 8.2 Calculation of averages of attractors of nonlinear periodic matrix population models. The upper half of the table shows averages over stages and over seasons when the dynamics are a fixed point on the inter-annual time scale, and thus a p-cycle on the seasonal time scale. The lower half of the table shows averages over all combinations of stages, seasons, and years, when the dynamics are a k-cycle on the inter-annual time scale, and thus a kp-cycle on the seasonal time scale

Average over	Formula	Vectors	Dimension
Stages	$\left(\mathbf{I}_p \otimes \mathbf{b}_s^\mathsf{T}\right)\mathbb{N}$	1	$p \times 1$
Seasons	$\left(\mathbf{b}_p^\mathsf{T} \otimes \mathbf{I}_s\right)\mathbb{N}$	1	$s \times 1$
Seasons and stages	$\left(\mathbf{b}_p^\mathsf{T} \otimes \mathbf{b}_s\right)\mathbb{N}$	1	1×1
Stages	$\left(\mathbf{I}_k \otimes \mathbf{I}_p \otimes \mathbf{b}_s^\mathsf{T}\right)\mathbb{N}$	1	$kp \times 1$
Seasons	$\left(\mathbf{I}_k \otimes \mathbf{b}_p^\mathsf{T} \otimes \mathbf{I}_s\right)\mathbb{N}$	k	$s \times 1$
Years	$\left(\mathbf{b}_k \otimes \mathbf{I}_p \otimes \mathbf{I}_s\right)\mathbb{N}$	p	$s \times 1$
Seasons and years	$\left(\mathbf{b}_k^\mathsf{T} \otimes \mathbf{b}_p^\mathsf{T} \otimes \mathbf{I}_s\right)\mathbb{N}$	1	$s \times 1$
Seasons and stages	$\left(\mathbf{I}_k \otimes \mathbf{b}_p^\mathsf{T} \otimes \mathbf{b}_s^\mathsf{T}\right)\mathbb{N}$	1	$k \times 1$
Years and stages	$\left(\mathbf{b}_k^\mathsf{T} \otimes \mathbf{I}_p \otimes \mathbf{b}_s^\mathsf{T}\right)\mathbb{N}$	1	$p \times 1$
Stages, seasons, years	$\left(\mathbf{b}_k^\mathsf{T} \otimes \mathbf{b}_p^\mathsf{T} \otimes \mathbf{b}_s^\mathsf{T}\right)\mathbb{N}$	1	1×1

Fig. 8.2 A periodic life cycle graph for a simple two-stage, two-season nonlinear model. J and A denote juveniles and adults, respectively
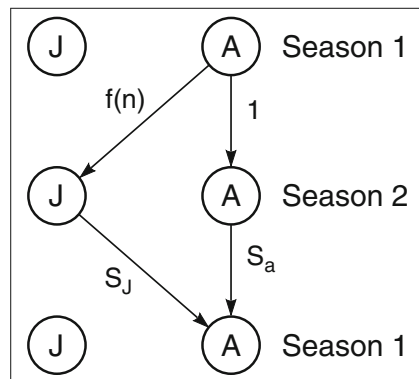
n_2 is the adult density in Season 1. In the harsher Season 2, juveniles and adults survive with probabilities s_j and s_a. A juvenile that survives to Season 1 matures into an adult.

This life cycle produces seasonal transition matrices $\mathbf{B}_1$ and $\mathbf{B}_2$:

$$\mathbf{B}_1[\mathbf{n}_1] = \begin{pmatrix} 0 & f[\mathbf{n}_1] \\ 0 & 1 \end{pmatrix} \tag{8.58}$$

$$\mathbf{B}_2 = \begin{pmatrix} 0 & 0 \\ s_j & s_a \end{pmatrix} \tag{8.59}$$
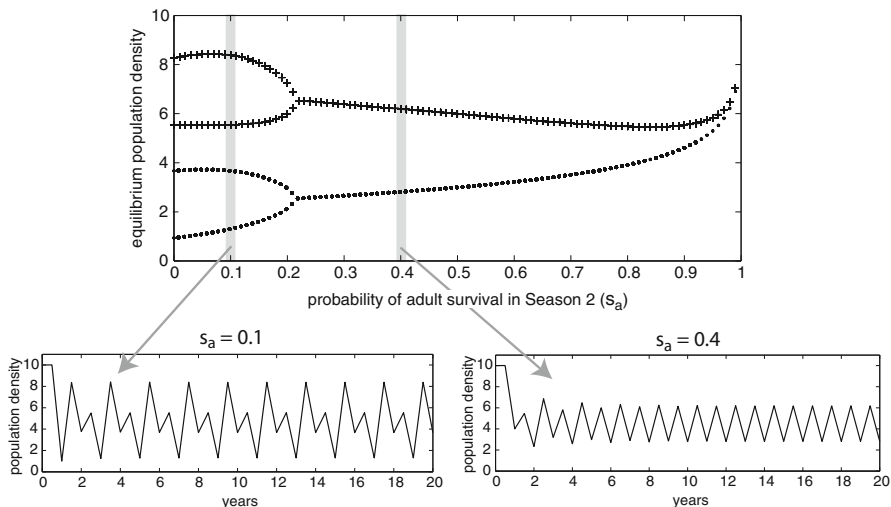
Fig. 8.3 A bifurcation diagram on the seasonal time scale for the two-season, two-stage model of Fig. 8.2. Total densities are plotted for Season 1 (•) and 2 (+). Parameters: $a = 20$, $b = 1$, $s_j = 0.5$; s_a varied from 0 to 1

and the nonlinear periodic model

$$\mathbf{n}_1(t + 1) = \mathbf{B}_2 \mathbf{n}_2(t)$$

$$\mathbf{n}_2(t) = \mathbf{B}_1 [\mathbf{n}_1(t)] \, \mathbf{n}_1(t) \tag{8.60}$$

Figure 8.3 is a bifurcation diagram for the system (8.60) in response to changes in adult survival s_a. When adults are long-lived ($s_a \gtrsim 0.22$) there is a 2-cycle on the seasonal scale, corresponding to a fixed point on the annual time scale, satisfying

$$\hat{\mathbf{n}}_1 = \mathbf{B}_2 \hat{\mathbf{n}}_2 \tag{8.61}$$

$$\hat{\mathbf{n}}_2 = \mathbf{B}_1[\hat{\mathbf{n}}_1]\hat{\mathbf{n}}_1 \tag{8.62}$$

At $s_a \approx 0.22$ this 2-cycle bifurcates to a 4-cycle on the seasonal time scale, corresponding to a 2-cycle on the annual scale.

To derive the block matrices $\mathbb{C}$ and $\mathbb{D}$ in Eqs. (8.49) and (8.50), define the parameter vector as

$$\boldsymbol{\theta} = \begin{pmatrix} s_j \; s_a \; a \; b \end{pmatrix}^{\top}. \tag{8.63}$$

The derivative matrices are

$$
\frac{d\,\text{vec}\,\mathbf{B}_1}{d\boldsymbol{\theta}^{\mathsf{T}}} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & e^{-b\hat{n}_2} & -a\hat{n}_2 e^{-b\hat{n}_2} \\ 0 & 0 & 0 & 0 \end{pmatrix} \tag{8.64}
$$

$$
\frac{d\,\text{vec}\,\mathbf{B}_2}{d\boldsymbol{\theta}^{\mathsf{T}}} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} \tag{8.65}
$$

$$
\frac{d\,\text{vec}\,\mathbf{B}_1}{d\mathbf{n}^{\mathsf{T}}} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & -abe^{-b\hat{n}_2} \\ 0 & 0 \end{pmatrix} \tag{8.66}
$$

$$
\frac{d\,\text{vec}\,\mathbf{B}_2}{d\mathbf{n}^{\mathsf{T}}} = \mathbf{0} \qquad (\text{dimension } 4 \times 2) \tag{8.67}
$$

We calculate the sensitivities of the equilibrium population at each phase of the cycle using Eq. (8.46) with $s_a = 0.4$ (a 2-cycle on the seasonal time scale; see Fig. 8.3) and with $s_a = 0.1$ (a 4-cycle on the seasonal time scale). The results, and the sensitivities of several averages, are shown in Fig. 8.4.

At the seasonal 2-cycle (annual fixed point), increases in s_j or s_a increase density in Season 1 and reduce density in Season 2, and have little effect on the
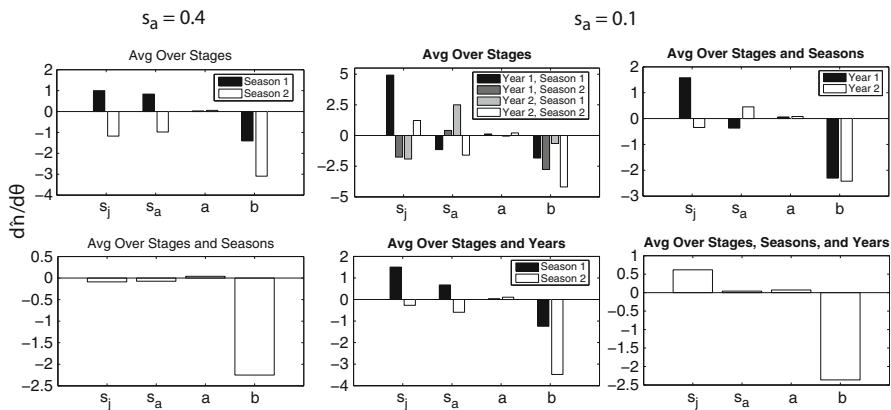
Fig. 8.4 Sensitivities of equilibrium total population size in Seasons 1 and 2, as well as the annual population average, to the demographic parameters s_j, s_a, a, and b. Left: sensitivities when $s_a = 0.4$ (seasonal 2-cycle, annual equilibrium). Right: sensitivities when $s_j = 0.1$ (seasonal 4-cycle, annual 2-cycle)

density averaged over seasons. The maximum fertility level a has little effect at either season, and the density-dependent parameter b has large negative effects throughout.

At the 4-point seasonal cycle (2-cycle on the annual time scale), the patterns are more complicated. We describe them in terms of the $kp = 4$ seasons in the cycle. The maximum fertility a has little effect at any point. The survival probabilities s_j and s_a have effects that are opposite in sign: an increase in s_j increases the density in seasons 1 and 4, and reduces it in seasons 2 and 3. An increase in s_a has the opposite effect. Averaged over years, both s_a and s_j increase density in season 1 and reduce it in season 2, thus increasing the amplitude of the oscillation. Averaged over seasons, s_a and s_j have opposite effects. When averaged over stages, seasons, and years, the effects of s_a cancel each other out, and only s_j and b have appreciable effects.

Even in this simple example, it is clear that parameter changes can have effects that differ among seasons and years. A set of MATLAB scripts to carry out these calculations appears in an online supplement to Caswell and Shyu (2012).

8.5 LTRE Decomposition Analysis

The LTRE decomposition analysis introduced in Sects. 2.9 and 4.5 can be extended to obtain the contributions, to any given outcome, of differences in parameters at each phase of the cycle.

Suppose that $\boldsymbol{\xi}$ is a $m \times 1$ dependent variable (scalar or vector-valued), a function of a parameter vector $\boldsymbol{\theta}$ that takes on values $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_p$ over the cycle. Use superscripts to denote two conditions,[3] which produce results $\boldsymbol{\xi}^{(1)}$ and $\boldsymbol{\xi}^{(2)}$:

$$\boldsymbol{\theta}_1^{(1)}, \ldots, \boldsymbol{\theta}_p^{(1)} \rightarrow \boldsymbol{\xi}^{(1)} \tag{8.68}$$

$$\boldsymbol{\theta}_1^{(2)}, \ldots, \boldsymbol{\theta}_p^{(2)} \rightarrow \boldsymbol{\xi}^{(2)} \tag{8.69}$$

To first order, the effect on $\boldsymbol{\xi}$ is

$$\boldsymbol{\xi}^{(2)} - \boldsymbol{\xi}^{(1)} \approx \sum_{k=1}^{p} \frac{d\boldsymbol{\xi}}{d\boldsymbol{\theta}_k^{\mathsf{T}}} \left(\boldsymbol{\theta}_k^{(2)} - \boldsymbol{\theta}_k^{(1)} \right) \tag{8.70}$$

The kth term in the summation in (8.70) is the total contribution, over all of the parameters in $\boldsymbol{\theta}$, of parameter differences in phase k of the cycle.

Define $\mathbf{R}_k$ as a $m \times p$ contribution matrix whose entries are the contributions of parameter θ_j in phase k to the effects on outcome variable ξ_i. Then

$$\mathbf{R}_k = \frac{d\boldsymbol{\xi}}{d\boldsymbol{\theta}_k^{\mathsf{T}}} \, \mathcal{D} \left(\boldsymbol{\theta}_k^{(2)} - \boldsymbol{\theta}_k^{(1)} \right) \tag{8.71}$$

---

[3]The extension to more than two conditions is easy; see Caswell (2001).

where the derivative is evaluated at the average of $\boldsymbol{\theta}^{(1)}$ and $\boldsymbol{\theta}^{(2)}$. These contributions are a decomposition of the approximate effect in (8.70),

$$\boldsymbol{\xi}^{(2)} - \boldsymbol{\xi}^{(1)} \approx \sum_{k=1}^{p} \mathbf{R}_k \mathbf{1}_p \qquad (8.72)$$

The contribution matrix (8.71) requires $d\boldsymbol{\xi}/d\boldsymbol{\theta}_k^{\mathsf{T}}$, that is, the derivative of $\boldsymbol{\xi}$ to the parameter at phase k of the cycle. In the linear model (8.2), this is given by the kth term in the summation in (8.11). In the case of the nonlinear model (8.36), the derivative is obtained from Eq. (8.46) by setting all blocks of $\mathbb{D}$, except those corresponding to phase k, to zero.

8.6   Discussion

The distinguishing feature of periodic models is that the dynamics over a projection interval are given by a periodic product of matrices. The periodic product may reflect the existence of multiple timescales (e.g., seasonal and annual), or the operation of multiple processes (e.g., demography and harvest), or express conditional probabilities, or arise from classifying individuals by multiple criteria. The sensitivity analysis of periodic models must account for the chain of causation (Fig. 8.1) from demographic parameters at each phase in the cycle to the corresponding projection matrices, and thence to the periodic matrix product over the whole cycle, and finally to demographic outcome $\boldsymbol{\xi}$. Matrix calculus makes this easy to do, starting with a simple chain rule expression (see Eq. (8.5)) and then using an appropriate version of (8.9) to calculate the derivative $d\mathrm{vec}\,\mathbf{A}/d\boldsymbol{\theta}^{\mathsf{T}}$.

Bibliography

Caswell, H. 2001. Matrix Population Models: Construction, Analysis, and Interpretation. 2nd edition. Sinauer Associates, Sunderland, MA.

Caswell, H. 2013. Sensitivity analysis of discrete Markov chains via matrix calculus. Linear Algebra and its Applications **438**:1727–1745.

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 E. Shyu. 2012. Sensitivity analysis of periodic matrix population models. Theoretical Population Biology **82**:329–339.

Caswell, H., and M. C. Trevisan. 1994. Sensitivity analysis of periodic matrix models. Ecology **75**:1299–1303.

Caswell, H., and S. F. van Daalen. 2016. A note on the vec operator applied to unbalanced block-structured matrices. Journal of Applied Mathematics **2016**:1–3.

Cushing, J. M. 2006. A juvenile-adult model with periodic vital rates. Journal of Mathematical Biology **53**:520–539.

Darwin, J. H., and R. M. Williams. 1964. The effect of time of hunting on the size of a rabbit population. New Zealand Journal of Science **7**:341–352.

Hauser, C. E., E. G. Cooch, and J.-D. Lebreton. 2006. Control of structured populations by harvest. Ecological Modelling **196**:462–470.

Henderson, H. V., and S. R. Searle. 1981. The vec-permutation matrix, the vec operator and Kronecker products: a review. Linear and Multilinear Algebra **9**:271–288.

Henson, S. M., and J. M. Cushing. 1997. The effect of periodic habitat fluctuations on a nonlinear insect population model. Journal of Mathematical Biology **36**:201–226.

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.

Hunter, C. M., and H. Caswell. 2005a. Selective harvest of sooty shearwater chicks: effects on population dynamics and sustainability. Journal of Animal Ecology **74**:589–600.

Hunter, C. M., and H. Caswell. 2005b. The use of the vec-permutation matrix in spatial matrix population models. Ecological Modelling **188**:15–21.

Jenouvrier, S., H. Caswell, C. Barbraud, and H. Weimerskirch. 2010. Mating behavior, population growth, and the operational sex ratio: a periodic two-sex model approach. The American Naturalist **175**:739–752.

Jenouvrier, S., M. Holland, J. Stroeve, M. Serreze, C. Barbraud, H. Weimerskirch, and H. Caswell. 2014. Projected continent-wide declines of the emperor penguin under climate change. Nature Climate Change **4**:715.

Lesnoff, M., P. Ezanno, and H. Caswell. 2003. Sensitivity analysis in periodic matrix models: a postscript to Caswell and Trevisan. Mathematical and Computer Modelling **37**:945–948.

MacArthur, R. H. 1968. Selectiondddsadsaafor life tables in periodic environments. The American Naturalist **102**:381–383.

Magnus, J. R., and H. Neudecker. 1979. The commutation matrix: some properties and applications. Annals of Statistics **7**:381–394.

Pico, F. X., H. de Kroon, and J. Retana. 2002. An extended flowering and fruiting season has few demographic effects in a Mediterranean perennial herb. Ecology **83**:1991–2004.

Shyu, E., E. A. Pardini, T. M. Knight, and H. Caswell. 2013. A seasonal, density-dependent model for the management of an invasive weed. Ecological Applications **23**:1893–1905.

Skellam, J. G. 1966. Seasonal periodicity in theoretical population ecology. Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability **4**:179–205.

Smith, M., H. Caswell, and P. Mettler-Cherry. 2005. Stochastic flood and precipitation regimes and the population dynamics of a threatened floodplain plant. Ecological Applications **15**:1036–1052.

van Daalen, S. F., and H. Caswell. 2017. Lifetime reproductive output: individual stochasticity, variance, and sensitivity analysis. Theoretical Ecology **10**:355–374.

Chapter 9
LTRE Decomposition of the Stochastic Growth Rate

9.1 Introduction

The basic unit of comparative demography is a study that reports the value of some demographic outcome in two populations that differ in a set of vital rates. One challenge of such studies is to account for the difference in outcomes by decomposing that difference into contributions from differences in each of the parameters. It frequently happens that small differences in some parameters make large contributions to the difference in outcomes, and vice-versa.

In some parts of the literature, such studies are called life table response experiment (or LTRE) analyses; versions of this analysis have appeared in Sect. 1.3.1 and Chaps. 2, 4, and 8. The term was introduced by in the context of laboratory studies of the population effects of pollutants, hence the use of the word "experiment" (Caswell 1989). The conditions among which the populations are compared will be called "treatments" here, but there is no restriction to experimental manipulations.

Similar decomposition analyses have been developed independently in ecology and human demography. For example, Pollard's (1988) study of life expectancy used methods very similar to LTRE analyses of the population growth rate. Horiuchi et al. (2008) developed a method for continuous variables that is essentially identical to that used by ecologists for regression LTRE calculations (Caswell 1996). Canudas Romo (2003) reviews the human demographic literature.

This chapter uses matrix calculus to extend LTRE analysis to stochastic models, by showing how to decompose differences in the stochastic growth rate, $\log \lambda_s$. Because stochastic models include both environmental fluctuations and the vital rate responses to those fluctuations, their structure is richer than that of time-

---

Chapter 9 is modified from: Caswell, H. 2010. Life table response experiment analysis of the stochastic growth rate. Journal of Ecology 98:324–333. ©Hal Caswell.

invariant models. Stochastic LTRE analysis thus requires a new approach to decomposing these differences. The payoffs, in terms of demographic and biological understanding, are great.

9.2   Decomposition with Derivatives

The familiar LTRE analysis uses derivatives to approximate the contributions of the vital rates to some (vector-valued) outcome $\boldsymbol{\xi}$ (dimension $q \times 1$), as described in Chap. 2. Suppose that $\boldsymbol{\xi}$ depends on a vector $\boldsymbol{\theta}$ of vital rates (dimension $p \times 1$), and that observations are available under two treatments, with

$$\boldsymbol{\theta}^{(1)} \longrightarrow \boldsymbol{\xi}^{(1)} \tag{9.1}$$

$$\boldsymbol{\theta}^{(2)} \longrightarrow \boldsymbol{\xi}^{(2)}. \tag{9.2}$$

Using matrix calculus notation, to first order,

$$\boldsymbol{\xi}^{(2)} - \boldsymbol{\xi}^{(1)} \approx \frac{d\boldsymbol{\xi}}{d\boldsymbol{\theta}^{\mathsf{T}}} \left(\boldsymbol{\theta}^{(2)} - \boldsymbol{\theta}^{(1)} \right). \tag{9.3}$$

where the derivative of $\boldsymbol{\xi}$ is evaluated at the mean of the two parameter vectors.

All the contributions to the difference $\boldsymbol{\xi}^{(2)} - \boldsymbol{\xi}^{(1)}$ are contained in a matrix $\mathbf{C}$ (dimension $q \times p$) given by

$$\mathbf{C} = \frac{d\boldsymbol{\xi}}{d\boldsymbol{\theta}^{\mathsf{T}}} \mathcal{D} \left(\boldsymbol{\theta}^{(2)} - \boldsymbol{\theta}^{(1)} \right) \tag{9.4}$$

where the derivative is evaluated at the mean of $\boldsymbol{\theta}^{(1)}$ and $\boldsymbol{\theta}^{(2)}$.

The entry $\mathbf{C}(i, j)$ of the contribution matrix is the contribution of the difference $\Delta\theta_j$ to the difference in $\Delta\xi_i$. The columns and rows of $\mathbf{C}$ give

$$\mathbf{C}(:, j) = \text{contribution of } \Delta\theta_j \text{ to } \Delta\boldsymbol{\xi} \tag{9.5}$$

$$\mathbf{C}(i, :) = \text{contribution of } \Delta\boldsymbol{\theta} \text{ to } \Delta\xi_i. \tag{9.6}$$

The sum over rows of $\mathbf{C}$ is the approximation (9.3) to the treatment effect on $\boldsymbol{\xi}$

$$\boldsymbol{\xi}^{(2)} - \boldsymbol{\xi}^{(1)} \approx \mathbf{C} \, \mathbf{1}_p. \tag{9.7}$$

The accuracy of this approximation gives a measure of the adequacy of the first-order assumption. Contributions can be small either because the treatment has little effect on θ_i or because ξ does not respond much to changes in θ_i.

The contribution matrix $\mathbf{C}$ takes advantage of matrix calculus to provide a simple calculation for decomposition of scalar-, vector- or matrix-valued differences.

Studies including more than two treatments or conditions are analyzed by defining a reference parameter vector $\boldsymbol{\theta}_r$ and calculating a matrix $\mathbf{C}_i$ for treatment i in terms of the parameter difference $\boldsymbol{\theta}_i - \boldsymbol{\theta}_r$. The reference treatment might be the average parameter set, or the parameters for a "control" condition, etc.

9.3 Kitagawa and Keyfitz: Decomposition Without Derivatives

In decomposing differences in the stochastic growth rate, we encounter variables for which the derivatives in (9.3) cannot be calculated. Fortunately, an alternative method for decomposition is available that does not rely on derivatives. It was introduced by Kitagawa (1955) to explore the effects of age-specific death rates and of age distribution on crude death rates. The method was later extended by Keyfitz to decompose differences in age distributions, dependency ratios, and population growth rates into contributions from the entire mortality and fertility schedules (Keyfitz 1968, Section 7.4; Keyfitz and Caswell 2005, Section 10.1). Canudas Romo (2003) summarizes more recent extensions of the approach in demography.

Suppose that ξ depends on two variables, with values (a, b) in Treatment 1 and (A, B) in Treatment 2. Thus

$$\xi^{(1)} = \xi[a, b] \tag{9.8}$$

$$\xi^{(2)} = \xi[A, B]. \tag{9.9}$$

To decompose the treatment effect $\xi[A, B] - \xi[a, b]$ into contributions from $A - a$ and $B - b$, the Kitagawa-Keyfitz method proceeds by exchanging variables between the two treatments and calculating ξ for all possible combinations. The effect of $A - a$, against the background of B, is $\xi[A, B] - \xi[a, B]$. The effect of $A - a$, against the background of b is $\xi[A, b] - \xi[a, b]$. The overall contribution of $A - a$ is obtained by averaging its effect against the two backgrounds:

$$C(A - a) = \frac{1}{2}\left(\xi[A, B] - \xi[a, B]\right)$$
$$+ \frac{1}{2}\left(\xi[A, b] - \xi[a, b]\right). \tag{9.10}$$

Similarly, the contribution of $B - b$ is

$$C(B - b) = \frac{1}{2}\left(\xi[A, B] - \xi[A, b]\right)$$
$$+ \frac{1}{2}\left(\xi[a, B] - \xi[a, b]\right). \tag{9.11}$$

If this appears familiar, it may be because this process of averaging differences across different backgrounds is precisely analogous to the calculation of main effects in a two-way ANOVA (e.g., Steel and Torrie 1960, Section 11.2).

9.4   Stochastic Population Growth

A stochastic model contains two components: a model for the dynamics of the environment and a model for the response of the vital rates to the environment (Cohen 1979; Tuljapurkar 1990; Caswell 2001). I focus here on the stochastic population growth rate, $\log \lambda_s$. Consider a population growing according to

$$\mathbf{n}(t + 1) = \mathbf{A}(t)\mathbf{n}(t) \qquad (9.12)$$

where the projection matrix $\mathbf{A}(t)$ is generated by a realization of an ergodic stochastic environment that produces, for every environmental state, a set of vital rates that satisfy certain regularity conditions. Then, the asymptotic long-term growth rate is, with probability one,

$$\log \lambda_s = \lim_{T \to \infty} \frac{1}{T} \log \left\| \mathbf{A}(T - 1) \cdots \mathbf{A}(0)\mathbf{n}_0 \right\| \qquad (9.13)$$

(Cohen 1976; Tuljapurkar and Orzack 1980; Tuljapurkar 1990). This growth rate plays a central role in demography and biodemography in stochastic environments, exactly analogous to the role played by the population growth rate λ or $r = \log \lambda$ in stable population theory in constant environments. Cohen (1986) and Lee and Tuljapurkar (1994) have incorporated models of the form (9.12), with the addition of immigration terms, into the context of human population projections, to provide estimates of confidence intervals more rigorous than the "high, medium, low" scenarios usually reported.

The additional component in stochastic environments adds an extra layer of complexity to the LTRE decomposition of the stochastic growth rate (Fig. 9.1). The differences in $\log \lambda_s$ between two treatments is partly due to differences in the environmental dynamics and partly to differences in the vital rates within each environmental state.

In this chapter, I consider the case in which the environment is described by a finite-state Markov chain. Ecological examples include years with our without fire (Silva et al. 1991), years since fire (Caswell and Kaye 2001), years with early or late floods, or with high or low precipitation (Smith et al. 2005) and years with good or poor sea ice conditions (Hunter et al. 2010; Jenouvrier et al. 2009b). The Markovian environment case also includes the situation where the environment is modelled implicitly by selecting randomly from a set of empirically-measured matrices (e.g., Bierzychudek 1982; Cohen et al. 1983; Jenouvrier et al. 2009a). Let $u(t)$ be the state of the environment at time t. The environmental dynamics are determined by the Markov chain transition matrix $\mathbf{P}$, where $p_{ij} = P\left[u(t + 1) = i \mid u(t) = j\right]$.

(a)

(b)
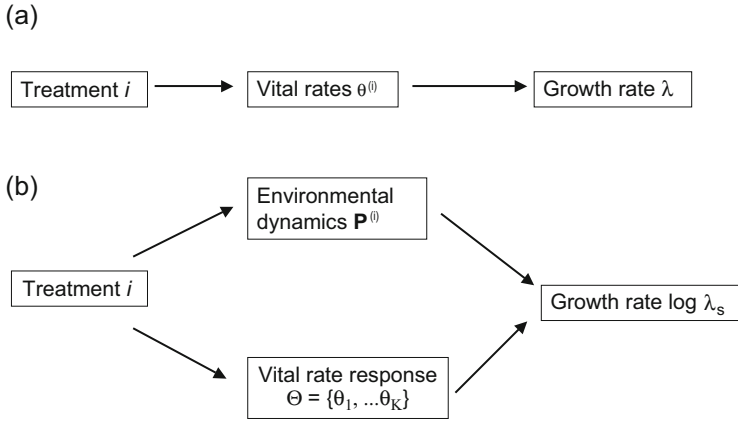
Fig. 9.1 The determination of population growth rate in (**a**) time-invariant and (**b**) stochastic models. The deterministic growth rate λ is defined by a set of vital rates, which are determined by the environment ("treatment"). The stochastic growth rate $\log \lambda_s$ requires an additional model for the stochastic dynamics of the environment and a function giving the response of the vital rates to the state of the environment

The second part of the model is the response of the vital rates to the environment. Let $\boldsymbol{\theta}$ be a vector of parameters that determine the projection matrix $\mathbf{A}$. The vectors $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_K$ correspond to environmental states $1, \ldots, K$. I will write the entire set of vital rates as

$$\boldsymbol{\Theta} = \{\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_K\}. \tag{9.14}$$

We write $\mathbf{A}(t) = \mathbf{A}[\boldsymbol{\theta}(t)]$, and the stochastic growth rate (9.13) becomes

$$\log \lambda_s [\mathbf{P}, \boldsymbol{\Theta}] = \lim_{T \to \infty} \frac{1}{T} \log \left\| \mathbf{A}[\boldsymbol{\theta}(T-1)] \cdots \mathbf{A}[\boldsymbol{\theta}(0)] \, \mathbf{n}_0 \right\| \tag{9.15}$$

where $\boldsymbol{\theta}(t)$ is the parameter vector created by the environmental state $u(t)$. I have written $\log \lambda_s$ as an explicit function of $\mathbf{P}$ and $\boldsymbol{\Theta}$ to emphasize that it depends on both the environment and the vital rate response.

9.4.1   Environment-Specific Sensitivities

The sensitivity of $\log \lambda_s$ to the vital rates was given by Tuljapurkar (1990). For the LTRE analysis, we require the derivatives of $\log \lambda_s$ with respect to the parameters in each state of the environment; i.e., to each of the vectors $\boldsymbol{\theta}_i$ in $\boldsymbol{\Theta}$. These environment-specific sensitivities were given by Caswell (2005) and independently by Horvitz et al. (2005), and have been applied by Gervais et al. (2006), Aberg

et al. (2009), and Svensson et al. (2009). Rewriting Tuljapurkar's (1990) formula in matrix calculus notation yields the derivative of $\log \lambda_s$ with respect to the vital rate vector in environment i:

$$\left. \frac{d \log \lambda_s}{d \boldsymbol{\theta}^{\mathsf{T}}} \right|_{u=i} = \lim_{T \to \infty} \frac{1}{T} \sum_{t=0}^{T-1} J_t \frac{\left[\mathbf{w}(t)^{\mathsf{T}} \otimes \mathbf{v}(t+1)^{\mathsf{T}} \right]}{R_t \mathbf{v}^{\mathsf{T}}(t+1) \mathbf{w}(t+1)} \frac{d \text{vec} \, \mathbf{A}[\boldsymbol{\theta}(t)]}{d \boldsymbol{\theta}^{\mathsf{T}}}. \tag{9.16}$$

This is the stochastic analogue of the expression (3.46) in Chap. 3, for the sensitivity of the deterministic growth rate. The vectors $\mathbf{w}(t)$ and $\mathbf{v}(t)$ are the stochastic analogues of the right and left eigenvectors of a deterministic model, and R_t is the growth of total population size from t to $t + 1$. See Caswell (2001, Section 14.4) for a step-by-step algorithm for the calculation.

To make sensitivity environment-dependent, J_t is an indicator variable, defined as

$$J_t = \begin{cases} 1 \text{ if } u(t) = i \\ 0 \text{ otherwise} \end{cases} \tag{9.17}$$

If the parameters $\boldsymbol{\theta}$ consist of the elements of $\mathbf{A}$, then $d \text{vec} \, \mathbf{A}/d \boldsymbol{\theta}^{\mathsf{T}} = \mathbf{I}$, where $\mathbf{I}$ is the identity matrix. If $\boldsymbol{\theta}$ contains lower-level parameters, then $d \text{vec} \, \mathbf{A}/d \boldsymbol{\theta}^{\mathsf{T}}$ contains the derivatives of $\mathbf{A}$ with respect to these parameters.

9.5 LTRE Decomposition Analysis for $\log \lambda_s$

Suppose now that we have two treatments, and want to decompose the difference,

$$\log \lambda_s^{(2)} - \log \lambda_s^{(1)} = \log \lambda_s \left[\mathbf{P}^{(2)}, \boldsymbol{\Theta}^{(2)} \right] - \log \lambda_s \left[\mathbf{P}^{(1)}, \boldsymbol{\Theta}^{(1)} \right] \tag{9.18}$$

into contributions. This difference compares growth in treatment 2 to growth in treatment 1. Treatment 1, the reference treatment, could be a control in a manipulative experiment, or some other specific condition of interest (as in the example to be considered below), or an average over treatments in a factorial experiment.

The treatment effect on $\log \lambda_s$ in (9.18) depends on both the differences in environmental dynamics (captured in the transition matrices $\mathbf{P}^{(1)}$ and $\mathbf{P}^{(2)}$) and the differences in the vital rate responses (captured in the parameter arrays $\boldsymbol{\Theta}^{(1)}$ and $\boldsymbol{\Theta}^{(2)}$). Because $\log \lambda_s$ is calculated numerically from (9.15) by simulation, it cannot be differentiated[1] with respect to $\mathbf{P}$, so we will use the Kitagawa-Keyfitz

---

[1] Well, not by me. But see Steinsaltz et al. (2011) for a rigorous development of the sensitivity analysis of stochastic growth rates that includes the effects of changes in the entries of $\mathbf{P}$.

decomposition for the environmental dynamics contribution, and environment-specific derivatives (9.16) for the vital rate response contributions

Let us consider three cases: the case where only the vital rate responses differ, the case where only the environmental dynamics differ, and finally the case where both differ.

9.5.1 Case 1: Vital Rates Differ, Environments Identical

Consider two treatments that affect the vital rate responses but not the environmental dynamics. For example, one might want to compare low and high fertility sites subjected to a common fire frequency. The transition matrix $\mathbf{P}$ is identical in the two sites, but the vital rates differ. The stochastic growth rates are

$$\log \lambda_s^{(1)} = \log \lambda_s \left[\mathbf{P}, \mathbf{\Theta}^{(1)} \right] \qquad (9.19)$$

$$\log \lambda_s^{(2)} = \log \lambda_s \left[\mathbf{P}, \mathbf{\Theta}^{(2)} \right]. \qquad (9.20)$$

The difference in $\log \lambda_s$ is composed of contributions from vital rate differences in each state of the environment. To first order,

$$\log \lambda_s^{(2)} - \log \lambda_s^{(1)} \approx \sum_{i=1}^{K} \left(\left. \frac{\partial \log \lambda_s}{\partial \boldsymbol{\theta}^{\mathsf{T}}} \right|_{u=i} \right) \left(\boldsymbol{\theta}_i^{(2)} - \boldsymbol{\theta}_i^{(1)} \right) \qquad (9.21)$$

where the derivatives are environment-specific sensitivities (9.16), and are evaluated at the mean of $\mathbf{\Theta}^{(1)}$ and $\mathbf{\Theta}^{(2)}$. The ith term of the summation in (9.21) is the contribution of differences in the ith environment. These can be written as the elements of a contribution matrix (dimension $1 \times p$)

$$\mathbf{C}(\boldsymbol{\theta}_i) = \left. \frac{\partial \log \lambda_s}{\partial \boldsymbol{\theta}^{\mathsf{T}}} \right|_{u=i} \mathcal{D} \left(\boldsymbol{\theta}_i^{(2)} - \boldsymbol{\theta}_i^{(1)} \right) \quad i = 1, \ldots, K. \qquad (9.22)$$

9.5.2 Case 2: Vital Rates Identical, Environments Differ

Now consider two treatments that affect the environmental dynamics (given by $\mathbf{P}^{(1)}$ and $\mathbf{P}^{(2)}$) but not the vital rate responses. For example, a comparison of population growth before and after implementing a fire control strategy that changes the frequency of fire, but has no effect on how the vital rates respond to fire. The stochastic growth rates are

$$\log \lambda_s^{(1)} = \log \lambda_s \left[\mathbf{P}^{(1)}, \mathbf{\Theta} \right] \qquad (9.23)$$

$$\log \lambda_s^{(2)} = \log \lambda_s \left[\mathbf{P}^{(2)}, \mathbf{\Theta} \right]. \qquad (9.24)$$

The matrices $\mathbf{P}^{(1)}$ and $\mathbf{P}^{(2)}$ may differ in their long-term frequencies of environmental states. Those long-term frequencies are given by the stationary distributions, i.e., the right eigenvector $\boldsymbol{\pi}$ corresponding to the dominant eigenvalue of $\mathbf{P}$ (which always equals 1), scaled so that $\boldsymbol{\pi}$ sums to 1. The same frequency of environmental states, however, can be obtained from processes with different autocorrelation patterns, from negative autocorrelation (where states tend to alternate) to positive autocorrelation (characterized by long runs of the same state; see Caswell and Kaye (2001, Fig. 2) for an example). So, $\mathbf{P}^{(1)}$ and $\mathbf{P}^{(2)}$ may differ in their stationary distributions, autocorrelation patterns, or both. To separate the contributions from these, using the Kitagawa-Keyfitz decomposition, we construct a Markov chain with the same stationary distribution $\boldsymbol{\pi}$ as $\mathbf{P}$, but in which successive environmental states are independent, and hence there is no autocorrelation. This chain has the transition matrix

$$\mathbf{Q} = \boldsymbol{\pi}\mathbf{1}^{\mathsf{T}} \tag{9.25}$$

where $\mathbf{1}$ is a vector of ones. Because the next state is independent of the previous state, and the same matrix is applied at each time, this process is called "independent and identically distributed," and abbreviated "iid."

The contribution to $\log \lambda_s^{(2)} - \log \lambda_s^{(1)}$ of differences in $\mathbf{P}$ is

$$C(\mathbf{P}) = \log \lambda_s \left[\mathbf{P}^{(2)}, \boldsymbol{\Theta} \right] - \log \lambda_s \left[\mathbf{P}^{(1)}, \boldsymbol{\Theta} \right]. \tag{9.26}$$

The contribution of the difference in the iid part of the environment is

$$C(\mathbf{Q}) = \log \lambda_s \left[\mathbf{Q}^{(2)}, \boldsymbol{\Theta} \right] - \log \lambda_s \left[\mathbf{Q}^{(1)}, \boldsymbol{\Theta} \right]. \tag{9.27}$$

The contribution of differences in environmental autocorrelation, denoted by $C(\mathbf{R})$, is obtained by subtraction;

$$C(\mathbf{R}) = C(\mathbf{P}) - C(\mathbf{Q}). \tag{9.28}$$

9.5.3   Case 3: Vital Rates and Environments Differ

Finally, consider two treatments that differ in both the environmental dynamics ($\mathbf{P}^{(1)}$ and $\mathbf{P}^{(2)}$) and the vital rate responses ($\boldsymbol{\Theta}^{(1)}$ and $\boldsymbol{\Theta}^{(2)}$). The stochastic growth rates are

$$\log \lambda_s^{(1)} = \log \lambda_s \left[\mathbf{P}^{(1)}, \boldsymbol{\Theta}^{(1)} \right] \tag{9.29}$$

$$\log \lambda_s^{(2)} = \log \lambda_s \left[\mathbf{P}^{(2)}, \boldsymbol{\Theta}^{(2)} \right]. \tag{9.30}$$

Our goal is to decompose $\log \lambda_s^{(2)} - \log \lambda_s^{(1)}$ into contributions from the differences in the stationary environmental frequencies ($C(\mathbf{Q})$), in the autocorrelation pattern ($C(\mathbf{R})$), and in the vital rates in each environmental state ($C(\boldsymbol{\theta}_1), \ldots, C(\boldsymbol{\theta}_K)$). The decomposition analysis proceeds in three steps.

1. Write the contributions of the environmental differences using the Kitagawa-Keyfitz method

$$
C(\mathbf{P}) = \frac{1}{2} \left(\log \lambda_s \left[\mathbf{P}^{(2)}, \boldsymbol{\Theta}^{(2)} \right] - \log \lambda_s \left[\mathbf{P}^{(1)}, \boldsymbol{\Theta}^{(2)} \right] \right.
$$

$$
\left. + \log \lambda_s \left[\mathbf{P}^{(2)}, \boldsymbol{\Theta}^{(1)} \right] - \log \lambda_s \left[\mathbf{P}^{(1)}, \boldsymbol{\Theta}^{(1)} \right] \right) \tag{9.31}
$$

$$
C(\mathbf{Q}) = \frac{1}{2} \left(\log \lambda_s \left[\mathbf{Q}^{(2)}, \boldsymbol{\Theta}^{(2)} \right] - \log \lambda_s \left[\mathbf{Q}^{(1)}, \boldsymbol{\Theta}^{(2)} \right] \right.
$$

$$
\left. + \log \lambda_s \left[\mathbf{Q}^{(2)}, \boldsymbol{\Theta}^{(1)} \right] - \log \lambda_s \left[\mathbf{Q}^{(1)}, \boldsymbol{\Theta}^{(1)} \right] \right) \tag{9.32}
$$

$$
C(\mathbf{R}) = C(\mathbf{P}) - C(\mathbf{Q}) \tag{9.33}
$$

Each of $C(\mathbf{P})$, $C(\mathbf{Q})$, and $C(\mathbf{R})$ is a scalar.

2. Write the contributions of the vital rate differences using the Kitagawa-Keyfitz method

$$
C(\boldsymbol{\Theta}) = \frac{1}{2} \left\{ \log \lambda_s \left[\mathbf{P}^{(2)}, \boldsymbol{\Theta}^{(2)} \right] - \log \lambda_s \left[\mathbf{P}^{(2)}, \boldsymbol{\Theta}^{(1)} \right] \right\}
$$

$$
+ \frac{1}{2} \left\{ \log \lambda_s \left[\mathbf{P}^{(1)}, \boldsymbol{\Theta}^{(2)} \right] - \log \lambda_s \left[\mathbf{P}^{(1)}, \boldsymbol{\Theta}^{(1)} \right] \right\} \tag{9.34}
$$

$C(\boldsymbol{\Theta})$ is a scalar, summing the effects of differences in all of the parameter responses at all states of the environment. It is decomposed further in the next step:

3. Use the environment-specific derivatives of $\log \lambda_s$ to decompose each term in (9.34) into contributions from the vital rates in each environment, using (9.22)

$$
\mathbf{C}(\boldsymbol{\theta}_i) = \frac{1}{2} \left(\left. \frac{\partial \log \lambda_s \left[\mathbf{P}^{(2)}, \bar{\boldsymbol{\Theta}} \right]}{\partial \boldsymbol{\theta}^\mathsf{T}} \right|_{u=i} \right) \mathcal{D} \left(\boldsymbol{\theta}_i^{(2)} - \boldsymbol{\theta}_i^{(1)} \right)
$$

$$
+ \frac{1}{2} \left(\left. \frac{\partial \log \lambda_s \left[\mathbf{P}^{(1)}, \bar{\boldsymbol{\Theta}} \right]}{\partial \boldsymbol{\theta}^\mathsf{T}} \right|_{u=i} \right) \mathcal{D} \left(\boldsymbol{\theta}_i^{(2)} - \boldsymbol{\theta}_i^{(1)} \right) \quad i = 1, \ldots, K
$$

$$
\tag{9.35}
$$

for $i = 1, \ldots, K$, with the derivatives evaluated at $\bar{\boldsymbol{\Theta}}$, the mean of the vital rates in the two treatments being compared. The matrix $\mathbf{C}(\boldsymbol{\theta}_i)$ is $(1 \times p)$ vector, whose entries give the contributions to the differences in $\log \lambda_s$ from each of the vital rates in environment i.

The total contribution of the parameter differences given in (9.34) is

$$C(\boldsymbol{\Theta}) = \sum_{i=1}^{K} \mathbf{C}\left(\boldsymbol{\theta}_i\right) \mathbf{1}_p. \tag{9.36}$$

These calculations are easily implemented by writing subroutines to calculate $\log \lambda_s$ and the environment-specific sensitivities given a transition matrix and a set of parameters. The accuracy of the approximations involved can be checked by comparing

$$\log \lambda_s^{(2)} - \log \lambda_s^{(2)} \overset{?}{\approx} C(\mathbf{Q}) + C(\mathbf{R}) + \sum_{i=1}^{K} \mathbf{C}\left(\boldsymbol{\theta}_i\right) \mathbf{1}_p. \tag{9.37}$$

9.6   An Example: Fire and an Endangered Plant

I know of no comparative studies of stochastic population growth that include differences in both the environmental dynamics and the vital rate responses, so here is an artificial example, based on a model for an endangered plant, *Lomatium bradshawii*, in a stochastic fire environment (Caswell and Kaye 2001). *L. bradshawii* (Apiaceae) is a polycarpic herbaceous perennial plant. It exists in only a few isolated populations in prairies of Oregon and Washington. These habitats were, until recent times, subject to natural and anthropogenic fires, to which *L. bradshawii* seems to have adapted. Fires increase plant size and seedling recruitment, but the effect fades within a few years. Populations in recently burned areas have higher growth rates and lower probabilities of extinction than unburned populations. For more information, see Pendergrass et al. (1999), Caswell and Kaye (2001), and Kaye et al. (2001).

A stochastic demographic model for *L. bradshawii* was developed by Caswell and Kaye (2001), based on data from an experimental burning study. Individuals were classified into six stages based on size and reproductive status: yearlings, small and large vegetative plants, and small, medium, and large reproductive plants. The environment was classified into four states defined by the time since the most recent fire: the year of a fire and 1, 2, and 3+ years post-fire, and vital rates were estimated in each of these environmental states. The matrices are given in Caswell and Kaye (2001).

Populations were studied in two sites: Fisher Butte (FB) and Rose Prairie (RP) in western Oregon. The two sites differed in quality for *L. bradshawii*, with RP

Table 9.1 The population		Fisher Butte	Rose Prairie
growth rate λ calculated from the environment-specific matrices $\mathbf{A}[\boldsymbol{\theta}_i]$ for *L. bradshawii*. (From Caswell and Kaye 2001)	Years post-fire	λ	λ
	0	1.020	1.155
	1	0.984	1.118
	2	0.662	0.483
	≥3	0.869	0.906

superior to FB. Population growth rates were generally higher at RP than at FP (Table 9.1), and the stochastic growth rate was higher in RP than FB at any fire frequency. The critical fire frequency required to maintain *L. bradshawii* populations was about 0.8–0.9 at FB, but only 0.4–0.5 at RP. The causes of the differences between the sites are not known (Pendergrass et al. 1999).

9.6.1 The Stochastic Fire Environment

The model for environmental dynamics is a two-state Markov chain for fires (each year is either fire or no fire). This generates a four-state Markov chain for the environmental states (0, 1, 2, and 3 or more years post-fire). Let f be the long-term frequency of fire, and ρ the temporal autocorrelation coefficient of the fire process (the magnitude of ρ determines the rate of decay of correlation as time increases, the sign of ρ determines whether the correlation is of one sign, or oscillates). In the two-state fire model, the probability of fire in year $t + 1$ if there was no fire in year t is $q = f(1 - \rho)$. The probability of a fire if there was a fire in year t is $p = q + \rho$ (see Caswell 2001, Section 14.1). The resulting transition matrix for the four environmental states is

$$\mathbf{P} = \begin{pmatrix} p & q & q & q \\ 1 - p & 0 & 0 & 0 \\ 0 & 1 - q & 0 & 0 \\ 0 & 0 & 1 - q & 1 - q \end{pmatrix}. \tag{9.38}$$

If $\rho < 0$, f must satisfy

$$\frac{-\rho}{1 - \rho} \le f \le \frac{1}{1 - \rho} \tag{9.39}$$

in order to keep probabilities bounded between 0 and 1. See Caswell and Kaye (2001). Note that even if the fire process is iid, so that $\rho = 0$, the environmental process given by (9.38) is not iid.

9.6.2 LTRE Analysis

There is no information on differences in fire dynamics at the two sites, so Caswell and Kaye (2001) studied the response of $\log \lambda_s$ to the frequency and autocorrelation of fires. Here, we use stochastic LTRE analysis to decompose the differences in $\log \lambda_s$ in three hypothetical scenarios of environmental differences. I will use the matrix entries as the vital rates $\boldsymbol{\theta}$, there being no natural lower-level parameterization in this model. MATLAB code for the calculations is available as an appendix to Caswell (2010).

The stochastic growth rate $\log \lambda_s$ increases with fire frequency for both species. The RP site has a growth advantage, with $\log \lambda_s^{(RP)} > \log \lambda_s^{(FB)}$ at all fire frequencies. The RP advantage, measured by $\log \lambda_s^{(RP)} - \log \lambda_s^{(RP)}$ increases from ≈ 0.02 when $f = 0$ to ≈ 0.13 when $f = 1$.

Differences in vital rates and environmental transitions (Case 3) Suppose that the two sites differ in both environmental dynamics and vital rate responses, with fire frequencies, autocorrelations, and resulting stochastic growth of

	FB	RP
f	0.5	0.7
ρ	-0.5	0.5
$\log \lambda_s$	-0.043	0.081

In this hypothetical scenario, the FB population tends to experience alternating years with and without fires; in RP, there is a tendency for long runs of years with and without fires. For additional scenarios, see Caswell (2010).

To decompose the treatment effect $\log \lambda_s^{(RP)} - \log \lambda_s^{(FB)}$, we construct the Markov chain transition matrices from (9.38), and calculate the stationary distributions $\boldsymbol{\pi}^{(RP)}$ and $\boldsymbol{\pi}^{(FB)}$ as eigenvectors of $\mathbf{P}$. For each site, we generate the iid transition matrix $\mathbf{Q}$ from (9.25), and compute the contributions $C(\mathbf{P})$ from (9.31), $C(\mathbf{Q})$ from (9.32), and $C(\mathbf{R})$ from (9.33). Then we compute the environment-specific sensitivities of $\log \lambda_s$ from (9.16), for both $\mathbf{P}^{(RP)}$ and $\mathbf{P}^{(FB)}$, and use these to calculate the contributions $C(\boldsymbol{\theta}_i)$ of the vital rates in each environmental state, using (9.35). Finally, we sum the $C(\boldsymbol{\theta}_i)$ to obtain the integrated effect of all vital rate differences in each environment.

Figure 9.2 shows these contributions. Most of the growth rate advantage of the RP site can be attributed to an RP advantage in $\mathbf{A}[\boldsymbol{\theta}_1]$ and $\mathbf{A}[\boldsymbol{\theta}_2]$ (the year of a fire and the year immediately following a fire). The difference in the long-term frequency of environmental states, and the differences in autocorrelation patterns, make relatively little contribution.
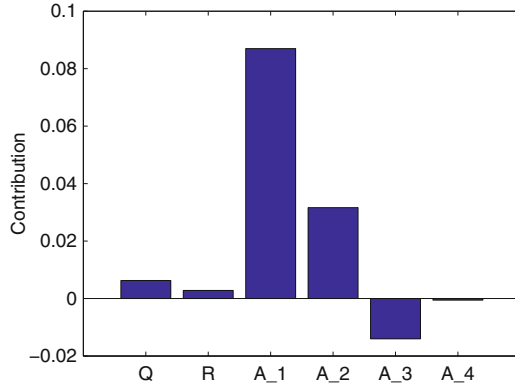
Fig. 9.2 The contributions of the iid component of the environment (**Q**), the autocorrelated component of the environment (**R**), and the projection matrix entries in each environmental state ($\mathbf{A}_1, \dots \mathbf{A}_4$) to the difference in the stochastic growth rate $\log \lambda_s$ between the Rose Prairie (RP) and Fisher Butte (FB) populations of *Lomatium bradshawii*. Calculations assume fire frequencies of 0.5 for FB and 0.7 for RP, and autocorrelations $\rho = -0.5$ for FB and $\rho = 0.5$ for RP

The accuracy of the approximations involved in the LTRE analysis is good. The sum of the contributions in Fig. 9.2 is 0.1192, while the actual difference in $\log \lambda_s$ is 0.1219 (an accuracy of 98%).

Alternatively, suppose that some kind of fire prevention program in the RP site reduced the fire frequency to $f = 0.1$ (well below the critical threshold for persistence), but a fire management program increased the fire frequency in the FP site to $f = 0.9$.

	FB	RP
f	0.9	0.1
ρ	0.0	0.0
$\log \lambda_s$	0.027	−0.113

Now $\log \lambda_s^{(FB)} > \log \lambda_s^{(RP)}$, despite the general advantage in vital rates of RP over FB in most environmental states. Figure 9.3 presents the contributions to $\Delta \log \lambda_s$ from differences in fire frequency, autocorrelation, and vital rates, and shows how the contributions of the vital rate differences are, in this case, overwhelmed by the RP disadvantage due to the stationary distribution of the environment.

The sum of the contributions in Fig. 9.3 is −0.1326, while the actual difference in $\log \lambda_s$ is −0.1395 (an accuracy of 95%, even with a very large difference in growth rate).
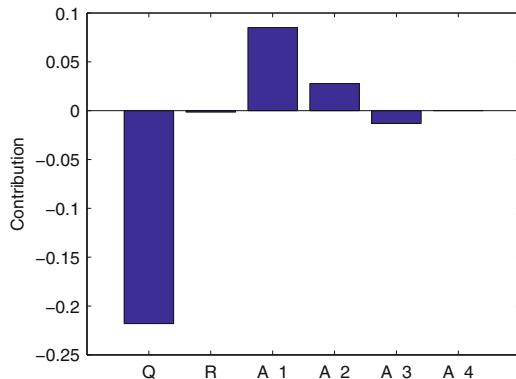
Fig. 9.3 The contributions of the iid component of the environment ($\mathbf{Q}$), the autocorrelated component of the environment ($\mathbf{R}$), and the matrix entries in each environmental state ($\mathbf{A}_1, \ldots \mathbf{A}_4$) to the difference in the stochastic growth rate $\log \lambda_s$ between the Rose Prairie (RP) and Fisher Butte (FB) populations of *Lomatium bradshawii*. Calculations assume fire frequencies of 0.9 for FB and 0.1 for RP, and autocorrelation $\rho = 0$ for both populations

9.7   Discussion

This application of matrix calculus provides a general framework for decomposition analysis of the stochastic growth rate in Markovian environments. It is a direct generalization of the familiar LTRE approaches for time-invariant and periodic models, but combined with the powerful Kitagawa-Keyfitz decomposition. Comparative studies of the stochastic growth rate require additional data on the stochastic dynamics of the environment, beyond that needed for time-invariant models (Fig. 9.1). Many stochastic studies present conditional results; for example, the study of *L. bradshawii* provides $\log \lambda_s$ as a function of f, ρ, and $\mathbf{\Theta}$, but does not estimate the value of $\log \lambda_s$ actually exhibited in either of the two sites. To do so would require long-term data on the stochastic environment, which is hard to come by. However, such information may possibly be extracted from historical data (e.g., Smith et al. 2005; Lawler et al. 2009), or projected from climate models (Hunter et al. 2010; Jenouvrier et al. 2009b).

The methods presented here are not limited to Markovian environments in which the environmental states have an interpretation (years since fire, flood conditions, etc.). They can also be used when matrices are randomly selected from a series collected over time (e.g., the early study of Bierzychudek (1982) based on two yearly matrices, or the study by Jenouvrier et al. (2009b) based on 44 years of matrices for emperor penguins). Although such models are indeed Markov chains, if years are simply a random sample of environmental variation, then it is of little interest to know the contribution of vital rate differences in, say, 1988 compared to 1989 or 1987. In these models, the mean and variance of the vital rates may be of more interest. Davison et al. (2010), drawing on the stochastic elasticity results of

Tuljapurkar et al. ([2003](#)), have presented an approach to LTRE analysis in terms of the contributions of differences in the mean and the variance of the vital rates. That method nicely complements the approach presented here.

In the analysis of *Lomatium bradshawii*, even large differences in environmental autocorrelation made small contributions to treatment effects on $\log \lambda_s$. This is not surprising, given the generally small impact of changes in autocorrelation on the stochastic growth rate in this model (Caswell and Kaye [2001](#)). It is, however, not guaranteed. Given the proper interaction between environmental states and the stage structure, autocorrelation can have dramatic impacts on the growth rate (Caswell [2001](#), Example 14.1). How often this happens in nature will only be revealed by further comparative studies.

Changing focus from plants in a fluctuating fire environment to human populations projected in response to stochastic fluctuations in mortality and fertility (e.g., Tuljapurkar [1992](#); Lee and Tuljapurkar [1994](#)), there are possibilities for applying this approach to population projections. However, such attempts will be challenging because the stochastic environments are not stationary, and the interest is not in asymptotic stochastic growth, but in short term transient dynamics. A combination of the transient analyses in Chap. [7](#) with the decomposition approach here might yield interesting results.

Bibliography

Aberg, P., C. J. Svensson, H. Caswell, and H. Pavia. 2009. Environment-specific elasticity and sensitivity analysis of the stochastic growth rate. Ecological Modelling **220**:605–610.

Bierzychudek, P. 1982. The demography of jack-in-the-pulpit, a forest perennial that changes sex. Ecological Monographs **52**:335–351.

Canudas Romo, V. 2003. Decomposition methods in demography. Population Studies, Rozenberg Publishers, Amsterdam, Netherlands.

Caswell, H. 1989. Analysis of life table response experiments I. Decomposition of effects on population growth rate. Ecological Modelling **46**:221–237.

Caswell, H., 1996. Demography meets ecotoxicology: untangling the population level effects of toxic substances. Pages 255–292 *in* M. C. Newman and C. H. Jagoe, editors. Ecotoxicology: a hierarchical treatment. Lewis Publishers, Boca Raton, Florida.

Caswell, H. 2001. Matrix Population Models: Construction, Analysis, and Interpretation. 2nd edition. Sinauer Associates, Sunderland, MA.

Caswell, H. 2005. Sensitivity analysis of the stochastic growth rate: three extensions. Australian and New Zealand Journal of Statistics **47**:75–85.

Caswell, H. 2010. Life table response experiment analysis of the stochastic growth rate. Journal of Ecology **98**:324–333.

Caswell, H., and T. N. Kaye. 2001. Stochastic demography and conservation of an endangered perennial plant (*Lomatium bradshawii*) in a dynamic fire regime. Advances in Ecological Research **32**:1–51.

Cohen, J. E. 1976. Ergodicity of age structure in populations with Markovian vital rates I: countable states. Journal of the American Statistical Association **71**:335–339.

Cohen, J. E. 1979. Ergodic theorems in demography. Bulletin of the American Mathematical Society **1**:275–295.

Cohen, J. E. 1986. Population forecasts and confidence intervals for Sweden: A comparison of model-based and empirical approaches. Demography **23**:105–126.

Cohen, J. E., S. W. Christensen, and C. P. Goodyear. 1983. A stochastic age-structured population model of striped bass (*Morone saxatilis*) in the Potomac River. Canadian Journal of Fisheries and Aquatic Sciences **40**:2170–2183.

Davison, R., H. Jacquemyn, D. Adriaens, O. Honnay, H. De Kroon, and S. Tuljapurkar. 2010. Demographic effects of extreme weather events on a short-lived calcareous grassland species: stochastic life table response experiments. Journal of Ecology **98**:255–267.

Gervais, J. A., C. M. Hunter, and R. G. Anthony. 2006. Interactive effects of prey and *p,p'*-DDE on burrowing owl population dynamics. Ecological Applications **16**:666–677.

Horiuchi, S., J. R. Wilmoth, and S. D. Pletcher. 2008. A decomposition method based on a model of continuous change. Demography **45**:785–801.

Horvitz, C. C., S. Tuljapurkar, and J. B. Pascarella. 2005. Plant-animal interactions in random environments: habitat-stage elasticity, seed predators, and hurricanes. Ecology **86**:3312–3322.

Hunter, C. M., H. Caswell, M. C. Runge, E. V. Regehr, S. C. Amstrup, and I. Stirling. 2010. Climate change threatens polar bear populations: a stochastic demographic analysis. Ecology **91**:2883–2897.

Jenouvrier, S., C. Barbraud, H. Weimerskirch, and H. Caswell. 2009a. Limitation of population recovery: a stochastic approach to the case of the emperor penguin. Oikos **118**:1292–1298.

Jenouvrier, S., H. Caswell, C. Barbraud, M. Holland, J. Strœve, and H. Weimerskirch. 2009b. Demographic models and IPCC climate projections predict the decline of an emperor penguin population. Proceedings of the National Academy of Sciences **106**:1844–1847.

Kaye, T. N., K. L. Pendergrass, K. Finley, and J. B. Kauffman. 2001. The effect of fire on the population viability of an endangered prairie plant. Ecological Applications **11**:1366–1380.

Keyfitz, N. 1968. Introduction to the Mathematics of Population. Addison-Wesley, Reading, Massachusetts, USA.

Keyfitz, N., and H. Caswell. 2005. Applied mathematical demography. 3rd edition. Springer, New York, New York.

Kitagawa, E. M. 1955. Components of a difference between two rates. Journal of the American Statistical Association **50**:1168–1194.

Lawler, R. L., H. Caswell, A. F. Richard, J. Ratsirarson, R. E. Dewar, and M. Schwartz. 2009. Population dynamics of Verreaux's sifaka in a stochastic rainfall environment. Oecologia **161**:491–504.

Lee, R. D., and S. Tuljapurkar. 1994. Stochastic population forecasts for the United States: Beyond high, medium, and low. Journal of the American Statistical Association **89**:1175–1189.

Pendergrass, K. L., P. M. Miller, J. B. Kauffman, and T. N. Kaye. 1999. The role of prescribed burning in maintenance of an endangered plant species, *Lomatium bradshawii*. Ecological Applications **9**:1420–1429.

Pollard, J. H. 1988. On the decomposition of changes in expectation of life and differentials in life expectancy. Demography **25**:265–276.

Silva, J. F., J. Raventos, H. Caswell, and M. C. Trevisan. 1991. Population responses to fire in a tropical savanna grass *Andropogon semiberbis*: a matrix model approach. Journal of Ecology **79**:345–356.

Smith, M., H. Caswell, and P. Mettler-Cherry. 2005. Stochastic flood and precipitation regimes and the population dynamics of a threatened floodplain plant. Ecological Applications **15**:1036–1052.

Steel, R. G. D., and J. H. Torrie. 1960. Principles and procedures of statistics. McGraw-Hill, New York, New York.

Steinsaltz, D., S. Tuljapurkar, and C. Horvitz. 2011. Derivatives of the stochastic growth rate. Theoretical Population Biology **80**:1–15.

Svensson, C. J., H. Pavia, and P. Aberg. 2009. Robustness of life history of the brown seaweed *Ascophyllum nodosum* (Fucales, Phaeophyceae) across large scales: effects of spatially and temporally induced variability on population growth. Marine Biology **156**:1139–1148.

Tuljapurkar, S. 1990. Population Dynamics in Variable Enviroinments. Springer-Verlag, New York, New York.

Tuljapurkar, S. 1992. Stochastic population forecasts and their uses. International Journal of Forecasting **8**:385–391.

Tuljapurkar, S., C. Horvitz, and J. B. Pascarella. 2003. The many growth rates and elasticities of populations in random environments. American Naturalist **162**:489–502.

Tuljapurkar, S., and S. H. Orzack. 1980. Population dynamics in variable environments I. Long-run growth rates and extinction. Theoretical Population Biology **18**:314–342.

Part IV
Nonlinear Models

Chapter 10
Sensitivity Analysis of Nonlinear Demographic Models

10.1   Introduction

Nonlinearities in demographic models arise due to density dependence, frequency dependence (in 2-sex models), feedback through the environment or the economy, recruitment subsidy due to immigration, and from the scaling inherent in calculations of proportional population structure. This chapter presents a series of analyses particular to nonlinear models: the sensitivity and elasticity of equilibria, cycles, ratios (e.g., dependency ratios), age averages and variances, temporal averages and variances, life expectancies, and population growth rates, for both age-classified and stage-classified models.

Nonlinearity is defined in contrast to linearity. If $\mathbf{x}$ is an age or stage distribution vector, and if the dynamics of $\mathbf{x}$ are given by

$$\mathbf{x}(t + 1) = f[\mathbf{x}(t)], \tag{10.1}$$

then the model is linear if $f(\cdot)$ is a linear function, i.e., if

$$f(a\mathbf{x}_1 + b\mathbf{x}_2) = af(\mathbf{x}_1) + bf(\mathbf{x}_2) \tag{10.2}$$

for any constants a and b and any vectors $\mathbf{x}_1$ and $\mathbf{x}_2$.

If a model is not linear, it is nonlinear. Not surprisingly, this covers a lot of territory, but nonlinearity in demographic models can be classified into four main sources: density dependence, environmental feedback, interactions between the sexes, and models that arise in calculation of proportional structure.

Density dependence: arises when one or more of the per-capita vital rates are functions of the numbers or density of the population. Such effects have been incorporated into demographic studies of plants (e.g., Solbrig et al. 1988; Gillman et al. 1993; Silva Matos et al. 1999; Pardini et al. 2009; Shyu et al. 2013) and animals (e.g., Pennycuick 1969; Clutton-Brock et al. 1997; Cushing et al. 2003; Bonenfant et al. 2009). Density dependence has been intensively studied in the laboratory (e.g., Pearl et al. 1927; Frank et al. 1957; Costantino and Desharnais 1991; Carey et al. 1995; Mueller and Joshi 2000; Cushing et al. 2003). It can arise from competition for food, space, or other resources, or from interactions (e.g., cannibalism) among individuals.

Simple density dependence is less often invoked by human demographers[1]. Weiss and Smouse (1976) proposed a density-dependent matrix model, and Wood and Smouse (1982) applied it to the Gainj people of Papua New Guinea. Density dependence is included in epidemiological feedback models applied to a rural English population in the sixteenth and seventeenth centuries by Scott and Duncan (1998).

The Easterlin effect (1961) produces density dependence in which fertility is a function of cohort size. Analysis of the Easterlin effect has focused mostly on the possibility that it could generate cycles in births (e.g., Lee 1974, 1976; Frauenthal and Swick 1983; Wachter and Lee 1989; Chu 1998).

Environmental (or economic) feedback.   Density-dependent models are often an attempt to sneak in, by the back door as it were, a feedback through the environment. A change in population size changes some aspect of the environment, which affects the vital rates, which in turn affect future population size. Models in which the feedback operates through resource consumption are the basis for the food chain and food web models that underlie models of global biogeochemistry (e.g.,. Hsu et al. 1977; Tilman 1982; Murdoch et al. 2003; Fennel and Neumann 2004). These models are typically unstructured, but there is a rich literature on structured models, written as partial differential equations, to incorporate physiological structure and resource feedback (de Roos and Persson 2013).

Feedback models are also invoked in human demography, with the feedback operating through the economy (Lee 1986, 1987; Chu 1998). An interesting aspect of these approaches is the possibility that, if larger populations support more robust economies, the feedback could be positive instead of negative (Lee 1986; Cohen 1995, Appendix 6). An exciting combination of ecological and

---

[1]Lee (1987) reviewed the situation and said "...we might say that human demography is all about Leslie matrices and the determinants of unconstrained growth in linear models, whereas animal population studies are all about Malthusian equilibrium through density dependence in nonlinear models ...". He admits that this is an exaggeration, and there clearly are nonlinear concerns in human demography (Bonneuil 1994), but a non-exhaustive survey finds no mention of density dependence in several contemporary human demography texts (e.g., Hinde 1998; Preston et al. 2001; Keyfitz and Caswell 2005).

economic feedback appears in the food ratio model recently proposed by Lee
and Tuljapurkar (2008).

Two-sex models.   To the extent that both males and females are required for
reproduction (and, in the bigger scheme of things, this is not always so),
demography is nonlinear because the marriage function or mating function
cannot satisfy (10.2). Nonlinear two-sex models have a long tradition in human
demography (see reviews in Keyfitz 1972; Pollard 1977) and have been applied
in ecology (e.g., Lindström and Kokko 1998; Legendre et al. 1999; Kokko and
Rankin 2006; Lenz et al. 2007; Jenouvrier et al. 2010, 2012). Their mathematical
properties have been investigated by e.g, Caswell and Weeks (1986), Chung
(1994) and Iannelli et al. (2005) and in a very abstract setting by Nussbaum
(1988, 1989).

   In their most basic form, two-sex models differ from density-dependent
models in that the vital rates depend only on the relative, not the absolute,
abundances of stages in the population (they are sometimes called frequency-
dependent for this reason). This has important implications for their dynamics.

Models for proportional population structure.   Even when the dynamics of
abundance are linear, the dynamics of *proportional* population structure are
nonlinear (e.g., Tuljapurkar 1997). This leads to some useful results on the
sensitivity of the stable age or stage distribution and the reproductive value.

   Linear models lead to exponential growth and convergence to a stable structure.
Much of their analysis focuses on the population growth rate λ or $r = \log \lambda$.
Nonlinear models do not usually lead to exponential growth (frequency-dependent
two-sex models are an exception). Instead, their trajectories converge to an attractor.
The attractor may be an equilibrium point, a cycle, an invariant loop (yielding
quasiperiodic dynamics), or a strange attractor (yielding chaotic dynamics); see
Cushing (1998) or Caswell (2001, Chapter 16) for a detailed discussion.

   This chapter analyzes the sensitivity and elasticity of equilibria and cycles.
Because the dynamic models considered here are discrete, solutions always exist
and are unique. The nature and the number of the attractors depends on the specific
model. Perturbation analysis always considers perturbations of *something*, so the
equilibria or cycles must be found before their perturbation properties can be
analyzed.

10.2   Density-Dependent Models

We begin with the basic discrete-time[2] density-dependent model, written as

$$\mathbf{n}(t+1) = \mathbf{A}[\boldsymbol{\theta}, \mathbf{n}(t)] \, \mathbf{n}(t) \qquad (10.3)$$

---

[2]It is possible to generalize to continuous-time models, that would be written

$$\frac{d\mathbf{n}}{dt} = \mathbf{A}[\boldsymbol{\theta}, \mathbf{n}(t)] \, \mathbf{n}(t)$$

where $\mathbf{n}(t)$ is a population vector of dimension $s \times 1$ and $\mathbf{A}$ is a population projection matrix of dimension $s \times s$. The matrix $\mathbf{A}$ depends on a $p \times 1$ vector $\boldsymbol{\theta}$ of parameters as well as on the current population vector $\mathbf{n}(t)$.[3]

10.2.1 Linearizations Around Equilibria

An equilibrium of (10.3) satisfies

$$\hat{\mathbf{n}} = \mathbf{A}\left[\boldsymbol{\theta}, \hat{\mathbf{n}}\right]\hat{\mathbf{n}}. \tag{10.4}$$

Such an equilibrium may be stable (small perturbations from $\hat{\mathbf{n}}$ eventually return to the equilibrium) or unstable.[4] That stability is determine by the linearization of the nonlinear system (10.3) near $\hat{\mathbf{x}}$. That is, define the deviation from $\hat{\mathbf{x}}$ as $\mathbf{z}(t) = \mathbf{x}(t) - \hat{\mathbf{x}}$. Then $\mathbf{z}(t)$ follows

$$\mathbf{z}(t+1)\mathbf{M}[\boldsymbol{\theta}, \hat{\mathbf{x}}]\mathbf{z}(t) \tag{10.5}$$

The matrix $\mathbf{M}$ is the Jacobian matrix,

$$\mathbf{M} = \left.\frac{\partial \mathbf{x}(t+1)}{\partial \mathbf{x}^{\mathsf{T}}(t)}\right|_{\hat{\mathbf{x}}} \tag{10.6}$$

To obtain $\mathbf{M}$, differentiate both sides of (10.3),

$$d\mathbf{x}(t+1) = (d\mathbf{A})\,\mathbf{x} + \mathbf{A}\,(d\mathbf{x}) \tag{10.7}$$

Applying the vec operator to both sides gives

$$d\mathbf{x}(t+1) = \left(\mathbf{x}^{\mathsf{T}} \otimes \mathbf{I}_s\right) d\mathrm{vec}\,\mathbf{A} + \mathbf{A}d\mathbf{x} \tag{10.8}$$

from which

$$\mathbf{M} = \left(\mathbf{x}^{\mathsf{T}} \otimes \mathbf{I}_s\right)\frac{d\mathrm{vec}\,\mathbf{A}}{d\mathbf{x}^{\mathsf{T}}} + \mathbf{A} \tag{10.9}$$

---

for some appropriately defined matrix function $\mathbf{A}$; see Verdy and Caswell (2008). Such models are less often used, but see Shyu and Caswell (2016a, 2018) for a two-sex model example.

[3]The explicit dependence on $\boldsymbol{\theta}$ and $\mathbf{n}(t)$ will be neglected when it is obvious from the context.

[4]A careful consideration of stability requires more care with the definition of these terms, but will not concern us here. See Caswell (2001) and Cushing (1998) for more details.

where $\mathbf{I}_s$ is an identity matrix of order s. The linearization at the equilibrium is obtained by evaluating $\mathbf{M}$ at $\mathbf{x} = \hat{\mathbf{x}}$:

$$\mathbf{M}\left[\boldsymbol{\theta}, \hat{\mathbf{x}}\right] = \left(\hat{\mathbf{x}}^\mathsf{T} \otimes \mathbf{I}_s\right) \frac{\partial \operatorname{vec} \mathbf{A}\left[\boldsymbol{\theta}, \hat{\mathbf{x}}\right]}{\partial \mathbf{x}^\mathsf{T}} + \mathbf{A}\left[\boldsymbol{\theta}, \hat{\mathbf{x}}\right] \tag{10.10}$$

If all the eigenvalues of $\mathbf{M}$ are less than one in magnitude, the equilibrium $\hat{\mathbf{x}}$ is locally asymptotically stable. The linearization also provides valuable information about short-term transient responses to perturbation; see Sect. 10.2.4.

10.2.2 Sensitivity of Equilibrium

Our goal is to find the derivatives of all the entries of $\hat{\mathbf{n}}$ with respect to all of the parameters in $\boldsymbol{\theta}$; these are the entries of the $s \times p$ matrix

$$\frac{d\hat{\mathbf{n}}}{d\boldsymbol{\theta}^\mathsf{T}}.$$

We begin by taking the differential of both sides of (10.4):

$$d\hat{\mathbf{n}} = (d\mathbf{A})\hat{\mathbf{n}} + \mathbf{A}(d\hat{\mathbf{n}}). \tag{10.11}$$

Rewrite this as

$$d\hat{\mathbf{n}} = \mathbf{I}_s(d\mathbf{A})\hat{\mathbf{n}} + \mathbf{A}(d\hat{\mathbf{n}}), \tag{10.12}$$

where $\mathbf{I}_s$ is an identity matrix of dimension s. Next apply the vec operator to both sides, remembering that since $\hat{\mathbf{n}}$ is a column vector, $\operatorname{vec} \hat{\mathbf{n}} = \hat{\mathbf{n}}$, and apply Roth's theorem, to obtain

$$d\hat{\mathbf{n}} = \left(\hat{\mathbf{n}}^\mathsf{T} \otimes \mathbf{I}_s\right) d\operatorname{vec} \mathbf{A} + \mathbf{A}d\hat{\mathbf{n}}. \tag{10.13}$$

However, $\mathbf{A}$ is a function of both $\boldsymbol{\theta}$ and $\hat{\mathbf{n}}$, so

$$d\operatorname{vec} \mathbf{A} = \frac{\partial \operatorname{vec} \mathbf{A}}{\partial \boldsymbol{\theta}^\mathsf{T}} d\boldsymbol{\theta} + \frac{\partial \operatorname{vec} \mathbf{A}}{\partial \mathbf{n}^\mathsf{T}} d\hat{\mathbf{n}}. \tag{10.14}$$

Substituting (10.14) into (10.13) and applying the chain rule leads to[5]

$$\frac{d\hat{\mathbf{n}}}{d\boldsymbol{\theta}^\mathsf{T}} = \left(\hat{\mathbf{n}}^\mathsf{T} \otimes \mathbf{I}_s\right) \left(\frac{\partial \operatorname{vec} \mathbf{A}}{\partial \boldsymbol{\theta}^\mathsf{T}} + \frac{\partial \operatorname{vec} \mathbf{A}}{\partial \mathbf{n}^\mathsf{T}} \frac{d\hat{\mathbf{n}}}{d\boldsymbol{\theta}^\mathsf{T}}\right) + \mathbf{A}\frac{d\hat{\mathbf{n}}}{d\boldsymbol{\theta}^\mathsf{T}}. \tag{10.15}$$

---

[5]It is reassuring to check that the dimensions of all these quantities are compatible:

Finally, solve (10.15) for $d\hat{\mathbf{n}}/d\boldsymbol{\theta}^\mathsf{T}$ to obtain

$$\frac{d\hat{\mathbf{n}}}{d\boldsymbol{\theta}^\mathsf{T}} = \left(\mathbf{I}_s - \mathbf{A} - \left(\hat{\mathbf{n}}^\mathsf{T} \otimes \mathbf{I}_s \right) \frac{\partial \operatorname{vec} \mathbf{A}}{\partial \mathbf{n}^\mathsf{T}} \right)^{-1} \left(\hat{\mathbf{n}}^\mathsf{T} \otimes \mathbf{I}_s \right) \frac{\partial \operatorname{vec} \mathbf{A}}{\partial \boldsymbol{\theta}^\mathsf{T}} \qquad (10.16)$$

where $\mathbf{A}$, $\partial \operatorname{vec} \mathbf{A}/\partial \boldsymbol{\theta}^\mathsf{T}$, and $\partial \operatorname{vec} \mathbf{A}/\partial \hat{\mathbf{n}}^\mathsf{T}$ are evaluated at $\hat{\mathbf{n}}$.

Comparing (10.16) and Eq. (10.10) for the linearization, we see that the sensitivity of equilibrium can be written

$$\frac{d\hat{\mathbf{n}}}{d\boldsymbol{\theta}^\mathsf{T}} = (\mathbf{I}_s - \mathbf{M})^{-1} \left(\hat{\mathbf{n}}^\mathsf{T} \otimes \mathbf{I}_s \right) \frac{\partial \operatorname{vec} \mathbf{A}}{\partial \boldsymbol{\theta}^\mathsf{T}}. \qquad (10.17)$$

The matrix $(\mathbf{I}_s - \mathbf{M})$ is singular if 1 is an eigenvalue of $\mathbf{M}$; i.e., at a bifurcation point when the equilibrium $\hat{\mathbf{n}}$ becomes unstable. At that point, quite appropriately, the sensitivity is not defined because the change in the equilibrium is not continuous.

The following example, applying (10.16) to a simple model, shows the basic steps and output of the analysis.

Example 1: A simple two-stage model The most basic distinction in the life cycle of many organisms is between non-reproducing juveniles and reproducing adults. A model based on these stages (Neubert and Caswell 2000) is parameterized by the juvenile survival σ_1, the adult survival σ_2, the growth or maturation probability γ (the expected time to maturity is $1/\gamma$), and the adult fertility f. The projection matrix is

$$\mathbf{A} = \begin{pmatrix} \sigma_1(1-\gamma) & f \\ \sigma_1\gamma & \sigma_2 \end{pmatrix}. \qquad (10.18)$$

Any of the vital rates could be density-dependent; here we suppose that juvenile survival σ_1 depends on total density:

$$\sigma_1(\mathbf{n}) = \tilde{\sigma} \exp(-\mathbf{1}^\mathsf{T}\mathbf{n}); \qquad (10.19)$$

where $\mathbf{1}$ is a vector of ones.

Define the parameter vector as $\boldsymbol{\theta} = \begin{pmatrix} f & \gamma & \tilde{\sigma} & \sigma_2 \end{pmatrix}^\mathsf{T}$. To apply (10.16) requires the derivatives of $\mathbf{A}[\boldsymbol{\theta}, \mathbf{n}]$ with respect to $\boldsymbol{\theta}$ and with respect to $\mathbf{n}$. These are

---

$$\underbrace{\frac{d\hat{\mathbf{n}}}{d\boldsymbol{\theta}^\mathsf{T}}}_{s \times p} = \underbrace{\left(\hat{\mathbf{n}}^\mathsf{T} \otimes \mathbf{I}_s \right)}_{s \times s^2} \left(\underbrace{\frac{\partial \operatorname{vec} \mathbf{A}}{\partial \boldsymbol{\theta}^\mathsf{T}}}_{s^2 \times p} + \underbrace{\frac{\partial \operatorname{vec} \mathbf{A}}{\partial \mathbf{n}^\mathsf{T}}}_{s^2 \times s} \underbrace{\frac{\partial \hat{\mathbf{n}}}{\partial \boldsymbol{\theta}^\mathsf{T}}}_{s \times p} \right) + \underbrace{\mathbf{A}}_{s \times s} \underbrace{\frac{d\hat{\mathbf{n}}}{d\boldsymbol{\theta}^\mathsf{T}}}_{s \times p}.$$

$$\frac{d\text{vec}\,\mathbf{A}}{df} = \text{vec}\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \tag{10.20}$$

$$\frac{d\text{vec}\,\mathbf{A}}{d\gamma} = \text{vec}\begin{pmatrix} -\sigma_1(\mathbf{n}) & 0 \\ \sigma_1(\mathbf{n}) & 0 \end{pmatrix} \tag{10.21}$$

$$\frac{d\text{vec}\,\mathbf{A}}{d\tilde{\sigma}} = \text{vec}\begin{pmatrix} (1-\gamma)\exp(-\mathbf{1}^\mathsf{T}\mathbf{n}) & 0 \\ \gamma\exp(-\mathbf{1}^\mathsf{T}\mathbf{n}) & 0 \end{pmatrix} \tag{10.22}$$

$$\frac{d\text{vec}\,\mathbf{A}}{d\sigma_2} = \text{vec}\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \tag{10.23}$$

$$\frac{d\text{vec}\,\mathbf{A}}{dn_1} = \frac{d\text{vec}\,\mathbf{A}}{dn_2} = \text{vec}\begin{pmatrix} -\sigma_1(\mathbf{n})(1-\gamma) & 0 \\ -\sigma_1(\mathbf{n})\gamma & 0 \end{pmatrix}. \tag{10.24}$$

The derivative of $\mathbf{A}$ with respect to the $\boldsymbol{\theta}$ is the 4×4 matrix

$$\frac{\partial\text{vec}\,\mathbf{A}}{\partial\boldsymbol{\theta}^\mathsf{T}} = \begin{pmatrix} 0 & -\sigma_1(\mathbf{n}) & (1-\gamma)\exp(-\mathbf{1}^\mathsf{T}\mathbf{n}) & 0 \\ 0 & \sigma_1(\mathbf{n}) & \gamma\exp(-\mathbf{1}^\mathsf{T}\mathbf{n}) & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \tag{10.25}$$

where each column corresponds to an entry of $\boldsymbol{\theta}$ and each row to an element of vec $\mathbf{A}$. The derivative of $\mathbf{A}$ with respect to $\mathbf{n}$ is

$$\frac{\partial\text{vec}\,\mathbf{A}}{\partial\mathbf{n}^\mathsf{T}} = \begin{pmatrix} -\sigma_1(\mathbf{n})(1-\gamma) & -\sigma_1(\mathbf{n})(1-\gamma) \\ -\sigma_1(\mathbf{n})\gamma & -\sigma_1(\mathbf{n})\gamma \\ 0 & 0 \\ 0 & 0 \end{pmatrix}. \tag{10.26}$$

Each column corresponds to an entry of $\mathbf{n}$ and each row to an element of vec $\mathbf{A}$.

Using some arbitrary parameter values (not unreasonable for humans or other large mammals)

$$f = 0.25$$
$$\gamma = 1/15$$
$$\tilde{\sigma} = 0.98$$
$$\sigma_2 = 0.95$$

leads to an equilibrium population

$$\hat{\mathbf{n}} = \begin{pmatrix} 0.1053 \\ 0.1109 \end{pmatrix}, \tag{10.27}$$

obtained by iterating the model to convergence.

These patterns reflect the life history, although comparative study of this dependence has scarcely begun. For example, if the demographic parameters were more appropriate for an insect, say with high fertility ($f = 70$), rapid maturation ($\gamma = 0.9$), and low juvenile survival ($\tilde{\sigma} = 0.1$), and in which most adults die after reproducing once ($\sigma_2 = 0.01$), then the equilibrium would become

$$\hat{\mathbf{n}} = \begin{pmatrix} 1.826 \\ 0.026 \end{pmatrix} \tag{10.28}$$

with sensitivities

$$\frac{d\hat{\mathbf{n}}}{d\boldsymbol{\theta}^{\mathsf{T}}} = \begin{pmatrix} 0.01 & 1.08\ 9.86\ 0.99 \\ -0.0002 & 0.02\ 0.14\ 0.01 \end{pmatrix}. \tag{10.29}$$

In this life history, increases in fertility have very small effects on the equilibrium population, and the effect of increased fertility on adult density is slightly negative. Changes in the maturation rate or in juvenile or adult survival have much larger impacts on juvenile density than on adult density. ∎

10.2.3   Dependent Variables: Beyond $\hat{\mathbf{n}}$

The equilibrium vector $\hat{\mathbf{n}}$ is usually not the only dependent variable of interest. If we write $\mathbf{m} = \mathbf{m}(\mathbf{n})$ for any vector- or scalar-valued transformation of $\mathbf{n}$, then the sensitivity of $\mathbf{m}$ is just

$$\frac{d\hat{\mathbf{m}}}{d\boldsymbol{\theta}^{\mathsf{T}}} = \frac{d\hat{\mathbf{m}}}{d\mathbf{n}^{\mathsf{T}}} \frac{d\hat{\mathbf{n}}}{d\boldsymbol{\theta}^{\mathsf{T}}}. \tag{10.30}$$

The possibilities for dependent variables are, roughly speaking, limited only by one's imagination. The following is a list of examples.

1. Weighted population density. Let $\mathbf{c} \geq 0$ be a vector of weights. Weighted population density is then $N(t) = \mathbf{c}^{\mathsf{T}}\mathbf{n}(t)$. Examples include total density ($\mathbf{c} = \mathbf{1}$), the density of a subset of stages ($c_i = 1$ for stages to be counted; $c_i = 0$ otherwise), biomass (c_i is the biomass of stage i), basal area, metabolic rate, etc. The sensitivity of $\hat{N}$ is

$$\frac{d\hat{N}}{d\boldsymbol{\theta}^{\mathsf{T}}} = \mathbf{c}^{\mathsf{T}} \frac{d\hat{\mathbf{n}}}{d\boldsymbol{\theta}^{\mathsf{T}}}. \tag{10.31}$$

2. Ratios, measuring the relative abundances of different stages. Let

$$R(t) = \frac{\mathbf{a}^{\mathsf{T}}\mathbf{n}(t)}{\mathbf{b}^{\mathsf{T}}\mathbf{n}(t)} \tag{10.32}$$

where $\mathbf{a} \geq 0$ and $\mathbf{b} \geq 0$ are weight vectors. Examples include the dependency ratio (in human populations, the ratio of the individuals below 15 or above 65 to those between 15 and 65; see Sect. 10.5.3), the sex ratio, and the ratio of juveniles to adults, which is used in wildlife management; see Skalski et al. (2005). Differentiating (10.32) gives

$$\frac{d\hat{R}}{d\boldsymbol{\theta}^{\mathsf{T}}} = \left(\frac{\mathbf{b}^{\mathsf{T}}\hat{\mathbf{n}}\mathbf{a}^{\mathsf{T}} - \mathbf{a}^{\mathsf{T}}\hat{\mathbf{n}}\mathbf{b}^{\mathsf{T}}}{\left(\mathbf{b}^{\mathsf{T}}\hat{\mathbf{n}}\right)^2} \right) \frac{d\hat{\mathbf{n}}}{d\boldsymbol{\theta}^{\mathsf{T}}}. \tag{10.33}$$

3. Age or stage averages. These include quantities such as the mean age or size in the stable population or at equilibrium and the mean age at reproduction in the stable population. Their perturbation analysis is presented in Sect. 10.5.4.
4. Properties of cycles. Nonlinear models may produce population cycles. Attention may focus on the mean, the variance, or higher moments of the population vector or of some scalar measure of density, over such cycles. The sensitivity of these moments is explored in Sect. 10.7.

10.2.4 Reactivity and Transient Dynamics

The asymptotic stability of an equilibrium is determined by the eigenvalues of the Jacobian matrix $\mathbf{M}$ in (10.9), evaluated at that equilibrium. In the short term, however, perturbations of the population away from the equilibrium can exhibit transient dynamics that differ from their asymptotic behavior. In particular, perturbations of stable equilibria, that are destined to eventually return to the equilibrium, may move (much) farther away before that return. Neubert and Caswell (1997) introduced three indices, each calculated from $\mathbf{M}$, to quantify these transient responses.[6] The *reactivity* of an asymptotically stable equilibrium is the maximum, over all perturbations, of the rate at which the trajectory departs from the equilibrium. At any time following a perturbation, there is a maximum (over all perturbations) deviation

---

[6]Because these indices are calculated from $\mathbf{M}$, they are properly considered properties of the system and its dynamics. Stott et al. (2011) and Stott (2016) have also considered indices of transient response that reflect the particular initial condition rather than the inherent dynamics of the system.

from the equilibrium. This maximum is the *amplification envelope*. It gives an upper bound on the extent of transient amplification as a function of time. The phrase "over all perturbations" in these definitions signals that the transient amplification depends on the direction of the perturbation. The perturbation that produces the maximum amplification at any specified time is the *optimal perturbation* (Verdy and Caswell 2008).[7]

The transient dynamics of the perturbed system are described by the evolution of the magnitude of $\mathbf{z}$, as measured by the Euclidean norm $\|\mathbf{z}\| = \sqrt{\mathbf{z}^\mathsf{T} \mathbf{z}}$. The reactivity is the maximum, over all perturbations, of the growth rate of $\|\mathbf{z}\|$, as $t \to 0$, and is given by

$$v_0 = \begin{cases} \lambda_1 \left[\mathbf{H}(\mathbf{M}) \right] \text{ continuous time} \\ \log \sigma_1 (\mathbf{M}) \text{ discrete time} \end{cases} \tag{10.34}$$

The matrix $\mathbf{H}(\mathbf{M}) = \left(\mathbf{M} + \mathbf{M}^\mathsf{T} \right) / 2$ is the Hermitian part of $\mathbf{M}$ and λ_1 denotes the eigenvalue with largest real part (Neubert and Caswell 1997). In discrete time, reactivity is the log of the largest singular value of $\mathbf{M}$, which we denote $\sigma_1(\mathbf{M})$.

The amplification envelope is

$$\rho(t) = \begin{cases} \sigma_1 \left(e^{\mathbf{M}t} \right) \text{ continuous} \\ \sigma_1 \left(\mathbf{M}^t \right) \text{ discrete} \end{cases} \tag{10.35}$$

The optimal perturbation, normalized to length 1, is given by the right singular vector corresponding to the singular value that defines $\rho(t)$.

Verdy and Caswell (2008) presented a complete sensitivity analysis of reactivity, the amplification envelope, and the optimal perturbation, in both continuous and discrete time. Suppose the ξ be one of the indices, and suppose that the model depends on a parameter vector $\boldsymbol{\theta}$. Changes in $\boldsymbol{\theta}$ will change the equilibrium vector, which will contribute to changes in the Jacobian matrix, so that the sensitivity of ξ to $\boldsymbol{\theta}$ is

$$\frac{d\xi}{d\boldsymbol{\theta}^\mathsf{T}} = \left(\frac{d\xi}{d\mathrm{vec}^\mathsf{T} \mathbf{M}} \right) \left(\frac{\partial \mathrm{vec}\, \mathbf{M}}{\partial \boldsymbol{\theta}^\mathsf{T}} + \frac{\partial \mathrm{vec}\, \mathbf{M}}{\partial \hat{\mathbf{n}}^\mathsf{T}} \frac{d\hat{\mathbf{n}}}{d\boldsymbol{\theta}^\mathsf{T}} \right) \tag{10.36}$$

The sensitivity of ξ in (10.36) requires four pieces: the linearization $\mathbf{M}$ at the equilibrium, which is given by (10.10), the sensitivity of the equilibrium $\hat{\mathbf{n}}$ to the parameters, which is given by (10.16), the sensitivity of the Jacobian matrix $\mathbf{M}$ to the parameters, and the sensitivity of the index ξ to the matrix $\mathbf{M}$. The sensitivity

---

[7]It is now known that reactivity is a common property of many ecological systems, including populations described by discrete matrix population models (Neubert and Caswell 1997; Chen and Cohen 2001; Neubert et al. 2004; Marvier et al. 2004; Caswell and Neubert 2005; Verdy and Caswell 2008).

of ξ to **M** depends on which index, but the calculations involve perturbations of eigenvalues, singular values, or the matrix exponential, and are given in Verdy and Caswell (2008). The derivative of the linearization **M** is obtained by differentiating all the terms in Eq. (10.10); the result, along with several examples, is given in Verdy and Caswell (2008, eq. (37)).

10.2.5   *Elasticity Analysis*

The derivatives in the matrix $d\hat{\mathbf{n}}/d\boldsymbol{\theta}^\mathsf{T}$ give the results of small additive perturbations of the parameters. It is often useful to study the elasticities, which give the proportional result of small proportional perturbations,

$$\frac{\epsilon\hat{\mathbf{n}}}{\epsilon\boldsymbol{\theta}^\mathsf{T}} = \mathcal{D}\left(\hat{\mathbf{n}}\right)^{-1} \frac{d\hat{\mathbf{n}}}{d\boldsymbol{\theta}^\mathsf{T}}\,\mathcal{D}\left(\boldsymbol{\theta}\right), \tag{10.37}$$

The elasticity of any other (scalar- or vector-valued) dependent variable $f(\hat{\mathbf{n}})$ is given by

$$\frac{\epsilon f(\hat{\mathbf{n}})}{\epsilon\boldsymbol{\theta}^\mathsf{T}} = \mathcal{D}\left(f(\hat{\mathbf{n}})\right)^{-1} \frac{df(\hat{\mathbf{n}})}{d\boldsymbol{\theta}^\mathsf{T}}\,\mathcal{D}\left(\boldsymbol{\theta}\right). \tag{10.38}$$

As usual, elasticities can only be calculated when $\boldsymbol{\theta} \geq 0$ and $f(\hat{\mathbf{n}}) > 0$.

Example 2: Metabolic population size in *Tribolium* Flour beetles of the genus *Tribolium* have been the subject of a long series of experiments on nonlinear population dynamics (reviewed by Cushing et al. 2003). *Tribolium* lives in stored flour. In addition to feeding on the flour, adults and larvae cannibalize eggs, and adults cannibalize pupae. These interactions are the source of nonlinearity in the demography, and are captured in a three-stage (larvae, pupae, and adults) model. The projection matrix is

$$\mathbf{A}[\boldsymbol{\theta}, \mathbf{n}] = \begin{pmatrix} 0 & 0 & b\exp(-c_{el}n_1 - c_{ea}n_3) \\ 1-\mu_l & 0 & 0 \\ 0 & \exp(-c_{pa}n_3) & 1-\mu_a \end{pmatrix} \tag{10.39}$$

where b is the clutch size, c_{ea}, c_{el}, and c_{pa} are rates of cannibalism (of eggs by adults, eggs by larvae, and pupae by adults, respectively), and μ_l and μ_a are larval and adult mortalities (the mortality of pupae, in these laboratory conditions, is effectively zero). Parameter values from an experiment reported by Costantino et al. (1997)

$$b = 6.598$$

$$c_{ea} = 1.155 \times 10^{-2}$$

$$c_{el} = 1.209 \times 10^{-2}$$

$$c_{pa} = 4.7 \times 10^{-3}$$

$$\mu_a = 7.729 \times 10^{-3}$$

$$\mu_l = 2.055 \times 10^{-1}$$

produce a stable equilibrium

$$\hat{\mathbf{n}} = \begin{pmatrix} 22.6 \\ 18.0 \\ 385.2 \end{pmatrix}. \tag{10.40}$$

The sensitivity of $\hat{\mathbf{n}}$ is calculated using (10.16). However, the damage caused by *Tribolium* as a pest of stored grain products might well depend more on metabolism than on numbers. Emekci et al. (2001) estimated the metabolic rates of larvae, pupae, and adults as 9, 1, and 4.5 μl CO_2 h^{-1}, respectively. We define the metabolic population size as $N_m(t) = \mathbf{c}^\mathsf{T}\mathbf{n}(t)$ where $\mathbf{c}^\mathsf{T} = \begin{pmatrix} 9 & 1 & 4.5 \end{pmatrix}$, and calculate the sensitivity and elasticity of $\hat{N}_m$ using (10.37) and (10.31).

Figure 10.1 shows the elasticity of $\hat{\mathbf{n}}$ and $\hat{N}_m$ to each of the parameters. The elasticities are diverse and perhaps counterintuitive. Increases in fecundity increase the equilibrium density of all stages; increases in the cannibalism of eggs by adults reduces the density of all stages. But increased cannibalism of pupae by adults increases the density of larvae and pupae, as does an increase in the mortality of adults.
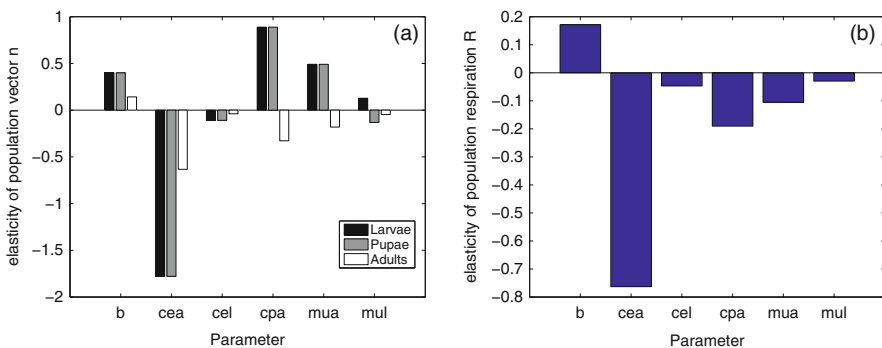
Fig. 10.1 Sensitivity analysis of equilibrium for the flour beetle *Tribolium* in Example 2. (**a**) The elasticity of the equilibrium $\hat{\mathbf{n}}$ to the parameters (see Example 2 for definitions). (**b**) The elasticity of the equilibrium population respiration rate $\hat{N}_m$ to the parameters

When the stages are weighted by their metabolic rate, the elasticity of $\hat{N}_m$ to fecundity is positive, but the elasticities to all the other parameters (cannibalism rates and mortalities) are negative. The positive effects of c_{pa} and μ_a on $\hat{\mathbf{n}}$ disappear when the stages are weighted according to metabolism. ∎

10.2.6 Continuous-Time Models

We have focused on discrete-time models throughout this book. An analogous perturbation analysis can be carried out on continuous-time models of the form

$$\frac{d\mathbf{n}}{dt} = \mathbf{A}\left[\mathbf{n}(t)\right]\mathbf{n}(t) \qquad (10.41)$$

Verdy and Caswell (2008) present a parallel presentation of the continuous and discrete models. The linearization at $\hat{\mathbf{n}}$ is, once again, given by (10.10). If all the eigenvalues of $\mathbf{M}$ have negative real parts, the equilibrium is locally stable.

The sensitivity of the equilibrium $\hat{\mathbf{n}}$ is

$$\frac{d\hat{\mathbf{n}}}{d\boldsymbol{\theta}^\mathsf{T}} = \left\{-\mathbf{A} - \left(\hat{\mathbf{n}}^\mathsf{T} \otimes \mathbf{I}_s\right) \frac{\partial \operatorname{vec} \mathbf{A}}{\partial \mathbf{n}^\mathsf{T}}\right\}^{-1} \left(\hat{\mathbf{n}}^\mathsf{T} \otimes \mathbf{I}_s\right) \frac{\partial \operatorname{vec} \mathbf{A}}{\partial \boldsymbol{\theta}^\mathsf{T}}, \qquad (10.42)$$

with $\mathbf{A}$ and all its derivatives evaluated at the equilibrium $\hat{\mathbf{n}}$. Substituting (10.10) for $\mathbf{M}$ gives

$$\frac{d\hat{\mathbf{n}}}{d\boldsymbol{\theta}^\mathsf{T}} = -\mathbf{M}^{-1} \left(\hat{\mathbf{n}}^\mathsf{T} \otimes \mathbf{I}_s\right) \frac{\partial \operatorname{vec} \mathbf{A}}{\partial \boldsymbol{\theta}^\mathsf{T}}, \qquad (10.43)$$

and $\mathbf{M}$ is nonsingular unless 0 is an eigenvalue of $\mathbf{M}$, which corresponds to a bifurcation point of the equilibrium.

10.3 Environmental Feedback Models

Environmental (or economic) feedback models write the vital rates as functions of some environmental variable, which in turn depends on population density. Feedback models may be static or dynamic. In static feedback models, the environment depends only on current conditions, with no inherent dynamics of its own. In dynamic feedback models, the environment can have dynamics as complicated as those of the population (e.g., if the environmental variable was the abundance of a prey species, affecting the dynamics of a predator species). The sensitivity analysis of dynamic feedback models is given in Sect. 10.8.

A static feedback model can be written

$$\mathbf{n}(t+1) = \mathbf{A}[\boldsymbol{\theta}, \mathbf{n}(t), \mathbf{g}(t)]\,\mathbf{n}(t) \tag{10.44}$$

$$\mathbf{g}(t) = \mathbf{g}[\boldsymbol{\theta}, \mathbf{n}(t)] \tag{10.45}$$

where $\mathbf{g}(t)$ is a vector (of dimension $q \times 1$) describing the ecological or economic aspects of the environment on which the vital rates depend. As written here, the model admits the possibility that the vital rates in $\mathbf{A}$ might depend directly on $\mathbf{n}$ as well as on the environment.

At equilibrium

$$\hat{\mathbf{n}} = \mathbf{A}[\boldsymbol{\theta}, \hat{\mathbf{n}}, \hat{\mathbf{g}}]\hat{\mathbf{n}} \tag{10.46}$$

$$\hat{\mathbf{g}} = \mathbf{g}[\boldsymbol{\theta}, \hat{\mathbf{n}}]. \tag{10.47}$$

Differentiating these expressions gives

$$d\hat{\mathbf{n}} = \mathbf{A}(d\hat{\mathbf{n}}) + (d\mathbf{A})\hat{\mathbf{n}} \tag{10.48}$$

$$d\hat{\mathbf{g}} = \frac{\partial \hat{\mathbf{g}}}{\partial \boldsymbol{\theta}^{\mathsf{T}}} d\boldsymbol{\theta} + \frac{\partial \hat{\mathbf{g}}}{\partial \mathbf{n}} d\hat{\mathbf{n}}. \tag{10.49}$$

Applying the vec operator to (10.48) and expanding $d\mathrm{vec}\,\mathbf{A}$ gives

$$d\hat{\mathbf{n}} = \left(\hat{\mathbf{n}}^{\mathsf{T}} \otimes \mathbf{I}_s\right)\left[\frac{\partial \mathrm{vec}\,\mathbf{A}}{\partial \boldsymbol{\theta}^{\mathsf{T}}} d\boldsymbol{\theta} + \frac{\partial \mathrm{vec}\,\mathbf{A}}{\partial \mathbf{g}^{\mathsf{T}}} d\hat{\mathbf{g}}\right] + \mathbf{A} d\hat{\mathbf{n}}. \tag{10.50}$$

Substituting (10.49) for $d\hat{\mathbf{g}}$ and rearranging gives

$$d\hat{\mathbf{n}} = \left(\hat{\mathbf{n}}^{\mathsf{T}} \otimes \mathbf{I}_s\right)\left[\frac{\partial \mathrm{vec}\,\mathbf{A}}{\partial \boldsymbol{\theta}^{\mathsf{T}}} + \frac{\partial \mathrm{vec}\,\mathbf{A}}{\partial \mathbf{g}^{\mathsf{T}}} \frac{\partial \hat{\mathbf{g}}}{\partial \boldsymbol{\theta}^{\mathsf{T}}}\right] d\boldsymbol{\theta}$$

$$+ \left[\mathbf{A} + \left(\hat{\mathbf{n}}^{\mathsf{T}} \otimes \mathbf{I}_s\right)\frac{\partial \mathrm{vec}\,\mathbf{A}}{\partial \mathbf{g}^{\mathsf{T}}} \frac{\partial \hat{\mathbf{g}}}{\partial \mathbf{n}^{\mathsf{T}}}\right] d\hat{\mathbf{n}}. \tag{10.51}$$

Solving for $d\hat{\mathbf{n}}$ and applying the identification theorem yields

$$\frac{d\hat{\mathbf{n}}}{d\boldsymbol{\theta}^{\mathsf{T}}} = \left[\mathbf{I}_s - \mathbf{A} - \left(\hat{\mathbf{n}}^{\mathsf{T}} \otimes \mathbf{I}_s\right)\frac{\partial \mathrm{vec}\,\mathbf{A}}{\partial \mathbf{g}^{\mathsf{T}}} \frac{\partial \hat{\mathbf{g}}}{\partial \mathbf{n}^{\mathsf{T}}}\right]^{-1}$$

$$\times \left(\hat{\mathbf{n}} \otimes \mathbf{I}_s\right)\left[\frac{\partial \mathrm{vec}\,\mathbf{A}}{\partial \boldsymbol{\theta}^{\mathsf{T}}} + \frac{\partial \mathrm{vec}\,\mathbf{A}}{\partial \mathbf{g}^{\mathsf{T}}} \frac{\partial \hat{\mathbf{g}}}{\partial \boldsymbol{\theta}^{\mathsf{T}}}\right]. \tag{10.52}$$

In this expansion, $\mathbf{A}$, $\mathbf{g}$, and all derivatives are evaluated at $(\hat{\mathbf{n}}, \hat{\mathbf{g}})$. A comparison of (10.52) with (10.16) shows that including the feedback mechanism has simply written $d\mathrm{vec}\,\mathbf{A}/d\mathbf{n}^{\mathsf{T}}$ and $d\mathrm{vec}\,\mathbf{A}/d\boldsymbol{\theta}^{\mathsf{T}}$ in terms of $\mathbf{g}$ using the chain rule.

The environmental variable $\mathbf{g}$ may be of interest in its own right (e.g., in the food ratio model of Lee and Tuljapurkar (2008), in which it is a measure of well-being, measured in terms of food per individual). The sensitivity of $\hat{\mathbf{g}}$ at equilibrium is

$$\frac{d\hat{\mathbf{g}}}{d\boldsymbol{\theta}^{\mathsf{T}}} = \frac{\partial \hat{\mathbf{g}}}{\partial \boldsymbol{\theta}^{\mathsf{T}}} + \frac{\partial \hat{\mathbf{g}}}{\partial \mathbf{n}} \frac{d\hat{\mathbf{n}}}{d\boldsymbol{\theta}^{\mathsf{T}}} \tag{10.53}$$

where $d\hat{\mathbf{g}}/d\boldsymbol{\theta}^{\mathsf{T}}$ is given by (10.49) and $(d\hat{\mathbf{n}}/d\boldsymbol{\theta}^{\mathsf{T}})$ by (10.52).

10.4 Subsidized Populations and Competition for Space

A subsidized population is one in which new individuals are recruited from elsewhere rather than (or in addition to) being generated by local reproduction. Subsidy is important in many plant and animal populations, especially of benthic marine invertebrates and fish. Many of these species produce planktonic larvae that may disperse very long distances (Scheltema 1971) before they settle and become sessile for the rest of their lives. Thus a significant part—maybe even all—of the recruitment at any location is independent of local fertility (e.g., Almany et al. 2007). Subsidized models have been used to analyze conservation programs in which captive-reared animals are released into a wild or re-established population (Sarrazin and Legendre 2000). They have been applied to the demography of human organizations; e.g., schools, businesses, learned societies (Gani 1963; Pollard 1968; Bartholomew 1982). They are also the basis of cohort-component population projections that include immigration.

In the simplest subsidized models, both local demography and recruitment are density-independent. Alternatively, recruitment may depend on some resource (e.g., space) whose availability depends on the local population, or the local demography after settlement may be density-dependent. All three cases can lead to equilibrium populations.

10.4.1 Density-Independent Subsidized Populations

The model,

$$\mathbf{n}(t + 1) = \mathbf{A}[\boldsymbol{\theta}]\mathbf{n}(t) + \mathbf{b}[\boldsymbol{\theta}], \tag{10.54}$$

includes a subsidy vector $\mathbf{b}$ giving the input of individuals to the population.[8] The parameters $\boldsymbol{\theta}$ may affect $\mathbf{A}$ or $\mathbf{b}$, or both. If the fertility appearing in $\mathbf{A}$ is below replacement, so that $\lambda < 1$, then a stable equilibrium $\hat{\mathbf{n}}$ exists.[9] This equilibrium satisfies

$$\hat{\mathbf{n}} = \mathbf{A}\hat{\mathbf{n}} + \mathbf{b} \qquad (10.55)$$

$$= (\mathbf{I}_s - \mathbf{A})^{-1}\,\mathbf{b}. \qquad (10.56)$$

Differentiating (10.55) and applying the vec operator yields

$$d\hat{\mathbf{n}} = \left(\hat{\mathbf{n}}^{\mathsf{T}} \otimes \mathbf{I}_s\right) d\operatorname{vec} \mathbf{A} + \mathbf{A}\left(d\hat{\mathbf{n}}\right) + d\mathbf{b} \qquad (10.57)$$

Solving for $d\hat{\mathbf{n}}$ and applying the chain rule gives the sensitivity of the equilibrium,

$$\frac{d\hat{\mathbf{n}}}{d\boldsymbol{\theta}^{\mathsf{T}}} = (\mathbf{I}_s - \mathbf{A})^{-1}\left\{\left(\hat{\mathbf{n}}^{\mathsf{T}} \otimes \mathbf{I}_s\right)\frac{d\operatorname{vec} \mathbf{A}}{d\boldsymbol{\theta}^{\mathsf{T}}} + \frac{d\mathbf{b}}{d\boldsymbol{\theta}^{\mathsf{T}}}\right\}. \qquad (10.58)$$

Example 3: The Australian Academy of Sciences Most human organizations are subsidized; recruits (new students in a school, new employees in a company) come from outside, not from local reproduction. In an early example of a subsidized population model, Pollard (1968) analyzed the age structure of the Australian Academy of Sciences, recruitment to which takes place by election.[10] The Academy had been founded in 1954, and between 1955 and 1963 had elected about 6 new Fellows each year, with an age distribution (Pollard 1968, Table 2) given by

Age	Percent
30–34	0.0
35–39	12.2
40–44	24.5
45–49	26.5
50–54	20.4
55–59	4.1
60–64	10.2
65–69	2.0

---

[8]The same model could describe harvest if $\mathbf{b} \le 0$ (e.g., Hauser et al. 2006). This form of harvest produces unstable equilibria, and is not considered further here.

[9]If $\lambda > 1$, the population grows exponentially and the subsidy eventually becomes negligible. The equilibrium in this case is non-positive (and hence biologically irrelevant) and unstable. If $\lambda = 1$ then the population would remain constant in the absence of subsidy; any non-zero subsidy will then lead to unbounded population growth.

[10]Pollard's paper is remarkable for its treatment of both deterministic and stochastic models, but here I consider only the deterministic case.

Pollard interpolated this distribution to 1-year age classes, and combined it with a 1954 life table for Australian males (only one woman, the redoubtable geologist Dorothy Hill in 1956, had been elected to the Academy prior to 1969) to construct a model of the form (10.54). He calculated the equilibrium size and age composition of the Academy. Here, I have used the male life table for Australia 1953–1955 in Keyfitz and Flieger (1968, p. 558) to construct an age-classified matrix $\mathbf{A}$ with age-specific probabilities of survival P_i on its subdiagonal and zeros elsewhere. Were these vital rates and the age distribution of the subsidy vector to remain constant, the Academy would reach an equilibrium size of $\hat{N} = 149.5$ with an age distribution $\hat{\mathbf{n}}$ shown in Fig. 10.2a.

As parameters, consider the age-specific mortality rates $\mu_i = -\log P_i$, and define the parameter vector $\boldsymbol{\theta} = \begin{pmatrix} \mu_1 & \mu_2 & \dots \end{pmatrix}^{\mathsf{T}}$. Equation (10.58) then gives the sensitivity of the equilibrium population to changes in age-specific mortality. The
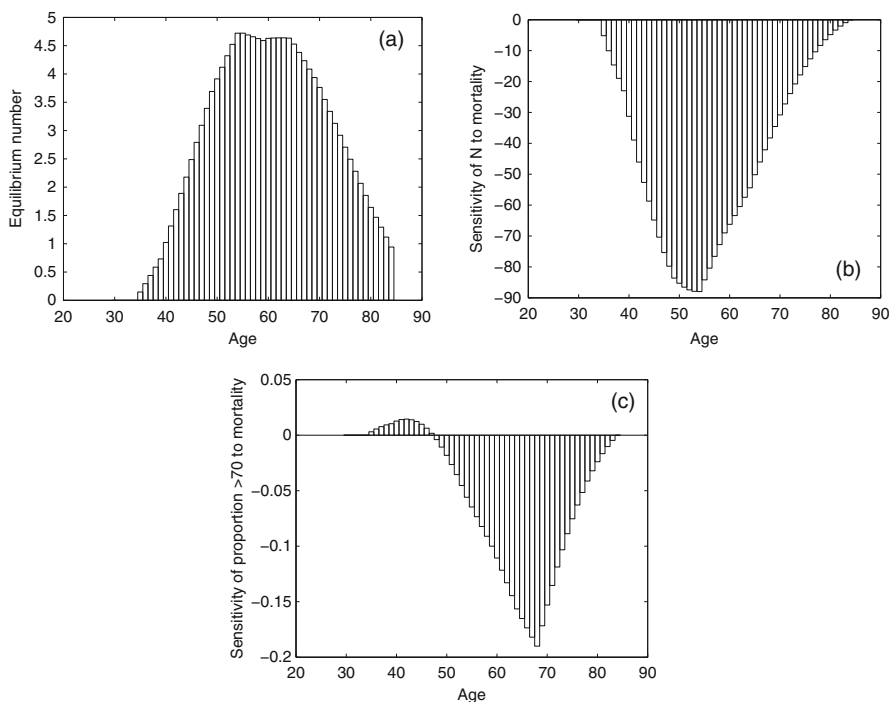
Fig. 10.2 Analysis of the equilibrium of a linear subsidized model for the Australian Academy of Science, based on Pollard (1968). (**a**) The equilibrium age structure of the Academy, assuming recruitment of 6 members per year. (**b**) The sensitivity, to changes in age-specific mortality, of the number of members. (**c**) The sensitivity, to changes in age-specific mortality, of the proportion of members over 70 years old

sensitivity of the total size of the Academy, $\hat{N} = \mathbf{1}^\mathsf{T}\hat{\mathbf{n}}$, calculated using (10.31), is shown in Fig. 10.2b. It shows that increases in mortality reduce $\hat{N}$ (not surprising), with the greatest effect coming from changes in mortality at ages 48–58.

Because learned societies are often concerned with their age distributions, Pollard (1968) examined the proportion of members over age 70. At equilibrium, this proportion is $\hat{R} = 0.26$. The sensitivity $d\hat{R}/d\boldsymbol{\theta}^\mathsf{T}$, calculated using (10.33), is shown in Fig. 10.2c. Increases in mortality before age 48 would increase the proportion of members over 70, while increases in mortality after age 48 would decrease it.[11]   ■

10.4.2   Linear Subsidized Models with Competition for Space

Recruitment in subsidized populations may be limited by the availability of a resource. Roughgarden et al. (1985; see also Pascual and Caswell 1991) presented a model for a population of sessile organisms, such as barnacles, in which recruitment is limited by available space. Barnacles[12] produce larvae that disperse in the plankton for several weeks before settling onto a rock surface or other suitable substrate, after which they no longer move.

Roughgarden's model supposes that settlement is proportional to the free space $F(t)$. Thus the subsidy vector is

$$\mathbf{b}(t) = \big(\phi F(t)\ 0 \cdots 0\big)^\mathsf{T}, \tag{10.59}$$

where ϕ is the settlement rate per unit of free space, and is determined by the pool of available larvae. The free space is the difference between the total area A and the space occupied by the population,

$$F(t) = A - \mathbf{g}^\mathsf{T}\mathbf{n}(t) \tag{10.60}$$

where $\mathbf{g}$ is a vector of stage-specific basal areas. Suppose that all locally-produced larvae are advected away, so that the first row of $\mathbf{A}$ is zero. Then, substituting (10.60) into (10.59) and rearranging gives

$$\mathbf{n}(t+1) = \mathbf{B}\mathbf{n}(t) + \big(\phi A\ 0 \cdots 0\big)^\mathsf{T} \tag{10.61}$$

---

[11]It is possible to calculate the average age of the Academy, and its sensitivity, using results to be introduced in Sect. 10.5.4. The response is very similar to that of the proportion over age 70.

[12]The temptation to draw analogies between barnacles and the members of learned academies is almost irresistible.

where

$$\mathbf{B} = \begin{pmatrix} -\phi g_1 & -\phi g_2 & \cdots & -\phi g_s \\ a_{21} & a_{22} & \cdots & a_{2s} \\ \vdots & \vdots & \ddots & \vdots \\ a_{s1} & a_{s2} & \cdots & a_{ss} \end{pmatrix}. \tag{10.62}$$

Although it includes competition for space, the model is linear. The equilibrium $\hat{\mathbf{n}}$ of (10.61) is stable if the spectral radius of $\mathbf{B}$ is less than one.[13] The formula (10.58) gives the sensitivity of this equilibrium to changes in the vital rates, the settlement rate, or the individual growth rate. This model might apply to any situation where the recruitment of new individuals depends on the availability of a resource (space, jobs, housing) that can be monopolized by residents.

Example 4: Intertidal barnacles Gaines and Roughgarden (1985) modelled a population of the barnacle *Balanus glandula* in central California. In one site (denoted KLM in their paper), they reported age-independent survival with a probability of $P_i = 0.985$ per week, $i = 1, \ldots, 52$. The growth in basal area of an individual barnacle could be described by $g_x = \pi(\rho x)^2$, where x is age in weeks and ρ is the radial growth rate ($\rho = 0.0041$ cm/wk). The mean settlement rate was $\phi = 0.107$. The matrix $\mathbf{B}$ contains survival probabilities P_i on the subdiagonal, terms of the form $-\phi g_i$ in the first row, and zeros elsewhere.

The equilibrium population $\hat{\mathbf{n}}$ has an exponential age distribution (Fig. 10.3a). It is scaled here relative to total area, so $A = 1$. The equilibrium proportion of free space is $\hat{F} = 0.865$.

To calculate sensitivities, let the parameters be age-specific survival probabilities, so that $\boldsymbol{\theta} = \begin{pmatrix} P_1 & \cdots & P_{52} \end{pmatrix}$. Some of the possible sensitivities are shown in Fig. 10.3. Increasing survival at age j (ages $j = 10, 20, 40$ are shown) reduces the abundance of ages younger than j and increases the abundance of ages older than j (Fig. 10.3b). A perturbation to a parameter, call it ξ, that affects survival at all ages would have the effect

$$\frac{d\hat{\mathbf{n}}}{d\xi} = \frac{d\hat{\mathbf{n}}}{d\boldsymbol{\theta}^{\mathsf{T}}} \frac{d\boldsymbol{\theta}}{d\xi} = \frac{d\hat{\mathbf{n}}}{d\boldsymbol{\theta}} \mathbf{1} \tag{10.63}$$

where $\mathbf{1}$ is a vector of ones. An increase in overall survival would reduce the abundance of age classes 1–6 and increase the abundance of older age classes (Fig. 10.3c).

---

[13]Because $\mathbf{B}$ contains negative elements, its dominant eigenvalue may be complex or negative, leading to oscillatory approach to the equilibrium.

The sensitivity of $\hat{\mathbf{n}}$ to the larval settlement rate ϕ is obtained from (10.58) by setting $d\mathrm{vec}\,\mathbf{B}/d\phi = \mathbf{0}_{s^2 \times 1}$, and

$$\frac{d\mathbf{b}}{d\phi} = \left(\hat{F}\ 0\ \cdots\ 0 \right)^{\mathsf{T}}$$

Not surprisingly, increases in ϕ increase $\hat{\mathbf{n}}$, with the largest effect on the young age classes (Fig. 10.3d). The sensitivity of $\hat{\mathbf{n}}$ to the radial growth rate ρ is obtained by writing

$$\frac{d\mathrm{vec}\,\mathbf{B}}{d\rho} = \frac{d\mathrm{vec}\,\mathbf{B}}{d\mathbf{g}^{\mathsf{T}}}\frac{d\mathbf{g}}{d\rho} \tag{10.64}$$

This sensitivity is negative, with the greatest impact on young age classes (Fig. 10.3e).
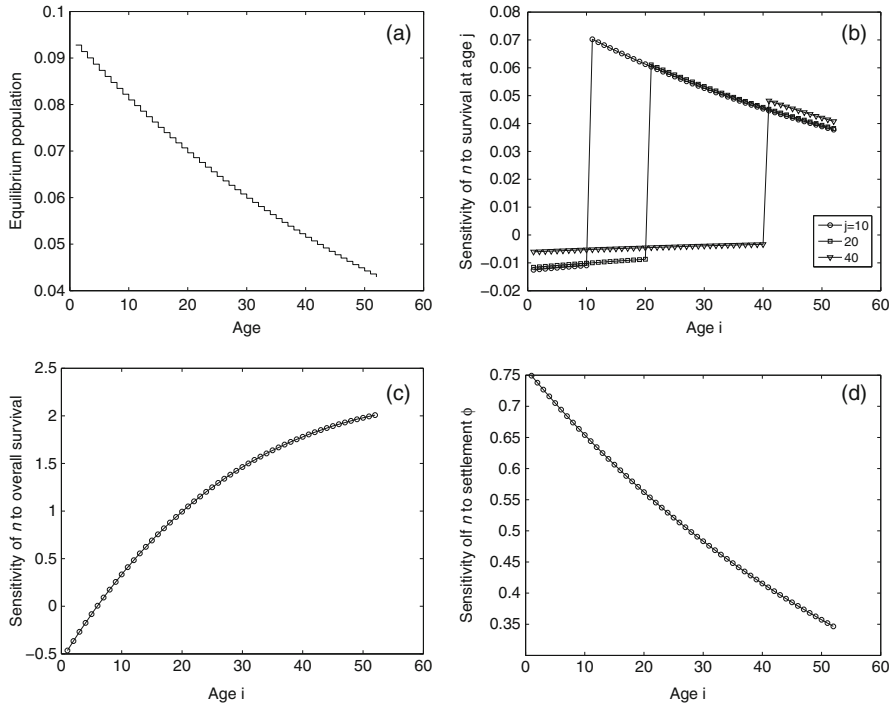
Fig. 10.3 Sensitivity analysis of a subsidized population of the intertidal barnacle *Balanus glandula*. (**a**) The equilibrium population $\hat{\mathbf{n}}$ (scaled relative to a unit of area $A = 1$). (**b**) The sensitivity of $\hat{b}on$ to a change in survival at ages $j = 10, 20, 40$. (**c**) The sensitivity of $\hat{\mathbf{n}}$ to changes in overall survival at all ages. (**d**) The sensitivity of $\hat{\mathbf{n}}$ to the settlement rate ϕ per unit area. A sensitivity analysis of a subsidized population of the intertidal barnacle *Balanus glandula*.
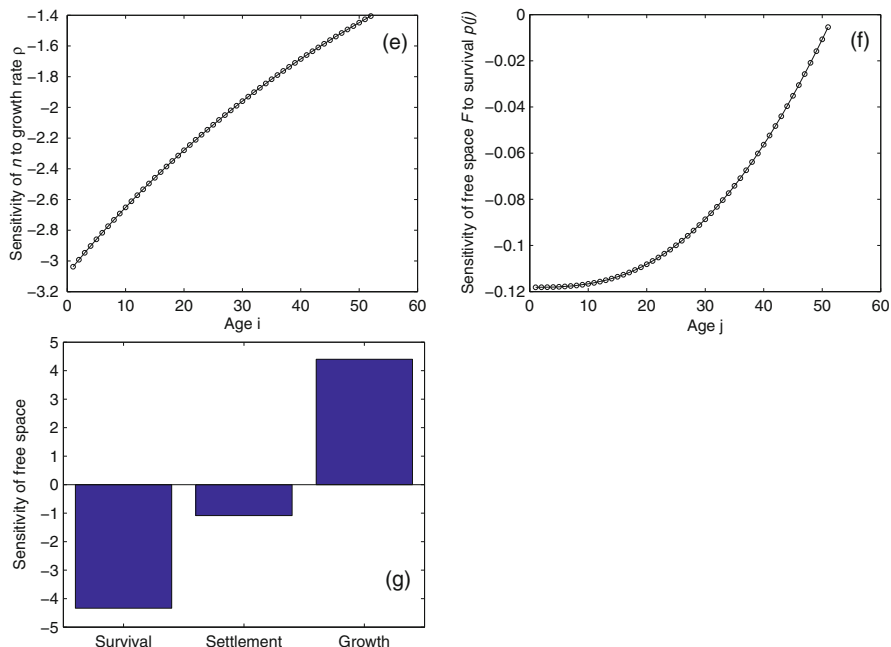
Fig. 10.3 (continued) (**e**) The sensitivity of $\hat{\mathbf{n}}$ to the radial growth rate ρ. (**f**) The sensitivity of the equilibrium free space $\hat{F}$ to age-specific survival. (**g**) The sensitivity of $\hat{F}$ to changes in overall survival, settlement rate, and radial growth rate. Based on data of Gaines and Roughgarden (1985)

Finally, the sensitivity of the equilibrium free space is given by

$$\frac{d\hat{F}}{d\boldsymbol{\theta}^{\mathsf{T}}} = \frac{d\hat{F}}{d\mathbf{n}^{\mathsf{T}}}\frac{d\hat{\mathbf{n}}}{d\boldsymbol{\theta}^{\mathsf{T}}} = -\mathbf{g}^{\mathsf{T}}\frac{d\hat{\mathbf{n}}}{d\boldsymbol{\theta}^{\mathsf{T}}} \tag{10.65}$$

Increases in survival reduce the amount of free space at equilibrium; the effect is largest for changes in survival of young age classes (Fig. 10.3f). Figure 10.3g compares the effect on $\hat{F}$ of changes in overall survival, settlement, and radial growth rate. It is not surprising that increases in survival or settlement will reduce free space, but perhaps surprising that increases in the radial growth rate actually increase $\hat{F}$. ∎

10.4.3   Density-Dependent Subsidized Models

Once individuals arrive in the population, they may experience a variety of density-dependent effects, that can be incorporated in a model

$$\mathbf{n}(t+1) = \mathbf{A}\left[\boldsymbol{\theta}, \mathbf{n}(t)\right]\mathbf{n}(t) + \mathbf{b}. \tag{10.66}$$

The sensitivity result (10.58) applies to this model by substituting

$$d\text{vec}\,\mathbf{A} = \frac{\partial \text{vec}\,\mathbf{A}}{\partial \boldsymbol{\theta}^{\mathsf{T}}} d\boldsymbol{\theta} + \frac{\partial \text{vec}\,\mathbf{A}}{\partial \mathbf{n}^{\mathsf{T}}} d\hat{\mathbf{n}} \tag{10.67}$$

into (10.57) and solving for $d\hat{\mathbf{n}}$, to obtain

$$\frac{d\hat{\mathbf{n}}}{d\boldsymbol{\theta}^{\mathsf{T}}} = \left(\mathbf{I}_s - \mathbf{A} - \left(\hat{\mathbf{n}}^{\mathsf{T}} \otimes \mathbf{I}_s \right) \frac{\partial \text{vec}\,\mathbf{A}}{\partial \mathbf{n}^{\mathsf{T}}} \right)^{-1} \left\{ \left(\hat{\mathbf{n}}^{\mathsf{T}} \otimes \mathbf{I}_s \right) \frac{\partial \text{vec}\,\mathbf{A}}{\partial \boldsymbol{\theta}^{\mathsf{T}}} + \frac{d\mathbf{b}}{d\boldsymbol{\theta}^{\mathsf{T}}} \right\}.$$
$$\tag{10.68}$$

where $\mathbf{A}$, $\mathbf{b}$, and all derivatives of $\mathbf{A}$ and $\mathbf{b}$ are evaluated at $\hat{\mathbf{n}}$.

10.5   Stable Structure and Reproductive Value

The linear model $\mathbf{n}(t+1) = \mathbf{A}\mathbf{n}(t)$ will, if $\mathbf{A}$ is primitive, converge to a stable age or stage distribution. But while the dynamics of the population vector $\mathbf{n}(t)$ are linear, the dynamics of the *proportional* population structure are nonlinear (Tuljapurkar 1997). We can take advantage of this to analyze the sensitivity of proportional structures by writing them as equilibria of nonlinear maps.

10.5.1   Stable Structure

The sensitivity of the stable stage distribution has been approached as an eigenvector perturbation problem (e.g., Caswell 1982, 2001; Kirkland and Neumann 1994), but those calculations are complicated. Analysis of the equilibrium of the nonlinear model (10.69) is much easier.

Let $\mathbf{p}$ denote the proportional stage structure vector ($\mathbf{p} \geq 0$, $\mathbf{1}^{\mathsf{T}}\mathbf{p} = 1$). The dynamics of $\mathbf{p}(t)$ satisfy

$$\mathbf{p}(t+1) = \frac{\mathbf{A}\mathbf{p}(t)}{\|\mathbf{A}\mathbf{p}(t)\|}. \tag{10.69}$$

The stable stage distribution is an equilibrium of (10.69); it satisfies

$$\hat{\mathbf{p}} = \frac{\mathbf{A}\hat{\mathbf{p}}}{\mathbf{1}^{\mathsf{T}}\mathbf{A}\hat{\mathbf{p}}} \tag{10.70}$$

where the 1-norm can be replaced by $\mathbf{1}^{\mathsf{T}}\mathbf{A}\hat{\mathbf{p}}$ because $\hat{\mathbf{p}}$ is non-negative. Differentiating both sides gives

$$d\hat{\mathbf{p}} = \frac{1}{\left(\mathbf{1}^{\mathsf{T}}\mathbf{A}\hat{\mathbf{p}}\right)^2} \left[\mathbf{1}^{\mathsf{T}}\mathbf{A}\hat{\mathbf{p}}(d\mathbf{A})\hat{\mathbf{p}} + \mathbf{1}^{\mathsf{T}}\mathbf{A}\hat{\mathbf{p}}\mathbf{A}(d\hat{\mathbf{p}}) - \mathbf{A}\hat{\mathbf{p}}\mathbf{1}^{\mathsf{T}}(d\mathbf{A})\hat{\mathbf{p}} - \mathbf{A}\hat{\mathbf{p}}\mathbf{1}^{\mathsf{T}}\mathbf{A}(d\hat{\mathbf{p}})\right]$$
(10.71)

Note that $\mathbf{A}\hat{\mathbf{p}} = \lambda\hat{\mathbf{p}}$ and $\mathbf{1}^{\mathsf{T}}\mathbf{A}\hat{\mathbf{p}} = \lambda$, where λ is the dominant eigenvalue of $\mathbf{A}$. Making these substitutions and applying the vec operator to both sides gives

$$\lambda\, d\hat{\mathbf{p}} = \left[\left(\hat{\mathbf{p}}^{\mathsf{T}} \otimes \mathbf{I}_s\right) - \left(\hat{\mathbf{p}}^{\mathsf{T}} \otimes \hat{\mathbf{p}}\mathbf{1}^{\mathsf{T}}\right)\right] d\mathrm{vec}\, \mathbf{A} + \left[\mathbf{A} - \hat{\mathbf{p}}\mathbf{1}^{\mathsf{T}}\mathbf{A}\right] d\hat{\mathbf{p}}$$
(10.72)

Solving for $d\hat{\mathbf{p}}$ and applying the chain rule gives

$$\frac{d\hat{\mathbf{p}}}{d\boldsymbol{\theta}^{\mathsf{T}}} = \left(\lambda\mathbf{I}_s - \mathbf{A} + \hat{\mathbf{p}}\mathbf{1}^{\mathsf{T}}\mathbf{A}\right)^{-1} \left(\hat{\mathbf{p}}^{\mathsf{T}} \otimes \mathbf{I}_s - \hat{\mathbf{p}}^{\mathsf{T}} \otimes \hat{\mathbf{p}}\mathbf{1}^{\mathsf{T}}\right) \frac{d\mathrm{vec}\, \mathbf{A}}{d\boldsymbol{\theta}^{\mathsf{T}}}$$
(10.73)

Example 5: A human age distribution As an example, consider the age distribution of the population of the United States in 1985 (data from Keyfitz and Flieger 1990). These vital rates yield a declining population ($\lambda = 0.975$) and an age distribution skewed towards older ages (Fig. 10.4). Applying (10.73) yields the sensitivity of $\hat{\mathbf{p}}$ to age-specific survival probabilities P_i and fertilities F_i, where age classes $i = 1, \ldots, 18$ correspond to ages 0–5, ..., 85–90. The overall patterns are familiar from previous sensitivity analyses of stable age distributions (e.g., Caswell 2001, Figure 9.22). Increasing survival probability at a given age increases the relative abundance of the next several age classes, at the expense of younger and older age classes. Increasing fertility at a given age increases the abundance of young age classes at the expense of older age classes. ∎

10.5.2   Reproductive Value

A similar approach gives the sensitivity of the reproductive value vector $\mathbf{v}$, given by the left eigenvector of $\mathbf{A}$ corresponding to λ. Reproductive value is customarily scaled so that $v_1 = 1$. Scaled in this way, $\mathbf{v}$ satisfies

$$\hat{\mathbf{v}}^{\mathsf{T}} = \frac{\hat{\mathbf{v}}^{\mathsf{T}}\mathbf{A}}{\hat{\mathbf{v}}^{\mathsf{T}}\mathbf{A}\mathbf{e}_1}$$
(10.74)

where $\mathbf{e}_1$ is a vector with 1 in the first entry and zeros elsewhere. Differentiating both sides gives

$$d\hat{\mathbf{v}}^{\mathsf{T}} = \frac{1}{\left(\hat{\mathbf{v}}^{\mathsf{T}}\mathbf{A}\mathbf{e}_1\right)^2} \left[\hat{\mathbf{v}}^{\mathsf{T}}\mathbf{A}\mathbf{e}_1(d\hat{\mathbf{v}}^{\mathsf{T}})\mathbf{A} + \hat{\mathbf{v}}^{\mathsf{T}}\mathbf{A}\mathbf{e}_1\hat{\mathbf{v}}^{\mathsf{T}}(d\mathbf{A})\right.$$
$$\left. - (d\hat{\mathbf{v}}^{\mathsf{T}})\mathbf{A}\mathbf{e}_1\hat{\mathbf{v}}^{\mathsf{T}}\mathbf{A} - \hat{\mathbf{v}}^{\mathsf{T}}(d\mathbf{A})\mathbf{e}_1\hat{\mathbf{v}}^{\mathsf{T}}\mathbf{A}\right]$$
(10.75)
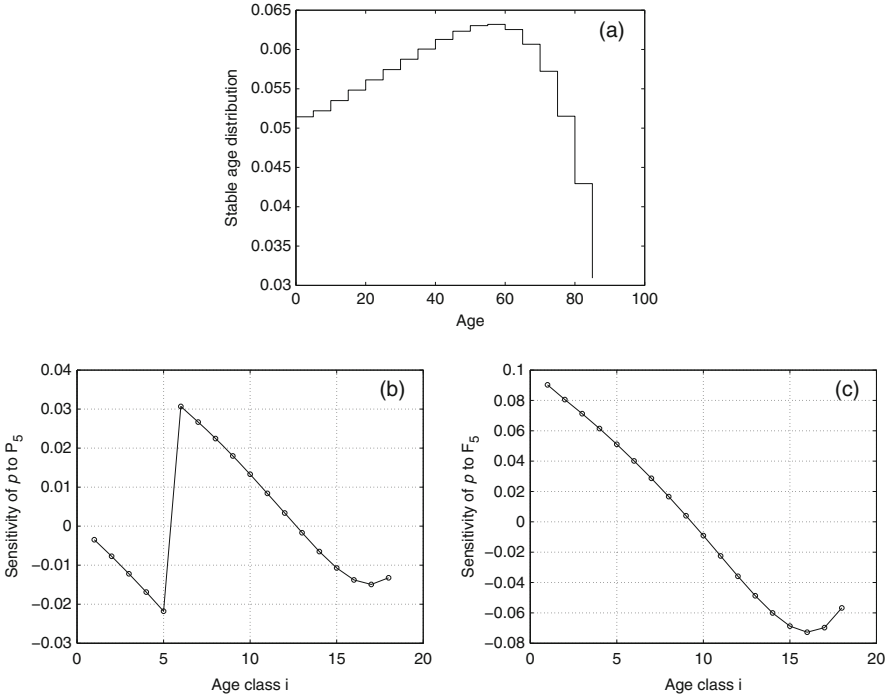
Fig. 10.4 Stable age distribution and sensitivity of stable age distribution to age-specific survival and fertility. (**a**) The stable age distribution. (**b**) The sensitivity of the stable age distribution to changes in survival (P_5) in age class 5. (**c**) Sensitivity of the stable age distribution to changes in fertility (F_5) in age class 5. Based on life table data for the United States in 1985 (Keyfitz and Flieger 1990)

But $\hat{\mathbf{v}}^\mathsf{T}\mathbf{A} = \lambda\hat{\mathbf{v}}^\mathsf{T}$ and $\hat{\mathbf{v}}^\mathsf{T}\mathbf{A}\mathbf{e}_1 = \lambda$. Making these substitutions and applying the vec operator (remembering that vec $\mathbf{v}^\mathsf{T} = \mathbf{v}$) gives

$$\lambda d\mathbf{v} = \left[\left(\mathbf{I}_s \otimes \hat{\mathbf{v}}^\mathsf{T}\right) - \left(\hat{\mathbf{v}}\mathbf{e}_1^\mathsf{T} \otimes \hat{\mathbf{v}}^\mathsf{T}\right)\right] d\text{vec }\mathbf{A} + \left(\mathbf{A}^\mathsf{T} - \hat{\mathbf{v}}\mathbf{e}_1^\mathsf{T}\mathbf{A}^\mathsf{T}\right) d\mathbf{v}. \qquad (10.76)$$

Solving for $d\mathbf{v}$ and using the chain rule gives

$$\frac{d\hat{\mathbf{v}}}{d\boldsymbol{\theta}^\mathsf{T}} = \left(\lambda\mathbf{I}_s - \mathbf{A}^\mathsf{T} + \hat{\mathbf{v}}\mathbf{e}_1^\mathsf{T}\mathbf{A}^\mathsf{T}\right)^{-1} \left[\left(\mathbf{I}_s \otimes \hat{\mathbf{v}}^\mathsf{T}\right) - \left(\hat{\mathbf{v}}\mathbf{e}_1^\mathsf{T} \otimes \hat{\mathbf{v}}^\mathsf{T}\right)\right] \frac{d\text{vec }\mathbf{A}}{d\boldsymbol{\theta}^\mathsf{T}} \qquad (10.77)$$

In stable population theory, in the calculation of second derivatives of population growth rate (Shyu and Caswell 2014), and in the analysis of multitype branching processes for demographic stochasticity (Caswell and Vindenes 2018), it is necessary to use the sensitivity of $\mathbf{v}$ subject to the scaling

$$\mathbf{v}^\mathsf{T}\mathbf{w} = 1. \qquad (10.78)$$

The resulting derivative is

$$\frac{d\mathbf{v}}{d\boldsymbol{\theta}^{\mathsf{T}}} = \left(\lambda\mathbf{I} - \mathbf{A}^{\mathsf{T}} + \lambda\mathbf{v}\mathbf{w}^{\mathsf{T}}\right)^{-1}$$

$$\times \left(\left[\left(\mathbf{I} - \mathbf{v}\mathbf{w}^{\mathsf{T}}\right) \otimes \mathbf{v}^{\mathsf{T}}\right] - \lambda\left(\mathbf{v} \otimes \mathbf{v}^{\mathsf{T}}\right)\frac{d\mathbf{w}}{d\mathrm{vec}^{\mathsf{T}}\mathbf{A}}\right)\frac{d\mathrm{vec}\,\mathbf{A}}{d\boldsymbol{\theta}^{\mathsf{T}}}. \quad (10.79)$$

(see Caswell and Vindenes 2018 for derivation).

10.5.3 Sensitivity of the Dependency Ratio

The dependency ratio characterizes an age distribution by the relative abundance of two groups, one assumed to be dependent and the other productive (Keyfitz and Flieger 1990, p. 32). It is often assumed that persons younger than 15 or older than 65 are dependent on productive individuals between 15 and 65. The dependency ratio is defined as

$$D = \frac{\mathbf{a}^{\mathsf{T}}\hat{\mathbf{p}}}{\mathbf{b}^{\mathsf{T}}\hat{\mathbf{p}}} \quad (10.80)$$

where $\mathbf{a}$ is a vector with ones for the dependent ages and zeros otherwise, and $\mathbf{b}$ is its complement. Applying Eq. (10.33) for the sensitivity of a ratio gives

$$\frac{dD}{d\boldsymbol{\theta}^{\mathsf{T}}} = \left(\frac{\mathbf{b}^{\mathsf{T}}\hat{\mathbf{p}}\mathbf{a}^{\mathsf{T}} - \mathbf{a}^{\mathsf{T}}\hat{\mathbf{p}}\mathbf{b}^{\mathsf{T}}}{\left(\mathbf{b}^{\mathsf{T}}\hat{\mathbf{p}}\right)^2}\right)\frac{d\hat{\mathbf{p}}}{d\boldsymbol{\theta}^{\mathsf{T}}}. \quad (10.81)$$

where $d\hat{\mathbf{p}}/\boldsymbol{\theta}^{\mathsf{T}}$ is given by (10.73).

   This result can be generalized in several ways. The analysis may be performed separately for the dependent young and the dependent old, by suitable modification of $\mathbf{a}$ and $\mathbf{b}$. These two components are likely to be influenced by different demographic factors and can respond to perturbations in opposite directions. The 0-1 vectors $\mathbf{a}$ and $\mathbf{b}$ may be replaced by vectors of weights; e.g., age-specific consumption and age-specific income (Fürnkranz-Prskawetz and Sambt 2014). For an example applied to a population projection for Spain, see Caswell and Sanchez Gassen (2015). The analysis also applies to stage-classified models, provided that dependent and independent stages can be identified. It also applies to nonlinear models, with the stable stage distribution $\hat{\mathbf{p}}$ replaced by the equilibrium population $\hat{\mathbf{n}}$ in (10.81). It can be extended to transient dynamics, where the age distribution, and thus the dependency ratio, varies over time (Caswell 2007), as is the case in population projections (Caswell and Sanchez Gassen 2015). Finally, the sensitivity (10.81) makes it possible to carry out LTRE analyses to decompose

differences in dependency ratios into components due to differences in each of the vital rates (see Chaps. 2, 8, and 9).

Example 5: (cont'd) Dependency ratios in human populations The United States in 1985 had a set of vital rates leading to a low growth rate ($\lambda = 0.975$), and a relatively low dependency ratio, dominated by the old. Kuwait in 1970, in contrast, had a high growth rate ($\lambda = 1.210$) and one of the highest dependency ratios listed in the compilation of Keyfitz and Flieger (1990), dominated by the young:

	U.S.A. 1985	Kuwait 1970
D	0.668	1.025
D_y	0.260	0.956
D_o	0.406	0.069

where D_y and D_o are the dependency ratios calculated for the young and old separately. The sensitivities of D, D_y, and D_o to changes in age-specific survival and fertility are shown in Fig. 10.5. The responses of D to changes in the vital rates differ between the two countries. In the U.S., increases in fertility would reduce D. In Kuwait, increases in fertility (especially at young ages) would increase D. In the U.S., increases in survival[14] before age 30 reduce D; increases after age 30 increase D. In Kuwait, increases in survival, except at very young and very old ages, reduce D.

Breaking D into its young and old components helps to explain these differences. In both countries, there is a crossover in survival effects. Increases in survival at early ages increase D_y and reduce D_o. At later ages, increases in survival reduce D_y and increase D_o. Increases in fertility increase D_y and reduce D_o. In the U.S. population, both these effects are large, with the negative effect on D_o larger than the positive effect on D_y. In the Kuwaiti population, the positive effect on D_y is much greater than the negative effect on D_o.   ∎

10.5.4   Sensitivity of Mean Age and Related Quantities

From an age distribution $\hat{\mathbf{p}}$, it is possible to compute the mean age of any age-specific property (e.g., production of children, collection of retirement benefits, exposure to toxic chemicals); see Chu (1998, p. 26) for general discussions. The most familiar of these is the mean age of reproduction, which is one measure of generation time (Coale 1972).

Let $\mathbf{f}$ be a vector of age-specific per-capita fertilities. The age distribution of offspring production is then $\mathbf{f} \circ \hat{\mathbf{p}}$, where $\circ$ is the Hadamard, or element-by-element product. The mean age of the mothers of these offspring is obtained by normalizing $\mathbf{f} \circ \hat{\mathbf{p}}$ to sum to 1 and taking the mean over the resulting distribution,

---

[14]Or, equivalently, reductions in mortality. For these parameter values, the sensitivity to mortality is approximately the sensitivity to survival with the opposite sign.
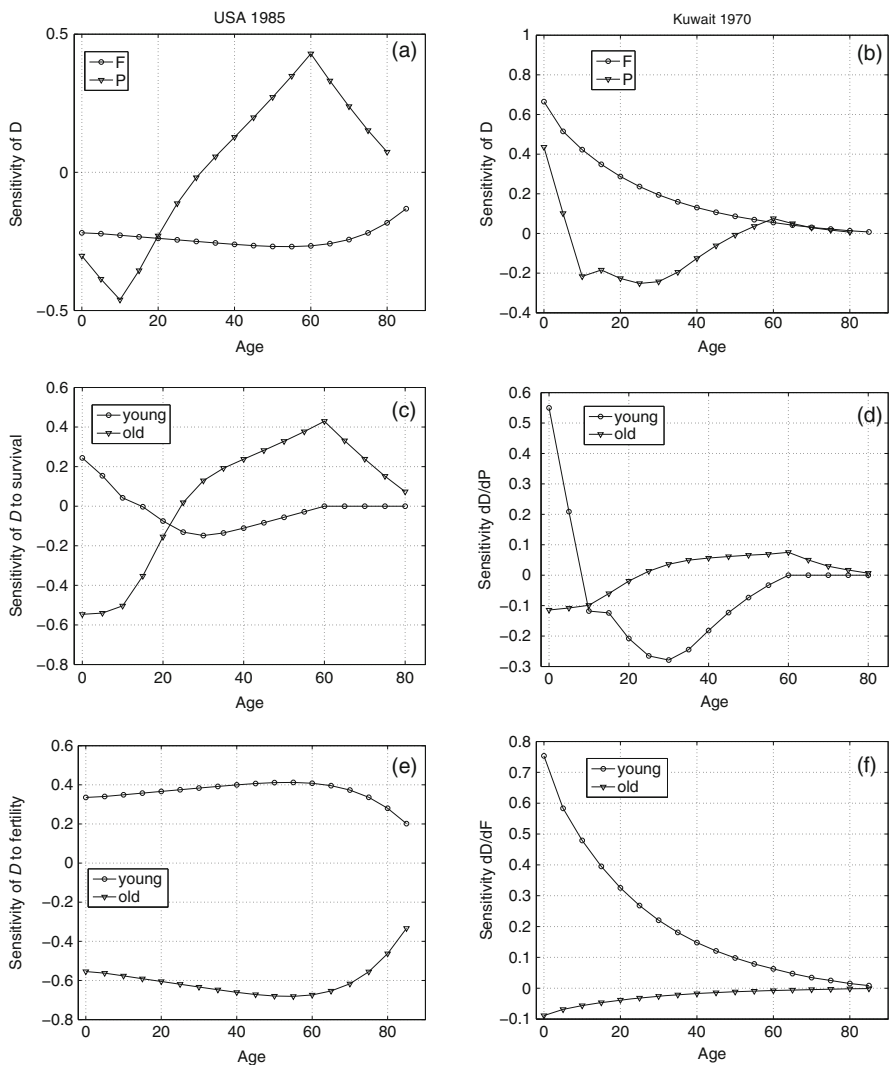
Fig. 10.5 Sensitivity of the dependency ratio D, and of its old and young components, to age-specific survival and fertility. Left: calculated from the stable age distribution of the United States in 1985. Right: calculated from the stable age distribution of Kuwait in 1970. (**a**) and (**b**): Sensitivity of D to survival (P_i) and fertility (F_i). (**c**) and (**d**): Sensitivity of the components of D to survival. (**e**) and (**f**): Sensitivity of the components of D to fertility. Life table data from Keyfitz and Flieger (1990)

$$\bar{a}_{\mathbf{f}} = \frac{\mathbf{c}^{\mathsf{T}} \left(\mathbf{f} \circ \hat{\mathbf{p}} \right)}{\mathbf{1}^{\mathsf{T}} \left(\mathbf{f} \circ \hat{\mathbf{p}} \right)} \tag{10.82}$$

where

$$\mathbf{c}^{\mathsf{T}} = \left(1 \; 2 \; \cdots \; s \right),$$

with s as the last age class.

Now differentiate $\bar{a}_{\mathbf{f}}$, following the now-familiar rules for ratios. The differential of the Hadamard product of two vectors is $d(\mathbf{a} \circ \mathbf{b}) = \mathcal{D}(\mathbf{a})d\mathbf{b} + \mathcal{D}(\mathbf{b})d\mathbf{a}$. The result is

$$\frac{d\bar{a}_{\mathbf{f}}}{d\boldsymbol{\theta}^{\mathsf{T}}} = \left(\frac{\mathbf{1}^{\mathsf{T}} \left(\mathbf{f} \circ \hat{\mathbf{p}} \right) \mathbf{c}^{\mathsf{T}} - \mathbf{c}^{\mathsf{T}} \left(\mathbf{f} \circ \hat{\mathbf{p}} \right) \mathbf{1}^{\mathsf{T}}}{\left(\mathbf{f}^{\mathsf{T}} \hat{\mathbf{p}} \right)^2} \right) \left(\mathcal{D}(\mathbf{f}) \frac{d\hat{\mathbf{p}}}{d\boldsymbol{\theta}^{\mathsf{T}}} + \mathcal{D}(\hat{\mathbf{p}}) \frac{d\mathbf{f}}{d\boldsymbol{\theta}^{\mathsf{T}}} \right) \tag{10.83}$$

where $d\hat{\mathbf{p}}/d\boldsymbol{\theta}^{\mathsf{T}}$ is given by (10.73).

This result can be generalized in several ways. Setting $\mathbf{f} = \mathbf{1}$ makes the age-specific property that of simply being alive, and $\bar{a}_1 = \mathbf{c}^{\mathsf{T}} \mathbf{1}$ is then the mean age of the stable population, the sensitivity of which is

$$\frac{d\bar{a}}{d\boldsymbol{\theta}^{\mathsf{T}}} = \mathbf{c}^{\mathsf{T}} \frac{d\hat{\mathbf{p}}}{d\boldsymbol{\theta}^{\mathsf{T}}} \tag{10.84}$$

The calculations can also be applied to the equilibrium population in a nonlinear model by substituting $\hat{\mathbf{n}}$ for $\hat{\mathbf{p}}$. They apply directly to stage-classified models with stages defined on an interval scale (e.g., size classes), in which case they give, e.g., the mean size at reproduction. If the stages are not evenly spaced, then $\mathbf{c}$ would be replaced by

$$\mathbf{c}^{\mathsf{T}} = \left(x_1 \; x_2 \; \cdots \; x_s \right) \tag{10.85}$$

where x_i is the value associated with stage i.

Example 5: (cont'd) Mean age of reproduction The mean age of reproduction in the stable age distribution of the United States in 1985 was $\bar{a}_{\mathbf{f}} = 24.02$ years (using the mid-points of the 5-year age intervals as the measure of age). The sensitivities of $\bar{a}_{\mathbf{f}}$ to changes in age-specific survival and fertility are shown in Fig. 10.6. Increases in survival prior to age 15 reduce $\bar{a}_{\mathbf{f}}$. Increases in survival after age 45 have almost no effect on $\bar{a}_{\mathbf{f}}$, because fertility is essentially zero after this age. Between age 15 and age 45, increases in survival increase the mean age of reproduction.

Increases in fertility reduce $\bar{a}_{\mathbf{f}}$ if they happen before age 25 and increase $\bar{a}_{\mathbf{f}}$ if they happen after age 25. These sensitivities are quite large, although this is somewhat irrelevant since the largest sensitivities are for ages at which fertility is zero and unlikely to be modified. ∎
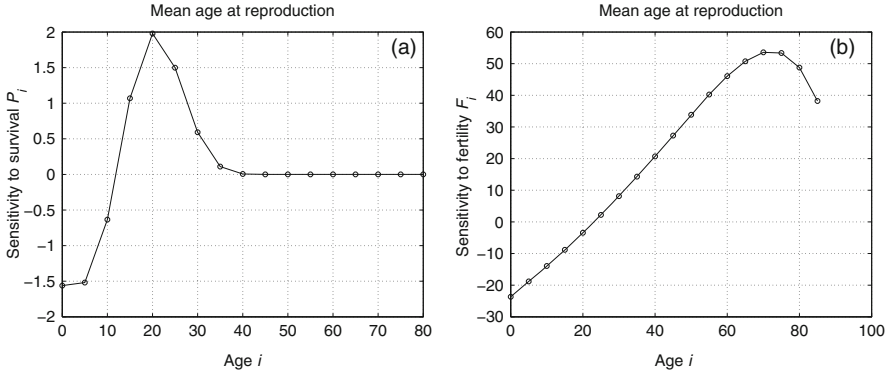
Fig. 10.6 Sensitivity of the mean age at reproduction to changes in age-specific survival and fertility, for the life table of the population of the United States, 1985. (Data from Keyfitz and Flieger 1990)

10.5.5 Sensitivity of Variance in Age

We can also calculate the sensitivity of the higher moments. For example, the variance in the age at reproduction is

$$V_{\mathbf{f}} = \overline{a_{\mathbf{f}}^2} - (\bar{a}_{\mathbf{f}})^2 \,. \tag{10.86}$$

This variance might, for example, be useful as a measure of the extent of iteroparity. The sensitivity of $V_{\mathbf{f}}$ to changes in parameters is obtained by writing the first term as

$$\overline{a_{\mathbf{f}}^2} = \frac{(\mathbf{c} \circ \mathbf{c})^{\mathsf{T}} \left(\mathbf{f} \circ \hat{\mathbf{p}}\right)}{\mathbf{1}^{\mathsf{T}} \left(\mathbf{f} \circ \hat{\mathbf{p}}\right)} \tag{10.87}$$

and then differentiating

$$dV_{\mathbf{f}} = d\left(\overline{a_{\mathbf{f}}^2}\right) - 2\bar{a}_{\mathbf{f}} \ (d\bar{a}_{\mathbf{f}}) \,. \tag{10.88}$$

The final result is

$$\frac{dV_{\mathbf{f}}}{d\boldsymbol{\theta}^{\mathsf{T}}} = \left(\frac{\mathbf{1}^{\mathsf{T}}(\mathbf{f} \circ \hat{\mathbf{p}})(\mathbf{c} \circ \mathbf{c})^{\mathsf{T}} - (\mathbf{c} \circ \mathbf{c})^{\mathsf{T}} \left(\mathbf{f} \circ \hat{\mathbf{p}}\right) \mathbf{1}^{\mathsf{T}}}{\left(\mathbf{f}^{\mathsf{T}}\hat{\mathbf{p}}\right)^2}\right)$$

$$\times \left(\mathcal{D}\left(\mathbf{f}\right)\frac{d\hat{\mathbf{p}}}{d\boldsymbol{\theta}^{\mathsf{T}}} + \mathcal{D}\left(\hat{\mathbf{p}}\right)\frac{d\mathbf{f}}{d\boldsymbol{\theta}^{\mathsf{T}}}\right) - 2\bar{a}_{\mathbf{f}}\frac{d\bar{a}_{\mathbf{f}}}{d\boldsymbol{\theta}^{\mathsf{T}}}. \tag{10.89}$$

where $d\hat{\mathbf{p}}/d\boldsymbol{\theta}^{\mathsf{T}}$ is given by (10.73) and $d\bar{a}_{\mathbf{f}}/d\boldsymbol{\theta}^{\mathsf{T}}$ is given by (10.83).

10.6   Frequency-Dependent Two-Sex Models

In sexually reproducing species, a particular sort of nonlinearity arises from the dependence of reproduction on the relative abundance of males and females. This dependence is captured in a marriage function or mating rule (e.g., McFarland 1972; Pollak 1987, 1990) When the vital rates depend only on the relative, rather than the absolute, abundance of males and females, then $\mathbf{A}[\boldsymbol{\theta}, \mathbf{n}]$ is homogeneous of degree 0 in $\mathbf{n}$; i.e.,

$$\mathbf{A}[\boldsymbol{\theta}, c\mathbf{n}] = \mathbf{A}[\boldsymbol{\theta}, \mathbf{n}] \qquad \text{for any } c \neq 0. \tag{10.90}$$

Such models have been called frequency-dependent (Caswell and Weeks 1986; Caswell 2001) to distinguish them from density-dependent nonlinear models that do not have this homogeneity property.

Because of the homogeneity of $\mathbf{A}[\boldsymbol{\theta}, \mathbf{n}]$, frequency-dependent models do not converge to an equilibrium density $\hat{\mathbf{n}}$. Instead, there may exist[15] a stable equilibrium proportional structure $\hat{\mathbf{p}}$ to which the population will converge, at which point it grows exponentially at a rate λ given by the dominant eigenvalue of $\mathbf{A}[\boldsymbol{\theta}, \hat{\mathbf{p}}]$. Thus sensitivity analysis of two-sex models must include both the population structure and the population growth rate.

Note that matrix models that include Mendelian genetics are also homogeneous of degree zero, but it is confusing to call them frequency-dependent, because doing so creates confusion with the genetic phenomenon of frequency-dependent fitness, which is a different thing altogether (de Vries and Caswell 2018).

10.6.1   Sensitivity of the Population Structure

The equilibrium proportional population structure $\hat{\mathbf{p}}$ satisfies

$$\hat{\mathbf{p}} = \frac{\mathbf{A}[\boldsymbol{\theta}, \hat{\mathbf{p}}] \, \hat{\mathbf{p}}}{\|\mathbf{A}[\boldsymbol{\theta}, \hat{\mathbf{p}}] \, \hat{\mathbf{p}}\|} \tag{10.91}$$

where $\hat{p}_i \geq 0$ and $\mathbf{1}^\mathsf{T} \hat{\mathbf{p}} = 1$. Differentiating (10.91) gives

$$d\hat{\mathbf{p}} = \frac{\mathbf{1}^\mathsf{T} \mathbf{A}\hat{\mathbf{p}} \Big[(d\mathbf{A})\hat{\mathbf{p}} + \mathbf{A}(d\hat{\mathbf{p}}) \Big] - \mathbf{A}\hat{\mathbf{p}} \Big[\mathbf{1}^\mathsf{T}(d\mathbf{A})\hat{\mathbf{p}} + \mathbf{1}^\mathsf{T}\mathbf{A}(d\hat{\mathbf{p}}) \Big]}{\left(\mathbf{1}^\mathsf{T}\mathbf{A}\hat{\mathbf{p}} \right)^2}. \tag{10.92}$$

---

[15]A sufficient, but not necessary, condition for the existence of an equilibrium is that $\mathbf{A}$ cannot map a nonzero vector $\mathbf{n}$ directly to zero; necessary conditions are more difficult (Nussbaum 1988, 1989). See also Martcheva (1999).

Making the substitutions $\mathbf{A}\hat{\mathbf{p}} = \lambda\hat{\mathbf{p}}$ and $\mathbf{1}^\mathsf{T}\mathbf{A}\hat{\mathbf{p}} = \lambda$ and rearranging gives

$$\lambda d\hat{\mathbf{p}} = (d\mathbf{A})\hat{\mathbf{p}} + \mathbf{A}(d\hat{\mathbf{p}}) - \hat{\mathbf{p}}\mathbf{1}^\mathsf{T}(d\mathbf{A})\hat{\mathbf{p}} - \hat{\mathbf{p}}\mathbf{1}^\mathsf{T}\mathbf{A}(d\hat{\mathbf{p}}). \qquad (10.93)$$

Applying the vec operator to both sides, expanding $d\mathrm{vec}\,\mathbf{A}$, invoking the chain rule, and solving for $d\hat{\mathbf{p}}/d\boldsymbol{\theta}^\mathsf{T}$ gives

$$\frac{d\hat{\mathbf{p}}}{d\boldsymbol{\theta}^\mathsf{T}} = \left[\lambda\mathbf{I}_s - \mathbf{A} + \hat{\mathbf{p}}\mathbf{1}^\mathsf{T}\mathbf{A} - \left[\hat{\mathbf{p}}^\mathsf{T} \otimes \left(\mathbf{I}_s - \hat{\mathbf{p}}\mathbf{1}^\mathsf{T}\right)\right]\frac{\partial\mathrm{vec}\,\mathbf{A}}{\partial\mathbf{p}^\mathsf{T}}\right]^{-1}$$
$$\times \left[\hat{\mathbf{p}}^\mathsf{T} \otimes \left(\mathbf{I}_s - \hat{\mathbf{p}}\mathbf{1}^\mathsf{T}\right)\right]\frac{\partial\mathrm{vec}\,\mathbf{A}}{\partial\boldsymbol{\theta}^\mathsf{T}} \qquad (10.94)$$

where $\mathbf{A}$ and all derivatives are evaluated at $\hat{\mathbf{p}}$. Note that (10.94) differs from the expression (10.73) for the stable stage distribution in the linear model only in the term involving $\partial\mathrm{vec}\,\mathbf{A}/\partial\mathbf{p}^\mathsf{T}$, which of course is zero in the linear model.

10.6.2 Population Growth Rate in Two-Sex Models

Because a population with the equilibrium structure grows exponentially, I once suggested treating $\mathbf{A}[\boldsymbol{\theta}, \hat{\mathbf{p}}]$ as a constant matrix and applying eigenvalue sensitivity analysis to it, in order to examine life history evolution in 2-sex models (Caswell 2001, p. 577). This was incorrect, because it ignored the effect of parameter changes on $\mathbf{A}$ through their effects on the equilibrium $\hat{\mathbf{p}}$. A correct calculation obtains the sensitivity of λ including effects of parameters on both $\mathbf{A}$ and $\hat{\mathbf{p}}$.

Note that $\hat{\mathbf{p}}$ is a right eigenvector of $\mathbf{A}[\boldsymbol{\theta}, \hat{\mathbf{p}}]$ corresponding to λ. Let $\mathbf{v}$ be the corresponding left eigenvector, where $\mathbf{v}^\mathsf{T}\mathbf{A}[\boldsymbol{\theta}, \hat{\mathbf{p}}] = \lambda\mathbf{v}^\mathsf{T}$ and $\mathbf{v}^\mathsf{T}\hat{\mathbf{p}} = 1$. Then

$$d\lambda = \mathbf{v}^\mathsf{T}(d\mathbf{A})\hat{\mathbf{p}} \qquad (10.95)$$

Caswell (1978). Apply the vec operator and Roth's theorem to get

$$d\lambda = \left(\hat{\mathbf{p}}^\mathsf{T} \otimes \mathbf{v}^\mathsf{T}\right) d\mathrm{vec}\,\mathbf{A}. \qquad (10.96)$$

Expanding $d\mathrm{vec}\,\mathbf{A}$ gives

$$\frac{d\lambda}{d\boldsymbol{\theta}^\mathsf{T}} = \left(\hat{\mathbf{p}}^\mathsf{T} \otimes \mathbf{v}^\mathsf{T}\right)\left[\frac{\partial\mathrm{vec}\,\mathbf{A}}{\partial\boldsymbol{\theta}^\mathsf{T}} + \frac{\partial\mathrm{vec}\,\mathbf{A}}{\partial\hat{\mathbf{p}}^\mathsf{T}}\frac{d\hat{\mathbf{p}}}{d\boldsymbol{\theta}^\mathsf{T}}\right] \qquad (10.97)$$

where $\mathbf{A}$, $\mathbf{v}$, and the derivatives of $\mathbf{A}$ are all evaluated at the equilibrium $\hat{\mathbf{p}}$, and $d\hat{\mathbf{p}}/d\boldsymbol{\theta}^\mathsf{T}$ is given by (10.94).

Note that λ is the invasion exponent for this model, and thus the sensitivity of λ to a parameter gives the selection gradient on that parameter. Tuljapurkar et al.

(2007) used this fact to explore the effect of male fertility patterns on the evolution
of aging; the sensitivity (10.97) could be used to generalize such results. Recent
work by Shyu has coupled these calculations to the methods of adaptive dynamics
to examine the evolution of sex ratios (Shyu and Caswell 2016a,b).

Although two-sex models are an important case of homogeneous models, they
are not the only case. Keyfitz's (1972) interpretation of the Easterlin hypothesis
describes fertility as dependent on only the relative, not absolute, size of a cohort. A
model based on this premise would be frequency-dependent (homogeneous) and
would lead to an exponentially growing population to which (10.97) would be
applicable.

Example 6: A two-sex model for passerine birds   Legendre et al. (1999) used a
frequency-dependent two-sex model to study the introductions of passerine birds
to New Zealand. The life cycle includes two age classes (first year and older) for
females and for males. The life cycle graph is shown in Fig. 10.7. The numbers of
females and males are $N_f = n_1 + n_2$ and $N_m = n_3 + n_4$, respectively.

Because passerines are typically monogamous within a breeding season, and
assuming that mating is indiscriminate with respect to age, Legendre et al. (1999)
used as a mating function

$$B(\mathbf{n}) = \min\left(N_f, N_m\right), \tag{10.98}$$

giving the number of matings as a function of the number of males and females. The
per-capita fertility of a female of age-class i is the number of matings divided by the
number of females and multiplied by the number of surviving offspring per mating.

$$F(\mathbf{n}) = \frac{\sigma_0 \phi_i B(\mathbf{n})}{N_f} \tag{10.99}$$

$$= \begin{cases} \sigma_0 \phi_i \frac{N_m}{N_f} & N_f \geq N_m \\ \sigma_0 \phi & N_f < N_m \end{cases} \tag{10.100}$$

where σ_0 is the probability of survival from fledging to age 1 and ϕ_i is the clutch size of age class i. When males are the scarcer sex (the avian equivalent of a marriage squeeze) fertility is proportional to the ratio of males to females. When females are the scarcer sex, all females are mated and fertility depends only on fecundity and neonatal survival.

Births are allocated to females and males according to a primary sex ratio ρ which gives the proportion female. The resulting two-sex projection matrix is

$$\mathbf{A[n]} = \begin{pmatrix} \rho F_1(\mathbf{n}) & \rho F_2(\mathbf{n}) & 0 & 0 \\ \sigma_1 & \sigma_2 & 0 & 0 \\ (1-\rho)F_1(\mathbf{n}) & (1-\rho)F_2(\mathbf{n}) & 0 & 0 \\ 0 & 0 & \sigma_3 & \sigma_4 \end{pmatrix} \tag{10.101}$$

Legendre et al. (1999) assigned typical values for passerine birds of $\sigma_0 = 0.2$, $\phi_i = 7$, and $\rho = 0.5$. They set male and female survival equal ($\sigma_1 = \sigma_3 = 0.35$, $\sigma_2 = \sigma_4 = 0.4$), but this is a pathological special case in this model, so instead I consider two cases, one in which male mortality is higher than female mortality, and one in which the difference is reversed.[16] The survival probabilities and equilibrium population structures are

$$\sigma = \begin{pmatrix} 0.35 \\ 0.5 \\ 0.25 \\ 0.4 \end{pmatrix} \qquad \hat{\mathbf{p}} = \begin{pmatrix} 0.320 \\ 0.226 \\ 0.320 \\ 0.134 \end{pmatrix} \tag{10.102}$$

$$\sigma = \begin{pmatrix} 0.25 \\ 0.4 \\ 0.35 \\ 0.5 \end{pmatrix} \qquad \hat{\mathbf{p}} = \begin{pmatrix} 0.320 \\ 0.134 \\ 0.320 \\ 0.226 \end{pmatrix} \tag{10.103}$$

The elasticities of $\hat{\mathbf{p}}$ to each of the parameters, calculated from (10.94), are shown in Table 10.1. Regardless of which sex is scarcer, increasing neonatal survival increases the proportion of young, at the expense of the proportion of adults, in both sexes. Increasing the sex ratio ρ increases the proportion of females at the expense of males. Increasing female survival (σ_1 or σ_2) increases the proportion of adult females at the expense of all other stages; increasing male survival has the opposite

---

[16]In a survey of the literature, adult mortality for female passerines exceeded that for males in 21 out of 28 cases (Promislow et al. 1992). Birds differ from mammals in this respect.

Table 10.1 Elasticity of $\hat{\mathbf{p}}$ to parameters in two-sex model for passerine birds, under two mortality scenarios. When male mortality is greater than female mortality, males are rarer than females and fertility at equilibrium is limited by the mating function. When male mortality is less than female mortality, females are rare and fertility is not affected by the mating function

Males rare								
Stage	σ_0	ρ	σ_1	σ_2	σ_3	σ_4	ϕ_1	ϕ_2
$\hat{p}_1$	0.455	0.453	−0.226	−0.229	0.000	0.000	0.266	0.189
$\hat{p}_2$	−0.890	1.799	0.774	0.783	−0.398	−0.268	−0.521	−0.369
$\hat{p}_3$	0.455	−1.547	−0.226	−0.229	0.000	0.000	0.266	0.189
$\hat{p}_4$	−0.664	−0.428	−0.226	−0.229	0.669	0.450	−0.389	−0.275
Females rare								
Stage	σ_0	ρ	σ_1	σ_2	σ_3	σ_4	ϕ_1	ϕ_2
$\hat{p}_1$	0.455	1.547	0.000	0.000	−0.226	−0.229	0.320	0.135
$\hat{p}_2$	−0.664	0.428	0.669	0.450	−0.226	−0.229	−0.467	−0.197
$\hat{p}_3$	0.455	−0.453	0.000	0.000	−0.226	−0.229	0.320	0.135
$\hat{p}_4$	−0.890	−1.799	−0.398	−0.268	0.774	0.783	−0.627	−0.264

effect. However, when females are rare, increasing female survival has no effect on the proportion of juveniles. When males are rare, increases in male survival have no effect on the proportion of juveniles. Increasing fecundity increases the proportion of juveniles, at the expense of adults, in both sexes and for either mortality pattern.

The elasticity of the population growth rate λ at equilibrium is shown in Table 10.2, and is compared to the naive calculation that treats $\mathbf{A}[\boldsymbol{\theta}, \hat{\mathbf{p}}]$ as a fixed matrix. When males are rare, so that fertility is limited by the mating function, the naive calculations are dramatically wrong. When calculated correctly, increases in the primary sex ratio ρ reduce λ, because they reduce the availability of males. Increases in female survival have no effect on λ, because the extra females produced have no opportunity to reproduce. Increases in male survival increase λ because they increase female fertility. In each case, the naive calculation leads, incorrectly, to the opposite conclusion.

When females are rare (which renders the model linear and female-dominant at equilibrium), the correct and the naive calculations agree. This is a consequence of using the minimum as a birth function. Some preliminary calculations using the harmonic mean birth function,

$$B(\mathbf{n}) = \frac{2N_f N_m}{N_f + N_m}, \tag{10.104}$$

in which both males and females influence fertility at all population structures, suggest that the naive elasticity calculations are always incorrect.

Sometimes the correct calculations lead to apparent paradoxes. Jenouvrier et al. (2010) developed a two-sex model for the Emperor penguin. It was a periodic model, with phases defined by events within the breeding cycle (cf. Chap. 8), and included a

Table 10.2 The elasticity of λ to parameters in the two-sex model for passerine birds, under two mortality scenarios. The correct calculation is based on (10.97). The naive calculation incorrectly treats $\mathbf{A}[\hat{\mathbf{p}}, \boldsymbol{\theta}]$ as a fixed matrix, ignoring the effect of parameters on the equilibrium population structure $\hat{\mathbf{p}}$

	Males rare		Females rare	
	Correct	Naive	Correct	Naive
σ_0	0.669	0.545	0.669	0.669
ρ	−0.669	0.545	0.669	0.669
σ_1	0	0.226	0.198	0.198
σ_2	0	0.229	0.133	0.133
σ_3	0.198	0	0	0
σ_4	0.133	0	0	0
ϕ_1	0.392	0.319	0.471	0.471
ϕ_2	0.277	0.226	0.198	0.198

mating function applied to adults at the breeding colony. Because Emperor penguins breed, and share parental care, in the midst of the Antarctic winter,[17] they must be strictly monogamous, and hence Jenouvrier used the minimum as a mating function.

Analysis of the equilibrium growth rate revealed that the sensitivity of λ to adult female survival was negative. This is impossible in a linear model, but happens in this frequency-dependent model because increasing adult female survival increases the proportion of females (already greater than the proportion of males) and thus decreases mating probability. The negative effect of reduced mating overwhelms the positive effect of improved adult survival; the net result is a reduction in population growth rate; see Jenouvrier et al. (2010) for details. ∎

10.6.3   The Birth Matrix-Mating Rule Model

Pollak (1987, 1990) introduced a powerful conceptual approach to two-sex models, which he called the birth matrix-mating rule (BMMR) model. This model separates the processes of mating, birth, and life cycle stage transitions, and treats them as a periodic process. When generalized to stage-structured models, it contains three main components:

1. A birth matrix whose entries give the expected number of male and female offspring produced by a mating of a male of age (or stage) i and a female of age j.
2. A mating rule function that gives the number of matings u_{ij} between males of age (or stage) i and females of age j.
3. A set of sex-specific mortality schedules, which project surviving individuals to the next age class, or, in our generalization, include other stage-specific life cycle transitions.

---

[17]Dramatically portrayed in the movie, *March of the Penguins*.

A matrix version of the BMMR has recently been developed, using a novel continuous-time formulation of periodic matrix models (Shyu and Caswell 2018). The mating, birth, and transition processes are described, respectively, by matrices **U**, **B**, and **T**. To explore the theoretical consequences of two-sex reproduction, the matrices are parameterized in terms of continuous-time rates rather than discrete-time probabilities. In continuous time, the periodic matrix product that would describe such a process in discrete time converges to a sum of the rate matrices. The dynamics of the population are given by

$$\frac{d\mathbf{n}(t)}{dt} = \mathbf{A}\left[\mathbf{n}(t)\right]\ \mathbf{n}(t) \tag{10.105}$$

where

$$\mathbf{A}\left[\mathbf{n}(t)\right] = \frac{1}{3}\left(\mathbf{T} + \mathbf{B} + \mathbf{U}[\mathbf{n}(t)]\right)\mathbf{n}(t) \tag{10.106}$$

That is, the projection matrix is the mean of the three component matrices, and is nonlinear because of the dependence of union formation (the matrix **U**) on **n**. Shyu and Caswell (2016a,b, 2018) explore this model in the context of sex ratio evolution and of sex-biased harvesting, deriving the sensitivity of the population growth rate as a measure of the selection gradient.

10.7   Sensitivity of Population Cycles

Equilibria are not the only attractors relevant in nature (e.g., Clutton-Brock et al. 1997) or the laboratory (Cushing et al. 2003). Cycles, invariant loops, and strange attractors also occur, and are sensitive to changes in parameters. This section examines the sensitivity of cycles.

10.7.1   Sensitivity of the Population Vector

A k-cycle is a sequence of population vectors $\hat{\mathbf{n}}_1, \ldots, \hat{\mathbf{n}}_k$, satisfying

$$\hat{\mathbf{n}}_{i+1} = \mathbf{A}\left[\boldsymbol{\theta}, \hat{\mathbf{n}}_i\right]\hat{\mathbf{n}}_i \qquad i = 1, \ldots, k-1$$

$$\hat{\mathbf{n}}_1 = \mathbf{A}\left[\boldsymbol{\theta}, \hat{\mathbf{n}}_k\right]\hat{\mathbf{n}}_k. \tag{10.107}$$

A change in parameters will modify each point in the cycle; the first goal of perturbation analysis is thus to find the sensitivities

$$\frac{d\hat{\mathbf{n}}_1}{d\boldsymbol{\theta}^\mathsf{T}}, \quad \ldots, \quad \frac{d\hat{\mathbf{n}}_k}{d\boldsymbol{\theta}^\mathsf{T}}. \tag{10.108}$$

The following is the derivation of these sensitivities for a 2-cycle. The extension to cycles of arbitrary length will follow. To simplify notation, define

$$\mathbf{A}_i \equiv \mathbf{A}\left[\boldsymbol{\theta}, \hat{\mathbf{n}}_i\right]. \tag{10.109}$$

The 2-cycle satisfies

$$\hat{\mathbf{n}}_1 = \mathbf{A}_2 \hat{\mathbf{n}}_2 \tag{10.110}$$

$$\hat{\mathbf{n}}_2 = \mathbf{A}_1 \hat{\mathbf{n}}_1 \tag{10.111}$$

Differentiating both equations, applying the vec operator, and expanding $d\text{vec}\,\mathbf{A}_i/d\boldsymbol{\theta}^\mathsf{T}$ yields a system of equations

$$\frac{d\hat{\mathbf{n}}_1}{d\boldsymbol{\theta}^\mathsf{T}} = \left(\hat{\mathbf{n}}_2^\mathsf{T} \otimes \mathbf{I}_s\right) \frac{\partial \text{vec}\,\mathbf{A}_2}{\partial \boldsymbol{\theta}^\mathsf{T}} + \left(\hat{\mathbf{n}}_2^\mathsf{T} \otimes \mathbf{I}_s\right) \frac{\partial \text{vec}\,\mathbf{A}_2}{\partial \mathbf{n}_2^\mathsf{T}} \left(\frac{d\hat{\mathbf{n}}_2}{d\boldsymbol{\theta}^\mathsf{T}}\right)$$

$$+\mathbf{A}_2 \left(\frac{d\hat{\mathbf{n}}_2}{d\boldsymbol{\theta}^\mathsf{T}}\right) \tag{10.112}$$

$$\frac{d\hat{\mathbf{n}}_2}{d\boldsymbol{\theta}^\mathsf{T}} = \left(\hat{\mathbf{n}}_1^\mathsf{T} \otimes \mathbf{I}_s\right) \frac{\partial \text{vec}\,\mathbf{A}_1}{\partial \boldsymbol{\theta}^\mathsf{T}} + \left(\hat{\mathbf{n}}_1^\mathsf{T} \otimes \mathbf{I}_s\right) \frac{\partial \text{vec}\,\mathbf{A}_1}{\partial \mathbf{n}_1^\mathsf{T}} \left(\frac{d\hat{\mathbf{n}}_1}{d\boldsymbol{\theta}^\mathsf{T}}\right)$$

$$+\mathbf{A}_1 \left(\frac{d\hat{\mathbf{n}}_1}{d\boldsymbol{\theta}^\mathsf{T}}\right) \tag{10.113}$$

This system can be written in block matrix form. Define $\mathbf{H}_i \equiv \hat{\mathbf{n}}_i^\mathsf{T} \otimes \mathbf{I}_s$. Then

$$\frac{d}{d\boldsymbol{\theta}^\mathsf{T}} \begin{pmatrix} \hat{\mathbf{n}}_1 \\ \hat{\mathbf{n}}_2 \end{pmatrix} = \begin{pmatrix} 0 & \mathbf{H}_2 \\ \mathbf{H}_1 & 0 \end{pmatrix} \begin{pmatrix} \dfrac{\partial \text{vec}\,\mathbf{A}_1}{\partial \boldsymbol{\theta}^\mathsf{T}} \\ \dfrac{\partial \text{vec}\,\mathbf{A}_2}{\partial \boldsymbol{\theta}^\mathsf{T}} \end{pmatrix}$$

$$+ \left[\begin{pmatrix} 0 & \mathbf{H}_2 \\ \mathbf{H}_1 & 0 \end{pmatrix} \begin{pmatrix} \dfrac{\partial \text{vec}\,\mathbf{A}_1}{\partial \mathbf{n}_1^\mathsf{T}} & 0 \\ 0 & \dfrac{\partial \text{vec}\,\mathbf{A}_2}{\partial \mathbf{n}_2^\mathsf{T}} \end{pmatrix} + \begin{pmatrix} 0 & \mathbf{A}_2 \\ \mathbf{A}_1 & 0 \end{pmatrix} \right]$$

$$\times \frac{d}{d\boldsymbol{\theta}^\mathsf{T}} \begin{pmatrix} \hat{\mathbf{n}}_1 \\ \hat{\mathbf{n}}_2 \end{pmatrix} \tag{10.114}$$

Solving for the sensitivities gives

$$
\frac{d}{d\boldsymbol{\theta}^{\mathsf{T}}}\begin{pmatrix}\hat{\mathbf{n}}_1\\\hat{\mathbf{n}}_2\end{pmatrix}=\left[\mathbf{I}_{2s}-\begin{pmatrix}0&\mathbf{H}_2\\\mathbf{H}_1&0\end{pmatrix}\begin{pmatrix}\dfrac{\partial\,\mathrm{vec}\,\mathbf{A}_1}{\partial\mathbf{n}_1^{\mathsf{T}}}&0\\0&\dfrac{\partial\,\mathrm{vec}\,\mathbf{A}_2}{\partial\mathbf{n}_2^{\mathsf{T}}}\end{pmatrix}\right.
$$

$$
\left.-\begin{pmatrix}0&\mathbf{A}_2\\\mathbf{A}_1&0\end{pmatrix}\right]^{-1}\begin{pmatrix}0&\mathbf{H}_2\\\mathbf{H}_1&0\end{pmatrix}\begin{pmatrix}\dfrac{\partial\,\mathrm{vec}\,\mathbf{A}_1}{\partial\boldsymbol{\theta}^{\mathsf{T}}}\\\dfrac{\partial\,\mathrm{vec}\,\mathbf{A}_2}{\partial\boldsymbol{\theta}^{\mathsf{T}}}\end{pmatrix}\tag{10.115}
$$

where the matrices $\mathbf{A}_i$ and the derivatives of $\mathbf{A}_i$ are all evaluated at $\hat{\mathbf{n}}_i$. The analogy with (10.16) is apparent.

This calculation can be extended to cycles of any period, in terms of block matrices as in (10.115). The pattern of the block matrices is clear from a 3-cycle. Define the following matrices:

$$
\mathbb{N}=\begin{pmatrix}\hat{\mathbf{n}}_1\\\hat{\mathbf{n}}_2\\\hat{\mathbf{n}}_3\end{pmatrix}\tag{10.116}
$$

$$
\mathbb{A}=\begin{pmatrix}0&0&\mathbf{A}_3\\\mathbf{A}_1&0&0\\0&\mathbf{A}_2&0\end{pmatrix}\tag{10.117}
$$

$$
\mathbb{H}=\begin{pmatrix}0&0&\mathbf{H}_3\\\mathbf{H}_1&0&0\\0&\mathbf{H}_2&0\end{pmatrix}\tag{10.118}
$$

$$
\mathbb{C}=\begin{pmatrix}\dfrac{\partial\,\mathrm{vec}\,\mathbf{A}_1}{\partial\mathbf{n}_1^{\mathsf{T}}}&0&0\\0&\dfrac{\partial\,\mathrm{vec}\,\mathbf{A}_2}{\partial\mathbf{n}_2^{\mathsf{T}}}&0\\0&0&\dfrac{\partial\,\mathrm{vec}\,\mathbf{A}_3}{\partial\mathbf{n}_3^{\mathsf{T}}}\end{pmatrix}\tag{10.119}
$$

$$\mathbb{D} = \begin{pmatrix} \dfrac{\partial \operatorname{vec} \mathbf{A}_1}{\partial \boldsymbol{\theta}^\mathsf{T}} \\[2ex] \dfrac{\partial \operatorname{vec} \mathbf{A}_1}{\partial \boldsymbol{\theta}^\mathsf{T}} \\[2ex] \dfrac{\partial \operatorname{vec} \mathbf{A}_1}{\partial \boldsymbol{\theta}^\mathsf{T}} \end{pmatrix}. \tag{10.120}$$

In terms of these matrices, the sensitivity of each point in the 3-cycle is given by

$$\frac{d\mathbb{N}}{d\boldsymbol{\theta}^\mathsf{T}} = [\mathbf{I}_{3s} - \mathbb{A} - \mathbb{H}\mathbb{C}]^{-1}\,\mathbb{H}\mathbb{D}. \tag{10.121}$$

10.7.2 Sensitivity of Weighted Densities and Time Averages

The matrix $d\mathbb{N}/d\boldsymbol{\theta}^\mathsf{T}$ contains the sensitivity of every stage to every parameter at every point in the cycle. This potential overload of information can be simplified by calculating the sensitivities of weighted densities and/or time averages over the cycle. To do this, it is convenient to write the points in the cycle as an array (of dimension $s \times k$)

$$\mathbf{G} = \begin{pmatrix} \hat{\mathbf{n}}_1 & \hat{\mathbf{n}}_2 & \cdots & \hat{\mathbf{n}}_k \end{pmatrix}. \tag{10.122}$$

The block vector $\mathbb{N}$ is

$$\mathbb{N} = \operatorname{vec} \mathbf{G}. \tag{10.123}$$

Weighted densities. Let $\mathbf{c}$ be a vector of weights, and let $\hat{N}_i = \mathbf{c}^\mathsf{T}\hat{\mathbf{n}}_i$ be the (scalar) weighted density at the ith point on the cycle. Then write

$$\hat{\mathbf{n}} = \begin{pmatrix} \hat{N}_1 \\ \vdots \\ \hat{N}_k \end{pmatrix} \tag{10.124}$$

The vector $\hat{\mathbf{n}}$ can be calculated from $\mathbb{N}$ as

$$\begin{aligned} \hat{\mathbf{n}} &= \begin{pmatrix} \mathbf{c}^\mathsf{T}\hat{\mathbf{n}}_1 & \cdots & \mathbf{c}^\mathsf{T}\hat{\mathbf{n}}_k \end{pmatrix}^\mathsf{T} \\ &= \operatorname{vec}\left(\mathbf{c}^\mathsf{T}\mathbf{G}\right) \\ &= \left(\mathbf{I}_k \otimes \mathbf{c}^\mathsf{T}\right)\operatorname{vec}\mathbf{G} \end{aligned}$$

$$= \left(\mathbf{I}_k \otimes \mathbf{c}^\mathsf{T} \right) \mathbb{N} \qquad \text{dimension} = k \times 1. \tag{10.125}$$

Time-averaged population vector.   Let $\mathbf{b}$ be a probability vector ($b_i \geq 0$, $\mathbf{1}^\mathsf{T}\mathbf{b} = 1$) and define the time-averaged population vector as

$$\bar{\mathbf{n}} = \sum_{i=1}^{k} b_i \hat{\mathbf{n}}_i. \tag{10.126}$$

Then

$$\begin{aligned}
\bar{\mathbf{n}} &= \mathbf{Gb} \\
&= \left(\mathbf{b}^\mathsf{T} \otimes \mathbf{I}_s \right) \text{vec } \mathbf{G} \\
&= \left(\mathbf{b}^\mathsf{T} \otimes \mathbf{I}_s \right) \mathbb{N} \qquad \text{dimension} = s \times 1 \tag{10.127}
\end{aligned}$$

Time-averaged weighted density.   Taking the time average of the $\hat{N}_i$ gives

$$\begin{aligned}
\bar{N} &= \sum_i b_i \mathbf{c}^\mathsf{T} \hat{\mathbf{n}}_i \\
&= \mathbf{c}^\mathsf{T} \mathbf{Gb} \\
&= \left(\mathbf{b}^\mathsf{T} \otimes \mathbf{c}^\mathsf{T} \right) \mathbb{N} \tag{10.128}
\end{aligned}$$

Thus the sensitivities of the weighted densities, the time-averaged population, and the time-averaged weighted density are obtained by differentiating (10.125), (10.127), and (10.128) as

$$\frac{d\hat{\mathbf{n}}}{d\boldsymbol{\theta}^\mathsf{T}} = \left(\mathbf{I}_k \otimes \mathbf{c}^\mathsf{T} \right) \frac{d\mathbb{N}}{d\boldsymbol{\theta}^\mathsf{T}} \tag{10.129}$$

$$\frac{d\bar{\mathbf{n}}}{d\boldsymbol{\theta}^\mathsf{T}} = \left(\mathbf{b}^\mathsf{T} \otimes \mathbf{I}_s \right) \frac{d\mathbb{N}}{d\boldsymbol{\theta}^\mathsf{T}} \tag{10.130}$$

$$\frac{d\bar{N}}{d\boldsymbol{\theta}^\mathsf{T}} = \left(\mathbf{b}^\mathsf{T} \otimes \mathbf{c}^\mathsf{T} \right) \frac{d\mathbb{N}}{d\boldsymbol{\theta}^\mathsf{T}} \tag{10.131}$$

where $d\mathbb{N}/\boldsymbol{\theta}^\mathsf{T}$ is given by (10.121).

Example 7 A 2-cycle in the *Tribolium* model A series of experiments on *Tribolium* reported by Dennis et al. (1995) produced stable 2-cycles by experimentally manipulating the adult mortality μ_a. Using the model in Example 2 and the estimated parameters

$$b = 11.677$$

$$c_{ea} = 1.100 \times 10^{-2}$$

$$c_{el} = 9.3 \times 10^{-3}$$

$$c_{pa} = 1.78 \times 10^{-2}$$

$$\mu_a = 1.108 \times 10^{-1}$$

$$\mu_l = 5.129 \times 10^{-1}$$

(Dennis et al. 1995, Table 1) leads to a 2-cycle

$$\hat{\mathbf{n}}_1 = \begin{pmatrix} 325.3 \\ 8.9 \\ 118.5 \end{pmatrix} \qquad \hat{\mathbf{n}}_2 = \begin{pmatrix} 18.2 \\ 158.4 \\ 106.4 \end{pmatrix}, \tag{10.132}$$

in which the population oscillates between a state dominated by larvae and adults and a state dominated by pupae and adults.

As an example of the rich sensitivity analyses possible for even such a simple model, consider the elasticity of the population vector $\hat{\mathbf{n}}_i$, of the total population $\hat{N}_i = \mathbf{1}^{\mathsf{T}}\hat{\mathbf{n}}_i$, of the total population respiration $\hat{R}_i = \mathbf{c}^{\mathsf{T}}\hat{\mathbf{n}}_i$ (with $\mathbf{c}$ the vector of stage-specific respiration rates from Example 2), and of the time averages $\bar{\mathbf{n}}$, $\bar{N}$, and $\bar{R}$. The results are collected in Fig. 10.8.

First, the elasticities of the $\hat{\mathbf{n}}_i$ differ from stage to stage and from one point on the cycle to another (Fig. 10.8a). Increases in fecundity, for example, increase the density of larvae and reduce the density of pupae in $\hat{\mathbf{n}}_1$, but have the opposite effects in $\hat{\mathbf{n}}_2$. The elasticities to b, c_{ea}, and c_{el} are much larger than those to the other parameters (cf. the elasticities of the equilibrium $\hat{n}$ in Fig. 10.1).

The elasticities of total population are similar at the two points in the cycle (Fig. 10.8b), except that larval mortality μ_l has a large negative effect on $\hat{N}_2$, but only a small effect on $\hat{N}_1$. The elasticities of total respiration $\hat{R}_i$, however, are different at the two points in the cycle (Fig. 10.8c).

The elasticities of the time-averaged population vector $\bar{\mathbf{n}}$ (Fig. 10.8d) are similar to those of the equilibrium vector in Fig. 10.1 (although they need not be). This pattern is not predictable from the patterns of the elasticities of the population vectors $\hat{\mathbf{n}}_1$ and $\hat{\mathbf{n}}_2$ (Fig. 10.8a).

Finally, the elasticities of the time averages, $\bar{N}$ and $\bar{R}$, of the weighted densities are similar to each other and to the elasticities of the time-averaged population $\bar{\mathbf{n}}$.

The sensitivity analysis of cycles thus depends very much on the dependent variables of interest. The matrix $d\mathbb{N}/d\boldsymbol{\theta}^{\mathsf{T}}$ (Fig. 10.8a) contains 36 pieces of information: the effects of 6 parameters on 3 stages at 2 points in the cycle. A focus on weighted density reduces this to 12 (Fig. 10.8b,c), but the results may depend very much on the particular weighting vector chosen. A focus on time averages reduces the information from 36 to 18 numbers (Fig. 10.8d), and the response of the time-averaged weighted densities finally are described by just 6 numbers. The good news
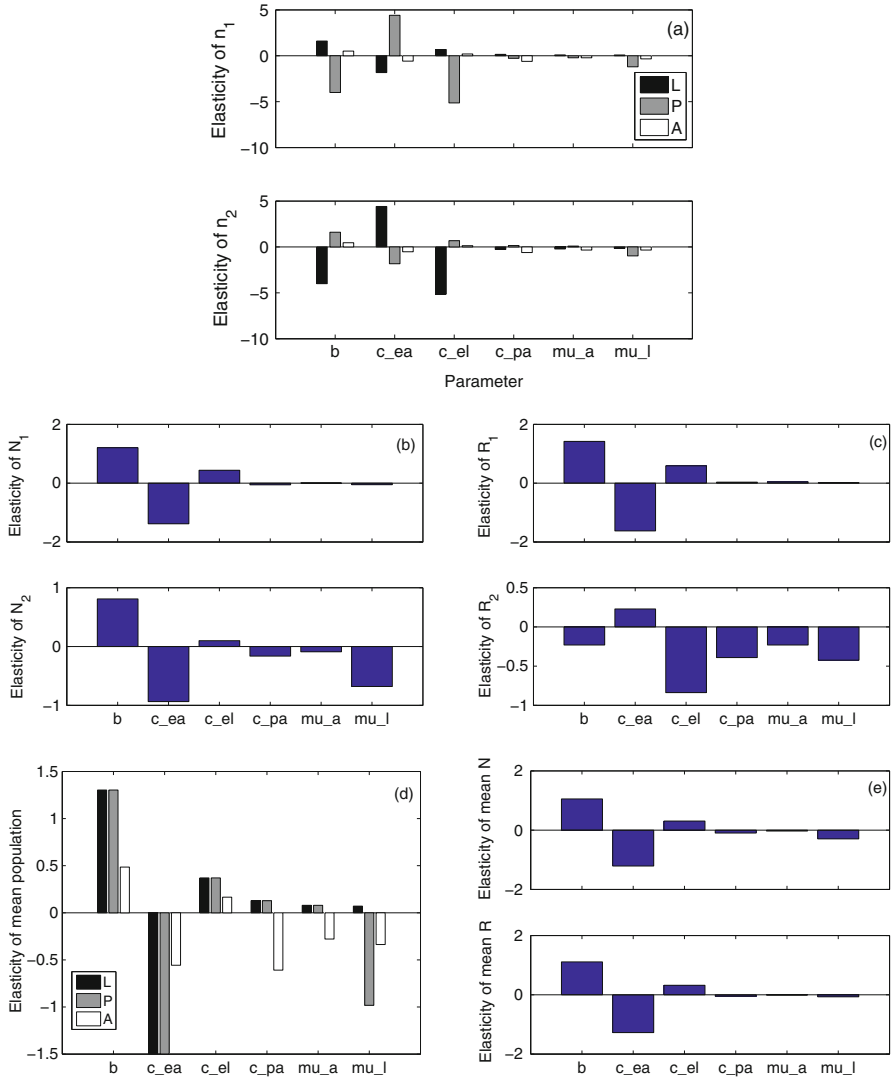
Fig. 10.8 Analysis of a 2-cycle in the *Tribolium* model. (**a**) Elasticity of the density of each stage, with respect to each parameter, at $\hat{\mathbf{n}}_1$ and $\hat{\mathbf{n}}_2$. (**b**) Elasticity of the total population $\hat{N}$ at each point in the cycle. (**c**) Elasticity of the total respiration $\hat{R}$ at each point in the cycle. (**d**) Elasticity of the time-averaged population $\bar{\mathbf{n}}$. (**e**) Elasticity of the time-averaged total population $\bar{N}$ and the time-averaged total respiration $\bar{R}$

is that Eqs. (10.121), (10.125), (10.127), and (10.128) make it easy to compute all these sensitivities. ∎

10.7.3   Sensitivity of Temporal Variance in Density

The variance over a cycle in a weighted density $\hat{N}$ can be written

$$V(\hat{N}) = E(\hat{N}^2) - \left[E(\hat{N})\right]^2 \tag{10.133}$$

where $E(\hat{N}) = \bar{N} = \mathbf{c}^{\mathsf{T}}\mathbf{Gb}$ and

$$E(\hat{N}^2) = \sum_{i=1}^{k} b_i \left(\mathbf{c}^{\mathsf{T}}\hat{\mathbf{n}}_i\right)^2 \tag{10.134}$$

$$= (\mathbf{c} \circ \mathbf{c})^{\mathsf{T}}(\mathbf{G} \circ \mathbf{G})\mathbf{b} \tag{10.135}$$

Taking the differential of $E(\hat{N}^2)$ and applying the vec operator gives

$$dE(\hat{N}^2) = 2\left[\mathbf{b}^{\mathsf{T}} \otimes (\mathbf{c} \circ \mathbf{c})^{\mathsf{T}}\right]\mathcal{D}\,(\mathbb{N})\,d\mathbb{N}. \tag{10.136}$$

Combining this with the differential of $E(\hat{N})^2$ gives the sensitivity of $V(\hat{N})$:

$$\frac{dV(\hat{N})}{d\boldsymbol{\theta}^{\mathsf{T}}} = 2\left\{\left[\mathbf{b}^{\mathsf{T}} \otimes (\mathbf{c} \circ \mathbf{c})^{\mathsf{T}}\right]\mathcal{D}\,(\mathbb{N}) - \bar{N}\left(\mathbf{b}^{\mathsf{T}} \otimes \mathbf{c}^{\mathsf{T}}\right)\right\}\frac{d\mathbb{N}}{d\boldsymbol{\theta}^{\mathsf{T}}} \tag{10.137}$$

where $d\mathbb{N}/d\boldsymbol{\theta}^{\mathsf{T}}$ is given by (10.121). The extension to higher moments, should one want to know, say, the sensitivity of the skewness of population size over a cycle, is possible.

10.7.4   Periodic Dynamics in Periodic Environments

Periodic environments (e.g., seasons within a year) are described by periodic products of matrices. If the environmental cycle contains p phases, then matrices $\mathbf{A}_1, \ldots \mathbf{A}_p$ describe the dynamics at each phase, and the periodic product $\mathbf{A}_p \cdots \mathbf{A}_1$ projects the population over an entire cycle. Nonlinear periodic models permit the $\mathbf{A}_i$ to depend on the population vector at any point in the cycle, including delayed dependence (e.g., the reproductive success of an individual plant in the fall may depend on the density it experienced in the spring). A fixed point on the inter-annual time scale is a p-cycle on the seasonal time scale. A k-cycle on the inter-annual

scale corresponds to a kp-cycle on the seasonal time scale. The sensitivity analysis of these models is given by Caswell and Shyu (2012) and presented here in Chap. 8. For an application to the dynamics of an invasive plant population, see Shyu et al. (2013).

10.8   Dynamic Environmental Feedback Models

The commonly encountered forms of density dependence are usually a shorthand for a feedback between a population and some aspect of its environment.[18] The static feedback model of Sect. 10.3 begins to incorporate environmental feedback, but assumed that the environmental variable $\mathbf{g}(t)$ had no inherent dynamics of its own. A more general, dynamic environmental feedback model can be written

$$\mathbf{n}(t+1) = \mathbf{A}[\boldsymbol{\theta}, \mathbf{n}(t), \mathbf{g}(t)]\mathbf{n}(t)$$
$$\mathbf{g}(t+1) = \mathbf{B}[\boldsymbol{\theta}, \mathbf{n}(t), \mathbf{g}(t)]\mathbf{g}(t) \qquad (10.138)$$

allowing for $\mathbf{n}(t)$ to depend on both the environment and on its own density, and likewise for the environmental factor.

The sensitivity of the equilibrium of (10.138) can be found using an approach similar to that applied above to cycles. At equilibrium,

$$\hat{\mathbf{n}} = \mathbf{A}[\boldsymbol{\theta}, \hat{\mathbf{n}}, \hat{\mathbf{g}}]\hat{\mathbf{n}} \qquad (10.139)$$

$$\hat{\mathbf{g}} = \mathbf{B}[\boldsymbol{\theta}, \hat{\mathbf{n}}, \hat{\mathbf{g}}]\hat{\mathbf{g}} \qquad (10.140)$$

Differentiating both sides of each equation, expanding $d\text{vec}\,\mathbf{A}$ and $d\text{vec}\,\mathbf{B}$, and applying the vec operator gives

$$d\hat{\mathbf{n}} = \mathbf{A}\left(d\hat{\mathbf{n}}\right) + \left(\hat{\mathbf{n}}^{\mathsf{T}} \otimes \mathbf{I}_s\right)\left(\frac{\partial \text{vec}\,\mathbf{A}}{\partial \boldsymbol{\theta}^{\mathsf{T}}}d\boldsymbol{\theta} + \frac{\partial \mathbf{A}}{\partial \mathbf{n}^{\mathsf{T}}}d\hat{\mathbf{n}} + \frac{\partial \mathbf{A}}{\partial \mathbf{g}^{\mathsf{T}}}d\hat{\mathbf{g}}\right) \quad (10.141)$$

$$d\hat{\mathbf{g}} = \mathbf{B}\left(d\hat{\mathbf{n}}\right) + \left(\hat{\mathbf{g}}^{\mathsf{T}} \otimes \mathbf{I}_q\right)\left(\frac{\partial \text{vec}\,\mathbf{B}}{\partial \boldsymbol{\theta}^{\mathsf{T}}}d\boldsymbol{\theta} + \frac{\partial \mathbf{B}}{\partial \mathbf{n}^{\mathsf{T}}}d\hat{\mathbf{n}} + \frac{\partial \mathbf{B}}{\partial \mathbf{g}^{\mathsf{T}}}d\hat{\mathbf{g}}\right). \quad (10.142)$$

Applying the identification theorem and the chain rule gives

---

[18]Early writers even interpreted the simple logistic equation as an interplay between a biotic potential for exponential growth and an environmental resistance due to lack of resources or interaction with predators (e.g., Chapman 1931). Incorporating a fully dynamic feedback greatly expands the range of phenomena that can be explained (see de Roos and Persson (2013) for an extensive development of this approach).

$$\frac{d\hat{\mathbf{n}}}{d\boldsymbol{\theta}^\mathsf{T}} = \mathbf{A}\frac{d\hat{\mathbf{n}}}{d\boldsymbol{\theta}^\mathsf{T}} + \left(\hat{\mathbf{n}} \otimes \mathbf{I}_s\right)\frac{\partial \text{vec}\,\mathbf{A}}{\partial \boldsymbol{\theta}^\mathsf{T}} + \left(\hat{\mathbf{n}} \otimes \mathbf{I}_s\right)\frac{\partial \text{vec}\,\mathbf{A}}{\partial \mathbf{n}^\mathsf{T}}\frac{d\hat{\mathbf{n}}}{d\boldsymbol{\theta}^\mathsf{T}}$$

$$+ \left(\hat{\mathbf{n}} \otimes \mathbf{I}_s\right)\frac{\partial \text{vec}\,\mathbf{A}}{\partial \mathbf{g}^\mathsf{T}}\frac{d\hat{\mathbf{g}}}{d\boldsymbol{\theta}^\mathsf{T}} \tag{10.143}$$

with a similar expression for $d\hat{\mathbf{g}}/d\boldsymbol{\theta}^\mathsf{T}$. All matrices and their derivatives are evaluated at the equilibrium $(\hat{\mathbf{n}}, \hat{\mathbf{g}})$. This system can be written in block matrix form by defining

$$\mathbf{H} \equiv \left(\hat{\mathbf{n}}^\mathsf{T} \otimes \mathbf{I}_s\right) \tag{10.144}$$

$$\mathbf{J} \equiv \left(\hat{\mathbf{g}}^\mathsf{T} \otimes \mathbf{I}_q\right) \tag{10.145}$$

Then define

$$\mathbb{A} = \left(\begin{array}{c|c} \mathbf{A} & 0 \\ \hline 0 & \mathbf{B} \end{array}\right) \tag{10.146}$$

$$\mathbb{H} = \left(\begin{array}{c|c} 0 & \mathbf{H} \\ \hline \mathbf{J} & 0 \end{array}\right) \tag{10.147}$$

$$\mathbb{C} = \left(\begin{array}{c|c} \dfrac{\partial \text{vec}\,\mathbf{B}}{\partial \mathbf{n}^\mathsf{T}} & \dfrac{\partial \text{vec}\,\mathbf{B}}{\partial \mathbf{g}^\mathsf{T}} \\ \hline \dfrac{\partial \text{vec}\,\mathbf{A}}{\partial \mathbf{n}^\mathsf{T}} & \dfrac{\partial \text{vec}\,\mathbf{A}}{\partial \mathbf{g}^\mathsf{T}} \end{array}\right) \tag{10.148}$$

$$\mathbb{D} = \left(\begin{array}{c} \dfrac{\partial \text{vec}\,\mathbf{A}}{\partial \boldsymbol{\theta}^\mathsf{T}} \\ \dfrac{\partial \text{vec}\,\mathbf{B}}{\partial \boldsymbol{\theta}^\mathsf{T}} \end{array}\right) \tag{10.149}$$

$$\mathbb{N} = \left(\begin{array}{c} \hat{\mathbf{n}} \\ \hat{\mathbf{g}} \end{array}\right) \tag{10.150}$$

In terms of these matrices,

$$\frac{d\mathbb{N}}{d\boldsymbol{\theta}^\mathsf{T}} = \mathbb{H}\mathbb{D} + (\mathbb{A} + \mathbb{H}\mathbb{C})\frac{d\mathbb{N}}{d\boldsymbol{\theta}^\mathsf{T}}. \tag{10.151}$$

Solving for $d\mathbb{N}/d\boldsymbol{\theta}^t r$ gives the sensitivity of both the population and the environmental factor,

$$\frac{d\mathbb{N}}{d\boldsymbol{\theta}^\mathsf{T}} = \left(\mathbf{I}_{s+q} - \mathbb{A} - \mathbb{HC}\right)^{-1} \mathbb{HD}. \qquad (10.152)$$

10.9 Stage-Structured Epidemics

The transmission of infectious diseases is a source of nonlinearity because the rate of transmission depends on the abundance of infected and non-infected individuals. When demographic structure is added to the picture, the models can become complicated because the transmission process, the recovery process, and the consequences of infection may all vary among age classes or stages.

Klepac and Caswell (2011) developed a general framework for stage-classified epidemics, using the vec-permutation formulation (e.g., Chaps. 5 and 6). Individuals were jointly classified by stage and infection category, and nonlinearity was introduced by the disease transmission process. Klepac and Caswell (2011) calculated sensitivities and elasticities of equilibria and cycles of the stage × infection distribution and, of stage-specific prevalence, to parameters specifying demographic, infection, and recovery processes.

Coupling demography and epidemiology requires attention to time scales. Suppose that the demographic processes operate on one time scale: say, years. For some diseases, the infection/recovery process might operate on a much longer time scale (decades). Or the disease might play out on a much shorter time scale (weeks). When the disease time scale is shorter than the demographic time scale, the matrices in Klepac's model that define disease transmission operate many times within a single year; the result is a periodic model on the infection time scale. See Klepac and Caswell (2011) for details.

10.10 Moments of Longevity in Nonlinear Models

The statistics of longevity (e.g., life expectancy) are traditionally calculated from linear age-classified models (see Chap. 4) or from linear stage-classified models (see Chap. 5). In a nonlinear model at equilibrium, the projection matrix is constant and an individual experiences a fixed schedule of vital rates, from which all the usual statistics of longevity can be calculated. Write the density-dependent projection matrix as

$$\mathbf{A}[\boldsymbol{\theta}, \mathbf{n}] = \mathbf{U}[\boldsymbol{\theta}, \mathbf{n}] + \mathbf{F}[\boldsymbol{\theta}, \mathbf{n}] \qquad (10.153)$$

where $\mathbf{U}$ contains the transition probabilities for individuals already present in the population and $\mathbf{F}$ describes the production of new individuals by reproduction. The matrix $\mathbf{U}$ is the transient matrix of an absorbing Markov chain, with death as an absorbing state. The fundamental matrix of this chain at equilibrium is

$$\mathbf{N}[\boldsymbol{\theta}, \hat{\mathbf{n}}] = \left(\mathbf{I}_s - \mathbf{U}[\boldsymbol{\theta}, \hat{\mathbf{n}}]\right)^{-1} \tag{10.154}$$

where the inverse is guaranteed to exist if the spectral radius of $\mathbf{U}$ is less than 1. The (i, j) element of $\mathbf{N}$ is the expected time spent in stage i, before death, by an individual in stage j.

As in Chap. 4, the vector $\boldsymbol{\eta}_1$ containing the mean longevity of each age class or stage is given by

$$\boldsymbol{\eta}_1^\mathsf{T} = \mathbf{1}_s^\mathsf{T} \mathbf{N}[\hat{\mathbf{n}}]. \tag{10.155}$$

The moments of longevity and other indices are calculated from $\mathbf{N}\left(\boldsymbol{\theta}, \hat{\mathbf{n}}\right)$ just as in the linear case. All the sensitivity results of Chaps. 4 and 5 apply directly, except that the derivative of $\mathbf{N}\left(\boldsymbol{\theta}, \hat{\mathbf{n}}\right)$ must include both the direct effects of $\boldsymbol{\theta}$ and the indirect effects through $\hat{\mathbf{n}}$. For convenience, write $\hat{\mathbf{N}}$ and $\hat{\mathbf{U}}$ for the matrices at equilibrium. Then

$$d\operatorname{vec}\hat{\mathbf{N}} = \left(\hat{\mathbf{N}}^\mathsf{T} \otimes \hat{\mathbf{N}}\right) d\operatorname{vec}\hat{\mathbf{U}} \tag{10.156}$$

$$= \left(\hat{\mathbf{N}}^\mathsf{T} \otimes \hat{\mathbf{N}}\right)\left[\frac{d\operatorname{vec}\hat{\mathbf{U}}}{d\boldsymbol{\theta}^\mathsf{T}}d\boldsymbol{\theta} + \frac{d\hat{\mathbf{U}}}{d\hat{\mathbf{n}}^\mathsf{T}}d\hat{\mathbf{n}}\right] \tag{10.157}$$

where $\hat{\mathbf{U}}$, $\hat{\mathbf{N}}$, and the derivatives of $\mathbf{U}$ are all evaluated at equilibrium and $d\hat{\mathbf{n}}/d\boldsymbol{\theta}^\mathsf{T}$ is given by (10.16). Comparing this with equation (4.34) shows that the nonlinearity adds an extra term, capturing the way that changes in $\boldsymbol{\theta}$ affect the vital rates through changes in equilibrium density.

This approach can be used to generalize the results for higher moments of longevity (Chaps. 4, 5, and 11) to the nonlinear case.

10.11 Summary

Table 10.3 lists the perturbation analysis results in this chapter; they comprise a fairly complete analysis for nonlinear demographic models. The nonlinearities may arise from density dependence, frequency dependence, environmental feedback, proportional population structure calculations, or recruitment subsidy. The sensitivity calculations accommodate a wide range of dependent variables and the calculation of both sensitivity and elasticity with respect to any kind of demographic parameters.

As in other chapters, most of the results in this chapter follow a straightforward method:

1. Write the model, specifying the dependence of the vital rates on $\boldsymbol{\theta}$ and $\mathbf{n}$.

Table 10.3 Summary of models and main sensitivity results of the chapter. Extending sensitivities to additional dependent variables (ratios, averages, rates, etc.) is described in Sect. 10.2.3. The extension from sensitivities to elasticities is given in Sect. 10.2.5

Model	Sensitivity of …		Equation[s]
Density-dependent	… equilibrium	$\dfrac{d\hat{\mathbf{n}}}{d\boldsymbol{\theta}^{\mathsf{T}}}$	(10.16)
$\mathbf{n}(t+1) = \mathbf{A}[\boldsymbol{\theta}, \mathbf{n}(t)]\,\mathbf{n}(t)$	… cycle	$\dfrac{d\mathbb{N}}{d\boldsymbol{\theta}^{\mathsf{T}}}$	(10.121)
	… weighted density, time average		(10.125)–(10.128)
	… temporal variance	$\dfrac{dV(\hat{N})}{d\boldsymbol{\theta}^{\mathsf{T}}}$	(10.137)
	… life expectancy	$\dfrac{d\eta_1}{d\boldsymbol{\theta}^{\mathsf{T}}}$	(10.155)
Environmental feedback			
$\mathbf{n}(t+1) = \mathbf{A}[\boldsymbol{\theta}, \mathbf{n}(t), \mathbf{g}(t)]\,\mathbf{n}(t)$			
$\mathbf{g}(t) = \mathbf{g}[\boldsymbol{\theta}, \mathbf{n}(t)]$	… equilibrium (static)	$\dfrac{d\hat{\mathbf{n}}}{d\boldsymbol{\theta}^{\mathsf{T}}}, \;\; \dfrac{d\hat{\mathbf{g}}}{d\boldsymbol{\theta}^{\mathsf{T}}}$	(10.52), (10.53)
$\mathbf{g}(t+1) = \mathbf{B}[\boldsymbol{\theta}, \mathbf{n}(t), \mathbf{g}(t)]\,\mathbf{g}(t)$	… equilibrium (dynamic)	$\dfrac{d\mathbb{N}}{d\boldsymbol{\theta}^{\mathsf{T}}}$	(10.152)
Frequency-dependent (two-sex)	… equilibrium structure	$\dfrac{d\hat{\mathbf{p}}}{d\boldsymbol{\theta}^{\mathsf{T}}}$	(10.94)
$\mathbf{n}(t+1) = \mathbf{A}[\boldsymbol{\theta}, \mathbf{n}(t)]\,\mathbf{n}(t)$	… population growth rate	$\dfrac{d\lambda}{d\boldsymbol{\theta}^{\mathsf{T}}}$	(10.97)
$\mathbf{A}[\boldsymbol{\theta}, \mathbf{n}]$ homogeneous of degree zero in $\mathbf{n}$			

Subsidized (linear or nonlinear) $\mathbf{n}(t+1) = \mathbf{A}[\boldsymbol{\theta}, \mathbf{n}(t)]\,\mathbf{n}(t) + \mathbf{b}[\boldsymbol{\theta}, \mathbf{n}]$ …equilibrium population	$\dfrac{d\hat{\mathbf{n}}}{d\boldsymbol{\theta}^{\mathsf{T}}}$	(10.58), (10.68)
Proportional structure …age- or stage distribution	$\dfrac{d\hat{\mathbf{p}}}{d\boldsymbol{\theta}^{\mathsf{T}}}$	(10.73)
$\mathbf{p}(t+1) = \dfrac{\mathbf{A}[\boldsymbol{\theta}]\mathbf{p}(t)}{\mathbf{1}^{\mathsf{T}}\mathbf{A}[\boldsymbol{\theta}]\mathbf{p}(t)}$ …reproductive value	$\dfrac{d\hat{\mathbf{v}}}{d\boldsymbol{\theta}^{\mathsf{T}}}$	(10.77)
…dependency ratio	$\dfrac{dD}{d\boldsymbol{\theta}^{\mathsf{T}}}$	(10.81)
…mean age of reproduction	$\dfrac{d\bar{a}_{\mathrm{f}}}{d\boldsymbol{\theta}^{\mathsf{T}}}$	(10.83)
…variance in age of reproduction	$\dfrac{dV_{\mathrm{f}}}{d\boldsymbol{\theta}^{\mathsf{T}}}$	(10.89)

2. Write a matrix expression for the demographic outcome of interest (e.g., the equilibrium population).
3. Differentiate this expression.
4. Use the vec operator and Roth's theorem to obtain an expression that involves only the differentials of vectors.
5. Use the chain rule for total differentials to expand the operators (e.g., $dvec\mathbf{A}$) that are functions of both $\boldsymbol{\theta}$ and $\mathbf{n}$, as in (10.14).
6. Use the first identification theorem and the chain rule to extend the results to sensitivities of any desired dependent variable with respect to any set of parameters

References

Almany, G. R., M. L. Berumen, S. R. Thorrold, S. Planes, and G. P. Jones. 2007. Local replenishment of coral reef fish populations in a marine reserve. Science **316**:742–744.

Bartholomew, D. J. 1982. Stochastic models for social processes. Wiley, New York, New York, USA.

Bonenfant, C., J.-M. Gaillard, T. Coulson, M. Festa-Bianchet, A. Loison, M. Garel, L. E. Loe, P. Blanchard, N. Pettorelli, J. Owen-Smith, Norman Du Toit, and P. Duncan. 2009. Empirical evidence of density-dependence in populations of large herbivores. Advances in Ecological Research **41**:313–357.

Bonneuil, N. 1994. Special issue: nonlinear models in demography. Mathematical Population Studies **5**:1.

Carey, J. R., P. Liedo, and J. W. Vaupel. 1995. Mortality dynamics of density in the Mediterranean fruit fly. Experimental Gerontology **30**:605–629.

Caswell, H. 1978. A general formula for the sensitivity of population growth rate to changes in life history parameters. Theoretical Population Biology **14**:215–230.

Caswell, H. 1982. Optimal life histories and the maximization of reproductive value: a general theorem for complex life cycles. Ecology **63**:1218–1222.

Caswell, H. 2001. Matrix Population Models: Construction, Analysis, and Interpretation. 2nd edition. Sinauer Associates, Sunderland, MA.

Caswell, H. 2007. Sensitivity analysis of transient population dynamics. Ecology Letters **10**:1–15.

Caswell, H., and M. G. Neubert. 2005. Reactivity and transient dynamics of discrete-time ecological systems. Journal of Difference Equations and Applications **11**:295–310.

Caswell, H., and N. Sanchez Gassen. 2015. The sensitivity analysis of population projections. Demographic Research **33**:801–840.

Caswell, H., and E. Shyu. 2012. Sensitivity analysis of periodic matrix population models. Theoretical Population Biology **82**:329–339.

Caswell, H., and Y. Vindenes. 2018. Demographic variance in heterogeneous populations: Matrix models and sensitivity analysis. Oikos **127**:648–663.

Caswell, H., and D. E. Weeks. 1986. Two-sex models: chaos, extinction, and other dynamic consequences of sex. The American Naturalist **128**:707–735.

Chapman, R. N. 1931. Animal Ecology with Especial Reference to Insects. McGraw-Hill, New York, New York, USA.

Chen, X., and J. E. Cohen. 2001. Transient dynamics and food–web complexity in the Lotka–Volterra cascade model. Proceedings of the Royal Society of London B: Biological Sciences **268**:869–877.

Chu, C. Y. C. 1998. Population dynamics: a new economic approach. Oxford University Press, New York, New York, USA.

Chung, R. 1994. Cycles in the two-sex problem: an investigation of a nonlinear demographic model. Mathematical Population Studies **5**:45–74.

Clutton-Brock, J. H., A. W. Illius, K. Wilson, B. T. Grenfell, A. D. C. MacColl, and S. D. Albon. 1997. Stability and instability in ungulate populations: an empirical analysis. American Naturalist **149**:195–219.

Coale, A. J. 1972. The growth and structure of human populations: a mathematical approach. Princeton University Press, Princeton, New Jersey.

Cohen, J. E. 1995. How many people can the earth support? W. W. Norton, New York, New York, USA.

Costantino, R. F., and R. A. Desharnais. 1991. Population dynamics and the *Tribolium* model: genetics and demography. Springer-Verlag, New York, New York, USA.

Costantino, R. F., R. A. Desharnais, J. M. Cushing, and B. Dennis. 1997. Chaotic dynamics in an insect population. Science **275**:389–391.

Cushing, J. M. 1998. An introduction to structured population dynamics. SIAM, Philadelphia, Pennsylvania, USA.

Cushing, J. M., R. F. Costantino, B. Dennis, R. A. Desharnais, and S. M. Henson. 2003. Chaos in ecology: experimental nonlinear dynamics. Academic Press, San Diego, California, USA.

de Roos, A. M., and L. Persson. 2013. Population and community ecology of ontogenetic development. Princeton University Press, Princeton, New Jersey, USA.

de Vries, C., and H. Caswell. 2018. Stage-stuctured evolutionary demography: linking life histories, population genetics, and ecological dynamics. (in prep.).

Dennis, B., R. A. Desharnais, J. M. Cushing, and R. F. Costantino. 1995. Nonlinear demographic dynamics: mathematical models, statistical methods, and biological experiments. Ecological Monographs **65**:261–282.

Easterlin, R. 1961. The American baby boom in historical perspective. American Economic Review **51**:860–911.

Emekci, M., S. Navarro, J. E. Donahaye, M. Rindner, and A. Azrieli, 2001. Respiration of stored product pests in hermetic conditions. Pages 26–35 *in* J. Donahaye, S. Navarro, and J. G. Leesch, editors. Proceedings of the International Conference on Controlled Atmosphere and Fumigation in Stored Products. Executive Printing Services, Clovis, California, USA.

Fennel, W., and T. Neumann. 2004. Introduction to the modelling of marine ecosystems. Elsevier, Amsterdam, Netherlands.

Frank, P. W., C. D. Boll, and R. W. Kelly. 1957. Vital statistics of laboratory cultures of *Daphnia pulex* DeGeer as related to density. Physiological Zoology **30**:287–305.

Frauenthal, J., and K. Swick. 1983. Limit cycle oscillations of the human population. Demography **20**:285–298.

Fürnkranz-Prskawetz, A., and J. Sambt. 2014. Economic support ratios and the demographic dividend in Europe. Demographic Research **30**:963–1010.

Gaines, S., and J. Roughgarden. 1985. Larval settlement rate: a leading determinant of structure in an ecological community of the marine intertidal zone. Proceedings of the National Academy of Sciences USA **82**:3707–3711.

Gani, J. 1963. Formulae for projecting enrolments and degrees awarded in universities. Journal of the Royal Statistical Society A **126**:400–409.

Gillman, M., J. M. Bullock, J. ilvertown, and B. C. Hill. 1993. A density-dependent model of *Cirsium vulgare* population dynamics using field-estimated parameter values. Oecologia **96**:282–289.

Hauser, C. E., E. G. Cooch, and J.-D. Lebreton. 2006. Control of structured populations by harvest. Ecological Modelling **196**:462–470.

Hinde, A. 1998. Demographic methods. Arnold, London, United Kingdom.

Hsu, S. B., S. P. Hubbell, and P. Waltman. 1977. A mathematical theory for single-nutrient competition in continuous cultures of microorganisms. SIAM Journal of Applied Mathematics **32**:366–383.

Iannelli, M., M. Martcheva, and F. A. Milner. 2005. Gender-structured population modeling: mathematical methods, numerics, and simulations. SIAM, Philadelphia, Pennsylvania, USA.

Jenouvrier, S., H. Caswell, C. Barbraud, and H. Weimerskirch. 2010. Mating behavior, population growth, and the operational sex ratio: a periodic two-sex model approach. The American Naturalist **175**:739–752.

Jenouvrier, S., M. Holland, J. Stroeve, C. Barbraud, H. Weimerskirch, M. Serreze, and H. Caswell. 2012. Effects of climate change on an emperor penguin population: analysis of coupled demographic and climate models. Global Change Biology **18**:2756–2770.

Keyfitz, N. 1972. The mathematics of sex and marriage. Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability **6**:89–108.

Keyfitz, N., and H. Caswell. 2005. Applied mathematical demography. 3rd edition. Springer, New York, New York.

Keyfitz, N., and W. Flieger. 1968. World population: an analysis of vital data. The University of Chicago Press, Chicago, Illinois, USA.

Keyfitz, N., and W. Flieger. 1990. World population growth and aging: demographic trends in the late twentieth century. University of Chicago Press, Chicago, Illinois, USA.

Kirkland, S. J., and M. Neumann. 1994. Convexity and concavity of the Perron root and vector of Leslie matrices with applications to a population model. SIAM Journal of Matrix Analysis and Applications **15**:1092–1107.

Klepac, P., and H. Caswell. 2011. The stage-structured epidemic: linking disease and demography with a multi-state matrix approach model. Theoretical Ecology **4**:301–319.

Kokko, H., and D. J. Rankin. 2006. Lonely hearts or sex in the city? Density-dependent effects in mating systems. Philosophical Transactions of the Royal Society B **361**:319–334.

Lee, C. T., and S. Tuljapurkar. 2008. Population and prehistory I: Food-dependent population growth in constant environments. Theoretical Population Biology **73**:473–482.

Lee, R. D. 1974. The formal dynamics of controlled populations and the echo, the boom, and the bust. Demography **11**:563–585.

Lee, R. D. 1976. Demographic forecasting and the Easterlin hypothesis. Population and Development Review **2**:459–468.

Lee, R. D., 1986. Malthus and Boserup: a dynamic synthesis. Pages 96–130 *in* D. Coleman and R. Schofield, editors. The state of population theory: forward from Malthus. Blackwell, Oxford, United Kingdom.

Lee, R. D. 1987. Population dynamics of humans and other animals. Demography **24**:443–465.

Legendre, S., J. Clobert, A. P. Møller, and G. Sorci. 1999. Demographic stochasticity and social mating system in the process of extinction of small populations: the case of passerines introduced to New Zealand. American Naturalist **153**:449–463.

Lenz, T. L., A. Jacob, and C. Wedekind. 2007. Manipulating sex ratio to increase population growth: the example of the Lesser Kestrel. Animal Conservation **10**:236–244.

Lindström, J., and H. Kokko, 1998. Sexual reproduction and population dynamics: the role of polygyny and demographic sex differences. Proceedings of the Royal Society B **265**:483–488.

Martcheva, M. 1999. Exponential growth in age-structured two-sex populations. Mathematical Biosciences **157**:1–22.

Marvier, M., P. Kareiva, and M. G. Neubert. 2004. Habitat destruction, fragmentation, and disturbance promote invasion by habitat generalists in a multispecies metapopulation. Risk analysis **24**:869–878.

McFarland, D. D., 1972. Comparison of alternative marriage models. *in* T. N. E. Greville, editor. Population dynamics. Academic Press, New York, New York, USA.

Mueller, L. D., and A. Joshi. 2000. Stability in model populations. Princeton University Press, Princeton, New Jersey, USA.

Murdoch, W. W., C. J. Briggs, and R. M. Nisbet. 2003. Consumer-resource dynamics. Princeton University Press, Princeton, New Jersey, USA.

Neubert, M. G., and H. Caswell. 1997. Alternatives to resilience for measuring the responses of ecological systems to perturbations. Ecology **78**:653–665.

Neubert, M. G., and H. Caswell. 2000. Density-dependent vital rates and their population dynamic consequences. Journal of Mathematical Biology **43**:103–121.

Neubert, M. G., T. Klanjscek, and H. Caswell. 2004. Reactivity and transient dynamics of predator-prey and food web models. Ecological Modelling **179**:29–38.

Nussbaum, R. D. 1988. Iterated nonlinear maps and Hilbert's projective metric. American Mathematical Society, Providence, Rhode Island, USA.

Nussbaum, R. D. 1989. Iterated nonlinear maps and Hilbert's projective metric, II. American Mathematical Society, Providence, Rhode Island, USA.

Pardini, E. A., J. M. Drake, J. M. Chase, and T. M. Knight. 2009. Complex population dynamics and control of the invasive biennial *Alliaria petiolata* (garlic mustard). Ecological Applications **19**:387–397.

Pascual, M., and H. Caswell. 1991. The dynamics of a size-classified benthic population with reproductive subsidy. Theoretical Population Biology **39**:129–147.

Pearl, R., J. R. Miner, and S. L. Parker. 1927. Experimental studies on the duration of life. XI. Density of population and life duration in *Drosophila*. American Naturalist **61**:289–318.

Pennycuick, L. 1969. A computer model of the Oxford Great Tit population. Journal of Theoretical Biology **22**:381–400.

Pollak, R. A. 1987. The two-sex problem with persistent unions: a generalization of the birth matrix-mating rule model. Theoretical Population Biology **32**:176–187.

Pollak, R. A. 1990. Two-sex demographic models. Journal of Political Economy **98**:399–420.

Pollard, J. H. 1968. A note on the age structures of learned societies. Journal of the Royal Statistical Society A **131**:569–578.

Pollard, J. H., 1977. The continuing attempt to incorporate both sexes into marriage analysis. Pages 291–308 *in* International Population Conference. International Union for the Scientific Study of Population.

Preston, S. H., P. Heuveline, and M. Guillot. 2001. Demography: measuring and modeling population processes. Blackwell, Oxford, United Kingdom.

Promislow, D. E. L., R. Montgomerie, and T. E. Martin. 1992. Mortality costs of sexual dimorphism in birds. Proceedings of the Royal Society of London B **250**:143–150.

Roughgarden, J., Y. Iwasa, and C. Baxter. 1985. Demographic theory for an open marine population with space-limited recruitment. Ecology **66**:54–67.

Sarrazin, F., and S. Legendre. 2000. Demographic approach to releasing adults versus young in reintroductions. Conservation Biology **14**:488–500.

Scheltema, R. S. 1971. Larval dispersal as a means of genetic exchange between geographically separated populations of shallow-water benthic marine gastropods. Biological Bulletin **140**:284–322.

Scott, S., and C. J. Duncan. 1998. Human demography and disease. Cambridge University Press, Cambridge, United Kingdom.

Shyu, E., and H. Caswell. 2014. Calculating second derivatives of population growth rates for ecology and evolution. Methods in Ecology and Evolution **5**:473–482.

Shyu, E., and H. Caswell. 2016a. A demographic model for sex ratio evolution and the effects of sex-biased offspring costs. Ecology and Evolution **6**:1470–1492.

Shyu, E., and H. Caswell. 2016b. Frequency-dependent two-sex models: a new approach to sex ratio evolution with multiple maternal conditions. Ecology and Evolution **6**:6855–6879.

Shyu, E., and H. Caswell. 2018. Mating, births, and transitions: a flexible two-sex matrix model for evolutionary demography. Population Ecology **60**:21–36.

Shyu, E., E. A. Pardini, T. M. Knight, and H. Caswell. 2013. A seasonal, density-dependent model for the management of an invasive weed. Ecological Applications **23**:1893–1905.

Silva Matos, D. M., R. P. Freckleton, and A. R. Watkinson. 1999. The role of density dependence in the population dynamics of a tropical palm. Ecology **80**:2635–2650.

Skalski, J. R., K. E. Ryding, and J. Millspaugh. 2005. Wildlife demography: analysis of sex, age, and count data. Academic Press, San Diego, California, USA.

Solbrig, O. T., R. Sarandon, and W. Bossert. 1988. A density-dependent growth model of a perennial herb, *Viola fimbriatula*. American Naturalist **131**:385–400.

Stott, I. 2016. Perturbation analysis of transient population dynamics using matrix projection models. Methods in Ecology and Evolution **7**:666–678.

Stott, I., S. Townley, and D. J. Hodgson. 2011. A framework for studying transient dynamics of population projection matrix models. Ecology Letters **14**:959–970.

Tilman, D. 1982. Resource competition and community structure. Princeton University Press, Princeton, New Jersey, USA.

Tuljapurkar, S., 1997. Stochastic matrix models. Pages 59–87 *in* S. Tuljapurkar and H. Caswell, editors. Structured-population models in marine terrestrial, and freshwater systems. Chapman and Hall, New York, New York, USA.

Tuljapurkar, S., C. O. Puleston, and M. D. Gurven. 2007. Why men matter: mating patterns drive evolution of human lifespan. PLoS ONE **2**:e785.

Verdy, A., and H. Caswell. 2008. Sensitivity analysis of reactive ecological dynamics. Bulletin of Mathematical Biology **70**:1634–1659.

Wachter, K. W., and R. D. Lee. 1989. U.S. births and limit cycle models. Demography **26**:99–115.

Weiss, K. M., and P. E. Smouse. 1976. The demographic stability of small human populations. Journal of Human Evolution **5**:59–73.

Wood, J. W., and P. E. Smouse. 1982. A method of analyzing density-dependent vital rates with an applicaion to the Gainj of Papua New Guinea. American Journal of Physical Anthropology **58**:403–411.

Part V
Markov Chains

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.

255

11.2   Absorbing Chains

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

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

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 ν_{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 ν_{ij} are entries of the fundamental matrix $\mathbf{N} = \mathbf{N}_1 = \left(E(\eta_{ij}) \right)$:

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

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

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

$$\mathbf{N}_2 = \left(2\mathbf{N}_{\mathrm{dg}} - \mathbf{I} \right) \mathbf{N}_1 \tag{11.4}$$

$$\mathbf{N}_3 = \left(6\mathbf{N}_{\mathrm{dg}}^2 - 6\mathbf{N}_{\mathrm{dg}} + \mathbf{I} \right) \mathbf{N}_1 \tag{11.5}$$

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

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

$$d\mathrm{vec}\,\mathbf{N}_1 = \left(\mathbf{N}_1^\mathsf{T} \otimes \mathbf{N}_1 \right) d\mathrm{vec}\,\mathbf{U} \tag{11.7}$$

$$dvec\,\mathbf{N}_2 = \left[2\left(\mathbf{I} \otimes \mathbf{N}_{dg}\right) - \mathbf{I}_{s^2}\right] dvec\,\mathbf{N}_1 + 2\left(\mathbf{N}^\mathsf{T} \otimes \mathbf{I}\right) dvec\,\mathbf{N}_{dg} \tag{11.8}$$

$$dvec\,\mathbf{N}_3 = \left[\mathbf{I} \otimes \left(6\mathbf{N}_{dg}^2 - 6\mathbf{N}_{dg} + \mathbf{I}\right)\right] dvec\,\mathbf{N}_1$$

$$+ \left[6\left(\mathbf{N}^\mathsf{T}\mathbf{N}_{dg} \otimes \mathbf{I}\right) + 6\left(\mathbf{N}^\mathsf{T} \otimes \mathbf{N}_{dg}\right) - 6\left(\mathbf{N}^\mathsf{T} \otimes \mathbf{I}\right)\right] dvec\,\mathbf{N}_{dg} \tag{11.9}$$

$$dvec\,\mathbf{N}_4 = \left[\mathbf{I} \otimes \left(24\mathbf{N}_{dg}^3 - 36\mathbf{N}_{dg}^2 + 14\mathbf{N}_{dg} - \mathbf{I}\right)\right] dvec\,\mathbf{N}_1$$

$$+ \left[24\left(\mathbf{N}^\mathsf{T}\mathbf{N}_{dg}^2 \otimes \mathbf{I}\right) + 24\left(\mathbf{N}^\mathsf{T}\mathbf{N}_{dg} \otimes \mathbf{N}_{dg}\right) + 24\left(\mathbf{N}^\mathsf{T} \otimes \mathbf{N}_{dg}^2\right)\right.$$

$$\left. -36\left(\mathbf{N}^\mathsf{T}\mathbf{N}_{dg} \otimes \mathbf{I}\right) - 36\left(\mathbf{N}^\mathsf{T} \otimes \mathbf{N}_{dg}\right) + 14\left(\mathbf{N}^\mathsf{T} \otimes \mathbf{I}\right)\right] dvec\,\mathbf{N}_{dg} \tag{11.10}$$

where (see Sect. 2.8)

$$d\mathbf{N}_{dg} = \mathbf{I} \circ d\mathbf{N}_1 \tag{11.11}$$

$$dvec\,\mathbf{N}_{dg} = \mathcal{D}\left(vec\,\mathbf{I}\right) dvec\,\mathbf{N}_1. \tag{11.12}$$

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

$$dvec\,\mathbf{N}_k = \frac{\partial vec\,\mathbf{N}_k}{\partial vec\,^\mathsf{T}\mathbf{N}_1} dvec\,\mathbf{N}_1 + \frac{\partial vec\,\mathbf{N}_k}{\partial vec\,^\mathsf{T}\mathbf{N}_{dg}} dvec\,\mathbf{N}_{dg}. \tag{11.13}$$

The two terms of (11.13) are the partial differentials of $vec\,\mathbf{N}_k$, 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 $\partial_{\mathbf{N}_{dg}}$ and $\partial_{\mathbf{N}_{dg}}$. Differentiating $\mathbf{N}_2$ in (11.4), gives

$$\partial_{\mathbf{N}_1}\mathbf{N}_2 = 2\mathbf{N}_{dg}\left(d\mathbf{N}_1\right) - d\mathbf{N}_1 \tag{11.14}$$

$$\partial_{\mathbf{N}_{dg}}\mathbf{N}_2 = 2\left(d\mathbf{N}_{dg}\right)\mathbf{N}_1. \tag{11.15}$$

Applying the vec operator gives

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

$$\partial_{\mathbf{N}_{dg}}vec\,\mathbf{N}_2 = 2\left(\mathbf{N}_1^\mathsf{T} \otimes \mathbf{I}\right) dvec\,\mathbf{N}_{dg}, \tag{11.17}$$

and (11.13) becomes

$$dvec\,\mathbf{N}_2 = \left[2\left(\mathbf{I} \otimes \mathbf{N}_{dg}\right) - \mathbf{I}_{s^2}\right] dvec\,\mathbf{N}_1 + 2\left(\mathbf{N}_1^\mathsf{T} \otimes \mathbf{I}\right) dvec\,\mathbf{N}_{dg}, \tag{11.18}$$

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.   □

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 $\boldsymbol{\eta}_k = E\left(\eta_1^k, \cdots, \eta_s^k\right)^{\mathsf{T}}$. The first several of these moments are (Iosifescu 1980, Thm. 3.2)

$$\boldsymbol{\eta}_1^{\mathsf{T}} = \mathbf{1}^{\mathsf{T}}\mathbf{N}_1 \tag{11.19}$$

$$\boldsymbol{\eta}_2^{\mathsf{T}} = \boldsymbol{\eta}_1^{\mathsf{T}}\left(2\mathbf{N}_1 - \mathbf{I}\right) \tag{11.20}$$

$$\boldsymbol{\eta}_3^{\mathsf{T}} = \boldsymbol{\eta}_1^{\mathsf{T}}\left(6\mathbf{N}_1^2 - 6\mathbf{N}_1 + \mathbf{I}\right) \tag{11.21}$$

$$\boldsymbol{\eta}_4^{\mathsf{T}} = \boldsymbol{\eta}_1^{\mathsf{T}}\left(24\mathbf{N}_1^3 - 36\mathbf{N}_1^2 + 14\mathbf{N}_1 - \mathbf{I}\right). \tag{11.22}$$

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

$$d\boldsymbol{\eta}_1 = \left(\mathbf{I} \otimes \mathbf{1}^{\mathsf{T}}\right) d\text{vec}\,\mathbf{N}_1 \tag{11.23}$$

$$d\boldsymbol{\eta}_2 = \left(2\mathbf{N}_1^{\mathsf{T}} - \mathbf{I}\right) d\boldsymbol{\eta}_1 + 2\left(\mathbf{I} \otimes \boldsymbol{\eta}_1^{\mathsf{T}}\right) d\text{vec}\,\mathbf{N}_1 \tag{11.24}$$

$$d\boldsymbol{\eta}_3 = \left(6\mathbf{n}^2 - 6\mathbf{N}_1 + \mathbf{I}\right)^{\mathsf{T}} d\boldsymbol{\eta}_1$$
$$+ \left[6\left(\mathbf{N}_1^{\mathsf{T}} \otimes \boldsymbol{\eta}_1^{\mathsf{T}}\right) + 6\left(\mathbf{I} \otimes \boldsymbol{\eta}_1^{\mathsf{T}}\mathbf{N}_1\right) - 6\left(\mathbf{I} \otimes \boldsymbol{\eta}_1^{\mathsf{T}}\right)\right] d\text{vec}\,\mathbf{N}_1 \tag{11.25}$$

$$d\boldsymbol{\eta}_4 = \left(24\mathbf{N}_1^3 - 36\mathbf{N}_1^2 + 14\mathbf{N}_1 - \mathbf{I}\right)^{\mathsf{T}} d\boldsymbol{\eta}_1$$
$$+ \left\{24\left[\left(\mathbf{N}_1^{\mathsf{T}}\right)^2 \otimes \boldsymbol{\eta}_1^{\mathsf{T}}\right] + 24\left(\mathbf{N}_1^{\mathsf{T}} \otimes \boldsymbol{\eta}_1^{\mathsf{T}}\mathbf{N}_1\right) + 24\left(\mathbf{I} \otimes \boldsymbol{\eta}_1^{\mathsf{T}}\mathbf{N}_1^2\right)\right.$$
$$\left. - 36\left(\mathbf{N}_1^{\mathsf{T}} \otimes \boldsymbol{\eta}_1^{\mathsf{T}}\right) - 36\left(\mathbf{I} \otimes \boldsymbol{\eta}_1^{\mathsf{T}}\mathbf{N}_1\right) + 14\left(\mathbf{I} \otimes \boldsymbol{\eta}_1^{\mathsf{T}}\right)\right\} d\text{vec}\,\mathbf{N}_1 \tag{11.26}$$

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

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

$$d\boldsymbol{\eta}_k = \frac{\partial \boldsymbol{\eta}_k}{\partial \boldsymbol{\eta}_1^\mathsf{T}}\, d\boldsymbol{\eta}_1 + \frac{\partial \boldsymbol{\eta}_k}{\partial \mathrm{vec}^\mathsf{T} \mathbf{N}_1}\, d\mathrm{vec}\,\mathbf{N}_1. \tag{11.27}$$

The partial differentials of $\boldsymbol{\eta}_2$ with respect to $\boldsymbol{\eta}_1$ and $\mathbf{N}_1$ are

$$\partial_{\boldsymbol{\eta}_1} \boldsymbol{\eta}_2^\mathsf{T} = \left(d\boldsymbol{\eta}_1^\mathsf{T}\right)(2\mathbf{N}_1 - \mathbf{I}) \tag{11.28}$$

$$\partial_{\mathbf{N}_1} \boldsymbol{\eta}_2^\mathsf{T} = 2\boldsymbol{\eta}_1^\mathsf{T}(d\mathbf{N}_1). \tag{11.29}$$

Applying the vec operator gives

$$\partial_{\boldsymbol{\eta}_1} \boldsymbol{\eta}_2 = \left(2\mathbf{N}_1^\mathsf{T} - \mathbf{I}\right) d\boldsymbol{\eta}_1 \tag{11.30}$$

$$\partial_{\mathbf{N}_1} \boldsymbol{\eta}_2 = 2\left(\mathbf{I} \otimes \boldsymbol{\eta}_1^\mathsf{T}\right) d\mathrm{vec}\,\mathbf{N}_1 \tag{11.31}$$

which combine according to (11.27) to yield (11.24). The derivations of $d\boldsymbol{\eta}_3$ and $d\boldsymbol{\eta}_4$ follow the same sequence of steps; the details are shown in Appendix A.   □

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 $\boldsymbol{\xi}_1 = E(\boldsymbol{\xi})$. Then

$$\boldsymbol{\xi}_1^\mathsf{T} = \mathbf{1}^\mathsf{T} \mathbf{N}_{\mathrm{dg}}^{-1} \mathbf{N}_1 \tag{11.32}$$

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

Theorem 11.2.3 *Let $\boldsymbol{\xi}_1 = E(\boldsymbol{\xi})$. The sensitivity of $\boldsymbol{\xi}$ is*

$$d\boldsymbol{\xi}_1 = \left[-\left(\mathbf{N}_1^\mathsf{T} \otimes \mathbf{1}^\mathsf{T}\right)\left(\mathbf{N}_{\mathrm{dg}}^{-1} \otimes \mathbf{N}_{\mathrm{dg}}^{-1}\right)\mathcal{D}\left(\mathrm{vec}\,\mathbf{I}\right) + \left(\mathbf{I} \otimes \mathbf{1}^\mathsf{T} \mathbf{N}_{\mathrm{dg}}^{-1}\right)\right] d\mathrm{vec}\,\mathbf{N}_1, \tag{11.33}$$

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

Proof Differentiating (11.32) yields

$$d\boldsymbol{\xi}_1^{\mathsf{T}} = \mathbf{1}^{\mathsf{T}} \left(d\mathbf{N}_{\mathrm{dg}}^{-1} \right) \mathbf{N}_1 + \mathbf{1}^{\mathsf{T}} \mathbf{N}_{\mathrm{dg}}^{-1} d\mathbf{N}_1. \tag{11.34}$$

Applying the vec operator yields

$$d\boldsymbol{\xi}_1 = \left(\mathbf{N}_1^{\mathsf{T}} \otimes \mathbf{1}^{\mathsf{T}} \right) d\mathrm{vec}\, \mathbf{N}_{\mathrm{dg}}^{-1} + \left(\mathbf{I} \otimes \mathbf{1}^{\mathsf{T}} \mathbf{N}_{\mathrm{dg}}^{-1} \right) d\mathrm{vec}\, \mathbf{N}_1. \tag{11.35}$$

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

$$d\boldsymbol{\xi}_1 = - \left(\mathbf{N}_1^{\mathsf{T}} \otimes \mathbf{1}^{\mathsf{T}} \right) \left(\mathbf{N}_{\mathrm{dg}}^{-1} \otimes \mathbf{N}_{\mathrm{dg}}^{-1} \right) \mathcal{D} \left(\mathrm{vec}\, \mathbf{I} \right) d\mathrm{vec}\, \mathbf{N}_1 + \left(\mathbf{I} \otimes \mathbf{1}^{\mathsf{T}} \mathbf{N}_{\mathrm{dg}}^{-1} \right) d\mathrm{vec}\, \mathbf{N}_1 \tag{11.36}$$

which simplifies to (11.33).   □

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 $b_{ij} = P \left[\text{absorption in } i \,|\text{starting in } j \right]$ for $i = 1, \ldots, a$ and $j = 1, \ldots, s$. The matrix $\mathbf{B} = \left(b_{ij} \right) = \mathbf{M} \mathbf{N}_1$ (Iosifescu 1980, Thm. 3.3).

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

$$d\mathrm{vec}\, \mathbf{B} = \left(\mathbf{N}_1^{\mathsf{T}} \otimes \mathbf{I} \right) d\mathrm{vec}\, \mathbf{M} + \left(\mathbf{N}_1^{\mathsf{T}} \otimes \mathbf{B} \right) d\mathrm{vec}\, \mathbf{U}. \tag{11.37}$$

Proof Differentiating $\mathbf{B}$ yields

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

Applying the vec operator gives

$$d\mathrm{vec}\, \mathbf{B} = \left(\mathbf{N}_1^{\mathsf{T}} \otimes \mathbf{I} \right) d\mathrm{vec}\, \mathbf{M} + (\mathbf{I} \otimes \mathbf{M}) d\mathrm{vec}\, \mathbf{N}_1. \tag{11.39}$$

Substituting (11.7) for $d\mathrm{vec}\, \mathbf{N}_1$ and simplifying gives (11.37).   □

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 jth unit vector of length s. Thus the derivative of $\mathbf{B}(:, j)$ is

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

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

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

where $\mathbf{e}_i$ is the ith 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

$$\mathbf{q}_a = \mathbf{w} \qquad (11.42)$$

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

$$\mathbf{q}_b = \frac{\mathbf{w} \circ \mathbf{v}}{\mathbf{w}^{\mathsf{T}}\mathbf{v}} \qquad (11.43)$$

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* $\mathbf{w}^{\mathsf{T}}\mathbf{v} = 1$*. Then*

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

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

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 $\mathbf{U}^{\mathsf{T}}$. $\square$

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*

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

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

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

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

Applying the vec operator gives

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

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 ith entry of $\boldsymbol{\eta}_1$, and the expected number of years lost over the distribution of age at death is $\eta^{\dagger} = \boldsymbol{\eta}_1^{\mathsf{T}}\mathbf{f}$. 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 $\eta^\dagger$ responds to changes in mortality.

Differentiating $\eta^\dagger$ gives

$$d\eta^\dagger = \left(d\boldsymbol{\eta}_1^\mathsf{T}\right)\mathbf{B}\mathbf{e}_1 + \boldsymbol{\eta}_1^\mathsf{T}\left(d\mathbf{B}\right)\mathbf{e}_1. \tag{11.50}$$

Applying the vec operator gives

$$d\eta^\dagger = \mathbf{e}_1^\mathsf{T}\mathbf{b}^\mathsf{T}d\boldsymbol{\eta}_1^\mathsf{T} + \left(\mathbf{e}_1^\mathsf{T}\otimes\boldsymbol{\eta}_1^\mathsf{T}\right)d\mathrm{vec}\,\mathbf{B}. \tag{11.51}$$

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

$$d\eta^\dagger = \mathbf{f}^\mathsf{T}\left(\mathbf{I}\otimes\mathbf{1}^\mathsf{T}\right)d\mathrm{vec}\,\mathbf{N}_1 + \left(\mathbf{e}_1^\mathsf{T}\otimes\boldsymbol{\eta}_1^\mathsf{T}\right)$$

$$\left[\left(\mathbf{N}_1^\mathsf{T}\otimes\mathbf{I}\right)d\mathrm{vec}\,\mathbf{M} + \left(\mathbf{N}_1^\mathsf{T}\otimes\mathbf{B}\right)d\mathrm{vec}\,\mathbf{U}\right] \tag{11.52}$$

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

$$\frac{d\eta^\dagger}{d\boldsymbol{\mu}^\mathsf{T}} = \left[\mathbf{f}^\mathsf{T}\left(\mathbf{N}_1^\mathsf{T}\otimes\boldsymbol{\eta}_1^\mathsf{T}\right) + \left(\mathbf{e}_1^\mathsf{T}\mathbf{N}_1^\mathsf{T}\otimes\boldsymbol{\eta}_1^\mathsf{T}\mathbf{B}\right)\right]$$

$$\frac{d\mathrm{vec}\,\mathbf{U}}{d\boldsymbol{\mu}^\mathsf{T}} + \left(\mathbf{e}_1^\mathsf{T}\mathbf{N}_1^\mathsf{T}\otimes\boldsymbol{\eta}_1^\mathsf{T}\right)\frac{d\mathrm{vec}\,\mathbf{M}}{d\boldsymbol{\mu}^\mathsf{T}} \tag{11.53}$$

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

$$\frac{\epsilon\eta^\dagger}{\epsilon\boldsymbol{\mu}^\mathsf{T}} = \frac{1}{\eta^\dagger}\frac{d\eta^\dagger}{d\boldsymbol{\mu}^\mathsf{T}}\mathcal{D}\left(\boldsymbol{\mu}\right). \tag{11.54}$$

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 $\eta^\dagger$, whereas reductions in old age mortality would increase $\eta^\dagger$. 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, $\eta^\dagger$, 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)
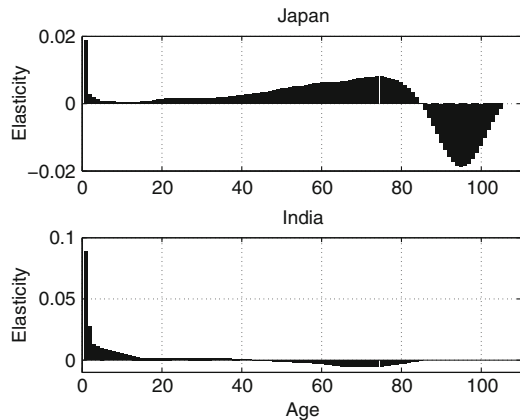
11.4   Ergodic Chains

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 $\mathbf{Z} = \left(\mathbf{I} - \mathbf{P} + \boldsymbol{\pi}\mathbf{1}^\mathsf{T}\right)^{-1}$ (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 $\mathbf{1}^\mathsf{T}\boldsymbol{\pi} = 1$. The sensitivity of $\boldsymbol{\pi}$ is*

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

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 $\mathbf{1}^\mathsf{T}\mathbf{P} = \mathbf{1}^\mathsf{T}$, gives $d\boldsymbol{\pi} = \mathbf{Z}\left[\boldsymbol{\pi}^\mathsf{T} \otimes \left(\mathbf{I}_s - \boldsymbol{\pi}\mathbf{1}^\mathsf{T}\right)\right] d\text{vec}\,\mathbf{P}$. 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 $(\mathbf{I} - \mathbf{P})^\# = \mathbf{Z} - \boldsymbol{\pi}\mathbf{1}^\mathsf{T}$ (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

$$dg(\boldsymbol{\pi}) = \frac{dg}{d\boldsymbol{\pi}^\mathsf{T}} \frac{d\boldsymbol{\pi}}{d\text{vec}^\mathsf{T}\mathbf{P}} \, d\text{vec}\,\mathbf{P} \tag{11.56}$$

Some examples will appear in Sect. 11.5.

11.4.2 The Fundamental Matrix

The fundamental matrix $\mathbf{Z} = \left(\mathbf{I} - \mathbf{P} + \boldsymbol{\pi}\mathbf{1}^\mathsf{T}\right)^{-1}$ 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*

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

Proof From (2.82),

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

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

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

11.4.3   The First Passage Time Matrix

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

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

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*

$$d\text{vec}\,\mathbf{R} = -\left[\mathbf{R}^{\mathsf{T}} \otimes \mathcal{D}\left(\boldsymbol{\pi}\right)^{-1}\right]\mathcal{D}\left(\text{vec}\,\mathbf{I}_s\right)\left(\mathbf{1} \otimes \mathbf{I}_s\right)d\boldsymbol{\pi}$$

$$-\left\{\left[\mathbf{I}_s \otimes \mathcal{D}\left(\boldsymbol{\pi}\right)^{-1}\right] - \left[\mathbf{E} \otimes \mathcal{D}\left(\boldsymbol{\pi}\right)^{-1}\right]\mathcal{D}\left(\text{vec}\,\mathbf{I}_s\right)\right\}d\text{vec}\,\mathbf{Z} \tag{11.61}$$

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

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

Applying the vec operator gives

$$d\text{vec}\,\mathbf{R} = \left[\left(\mathbf{I} - \mathbf{Z} + \mathbf{Z}_{dg}\mathbf{E}\right)^{\mathsf{T}} \otimes \mathbf{I}_s\right]d\text{vec}\left[\mathcal{D}\left(\boldsymbol{\pi}\right)^{-1}\right]$$

$$-\left[\mathbf{I}_s \otimes \mathcal{D}\left(\boldsymbol{\pi}\right)^{-1}\right]d\text{vec}\,\mathbf{Z} + \left[\mathbf{E} \otimes \mathcal{D}\left(\boldsymbol{\pi}\right)^{-1}\right]d\text{vec}\,\mathbf{Z}_{dg}. \tag{11.63}$$

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

$$d\text{vec}\,\mathbf{R} = -\left[\mathbf{R}^{\mathsf{T}}\mathcal{D}\left(\boldsymbol{\pi}\right) \otimes \mathbf{I}_s\right]\left[\mathcal{D}\left(\boldsymbol{\pi}\right)^{-1} \otimes \mathcal{D}\left(\boldsymbol{\pi}\right)^{-1}\right]\mathcal{D}\left(\text{vec}\,\mathbf{I}\right)\left(\mathbf{1} \otimes \mathbf{I}\right)d\boldsymbol{\pi}$$

$$-\left[\mathbf{I} \otimes \mathcal{D}\left(\boldsymbol{\pi}\right)^{-1}\right]d\text{vec}\,\mathbf{Z} + \left[\mathbf{E} \otimes \mathcal{D}\left(\boldsymbol{\pi}\right)^{-1}\right]\mathcal{D}\left(\text{vec}\,\mathbf{I}\right)d\text{vec}\,\mathbf{Z} \tag{11.64}$$

which simplifies to give (11.61).   □

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 π. 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*

$$dK = (\text{vec}\,\mathbf{I}_s)^{\mathsf{T}}\,d\text{vec}\,\mathbf{Z}. \tag{11.65}$$

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

$$dK = \mathbf{1}^{\mathsf{T}}\,(\mathbf{I} \circ d\mathbf{Z})\,\mathbf{1}. \tag{11.66}$$

Applying the vec operator gives

$$dK = \left(\mathbf{1}^{\mathsf{T}} \otimes \mathbf{1}^{\mathsf{T}}\right) \mathcal{D}\,(\text{vec}\,\mathbf{I})\,d\text{vec}\,\mathbf{Z} \tag{11.67}$$

which simplifies to (11.65). □

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

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

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

$$dy = \frac{dy}{d\mathbf{p}^{\mathsf{T}}} \frac{d\mathbf{p}}{d\boldsymbol{\theta}^{\mathsf{T}}} d\boldsymbol{\theta} \tag{11.69}$$

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

Additive compensation For the case of additive compensation, we write

$$p_1(\boldsymbol{\theta}) = p_1(0) + \theta_1 - \frac{\theta_2}{s-1} - \cdots - \frac{\theta_s}{s-1}$$

$$p_2(\boldsymbol{\theta}) = p_2(0) - \frac{\theta_1}{s-1} + \theta_2 - \cdots - \frac{\theta_s}{s-1}$$

$$\vdots \tag{11.70}$$

$$p_s(\boldsymbol{\theta}) = p_s(0) - \frac{\theta_1}{s-1} - \frac{\theta_2}{s-1} - \cdots + \theta_s$$

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

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

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

$$\mathbf{A} = \mathbf{I} - \frac{1}{s-1}\mathbf{C} \tag{11.72}$$

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

$$p_1(\boldsymbol{\theta}) = p_1(0) + \theta_1 - \frac{p_1\theta_2}{1-p_2} - \cdots - \frac{p_1\theta_s}{1-p_s}$$

$$p_2(\boldsymbol{\theta}) = p_2(0) - \frac{p_2\theta_1}{1-p_1} + \theta_2 - \cdots - \frac{p_2\theta_s}{1-p_s}$$

$$\vdots \tag{11.73}$$

$$p_s(\boldsymbol{\theta}) = p_s(0) - \frac{p_s\theta_1}{1-p_1} - \frac{p_s\theta_2}{1-p_2} - \cdots + \theta_s$$

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

Equation (11.73) can be written

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

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

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

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

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

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

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*

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

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

If compensation is proportional, then

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

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

Proof $\mathbf{P}(\boldsymbol{\Theta})$ 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

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

Using the vec operator gives (11.83).   □

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

$$\left.\frac{dy}{d\text{vec}^{\mathsf{T}}\mathbf{P}}\right|_{\text{comp}} = \frac{dy}{d\text{vec}^{\mathsf{T}}\mathbf{P}}\,\frac{d\text{vec}\,\mathbf{P}}{d\text{vec}^{\mathsf{T}}\boldsymbol{\Theta}} \tag{11.85}$$

where $d\text{vec}\,\mathbf{P}/d\text{vec}^{\mathsf{T}}\boldsymbol{\Theta}$ 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 **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 determing **P**.

11.5.1   Biotic Diversity

The stationary distribution $\boldsymbol{\pi}$, 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, $H(\boldsymbol{\pi}) = -\boldsymbol{\pi}^{\mathsf{T}}(\log \boldsymbol{\pi})$, 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

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

Most ecologists, however, would not include bare substrate in a measure of biodiversity, so we define instead a "biotic diversity" $H_b(\boldsymbol{\pi}) = H(\boldsymbol{\pi}_b)$ where

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

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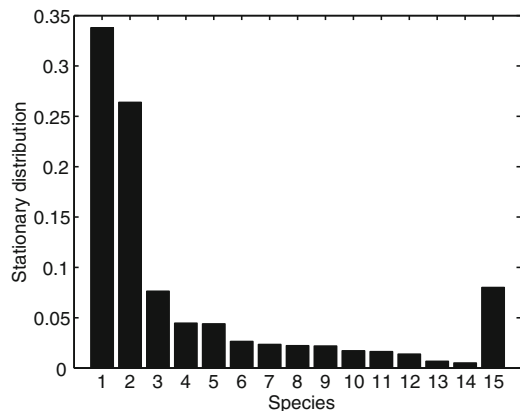
The matrix $\mathbf{G}$, of dimension 14×15, is a 0–1 matrix that selects rows 1–14 of $\boldsymbol{\pi}$. Because $\boldsymbol{\pi}$ is positive, $\|\mathbf{G}\boldsymbol{\pi}\| = \mathbf{1}^{\mathsf{T}}\mathbf{G}\boldsymbol{\pi}$. Differentiating $\boldsymbol{\pi}_b$ gives

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

which simplifies to

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

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

$$\left. \frac{dH_b}{d\text{vec}^{\mathsf{T}}\mathbf{P}} \right|_{\text{comp}} = \underbrace{\frac{dH_b}{d\boldsymbol{\pi}_b^{\mathsf{T}}}}_{1} \underbrace{\frac{d\boldsymbol{\pi}_b}{d\boldsymbol{\pi}^{\mathsf{T}}}}_{2} \underbrace{\frac{d\boldsymbol{\pi}}{d\text{vec}^{\mathsf{T}}\mathbf{P}}}_{3} \underbrace{\frac{d\text{vec}\,\mathbf{P}}{d\text{vec}^{\mathsf{T}}\boldsymbol{\Theta}}}_{4} \tag{11.90}$$

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

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

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

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

$$P[\text{persistence of sp. } i] = p_{ii}$$
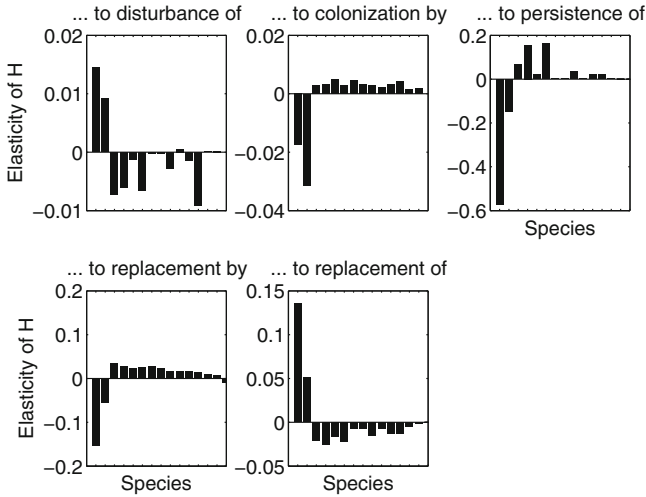
Fig. 11.3 The elasticity of the biotic diversity $H_b(\boldsymbol{\pi})$ calculated over the biotic states of the stationary distribution of the subtidal benthic community succession model of Hill et al. (2004). States 1–14 correspond to species, numbered in decreasing order of abundance in the stationary 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 $\frac{\epsilon H_b}{\epsilon \text{vec}^{\mathsf{T}} \mathbf{P}}$ 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
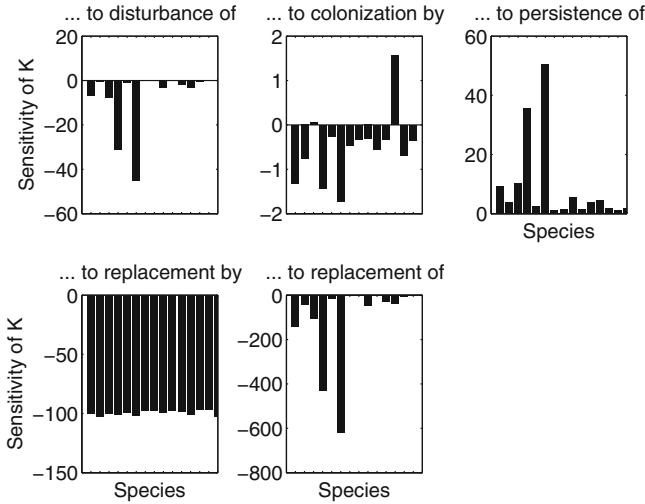
Fig. 11.4 The sensitivity of the Kemeny constant K of the subtidal benthic community succession model of Hill et al. (2004). States 1–14 correspond to species, numbered in decreasing order of abundance in the stationary distribution. State 15 is bare rock, unoccupied by any species. For the identity 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

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

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 $dK/d\text{vec}^{\mathsf{T}}\mathbf{P}$, 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 $\mathbf{N}_{\mathrm{dg}}$, to obtain

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

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

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

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

$$\partial_{\mathbf{N}_{\mathrm{dg}}} \mathrm{vec}\, \mathbf{N}_3 = \left[6\left(\mathbf{N}_1^\mathsf{T}\mathbf{N}_{\mathrm{dg}} \otimes \mathbf{I}\right) + 6\left(\mathbf{N}^\mathsf{T} \otimes \mathbf{N}_{\mathrm{dg}}\right) - 6\left(\mathbf{N}_1^\mathsf{T} \otimes \mathbf{I}\right) \right]$$
$$d\mathrm{vec}\,\mathbf{N}_{\mathrm{dg}}. \tag{11.96}$$

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

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

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

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

$$-36\left(d\mathbf{N}_{\mathrm{dg}}\right) \mathbf{N}_{\mathrm{dg}} \mathbf{N}_1 - 36\mathbf{N}_{\mathrm{dg}} \left(d\mathbf{N}_{\mathrm{dg}}\right) \mathbf{N}_1 + 14\left(d\mathbf{N}_{\mathrm{dg}}\right) \mathbf{N}_1. \tag{11.98}$$

Applying the vec operator yields

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

$$\partial_{\mathbf{N}_{\mathrm{dg}}} \mathrm{vec}\, \mathbf{N}_4 = \left[24\left(\mathbf{N}_1^\mathsf{T}\mathbf{N}_{\mathrm{dg}}^2 \otimes \mathbf{I}\right) + 24\left(\mathbf{N}_1^\mathsf{T}\mathbf{N}_{\mathrm{dg}} \otimes \mathbf{N}_{\mathrm{dg}}\right) + 24\left(\mathbf{N}_1^\mathsf{T} \otimes \mathbf{N}_{\mathrm{dg}}^2\right) \right.$$

$$-36\left(\mathbf{N}_1^\mathsf{T}\mathbf{N}_{\mathrm{dg}} \otimes \mathbf{I}\right) - 36\left(\mathbf{N}_1^\mathsf{T} \otimes \mathbf{N}_{\mathrm{dg}}\right)$$

$$\left. +14\left(\mathbf{N}_1^\mathsf{T} \otimes \mathbf{I}\right) \right] d\mathrm{vec}\,\mathbf{N}_{\mathrm{dg}}. \tag{11.100}$$

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 $\boldsymbol{\eta}_3$, in (11.21) with respect to $\boldsymbol{\eta}_1$ and $\mathbf{N}_1$, to obtain

$$\partial_{\boldsymbol{\eta}_1} \boldsymbol{\eta}_3^\mathsf{T} = \left(d\boldsymbol{\eta}_1^\mathsf{T}\right)\left(6\mathbf{N}_1^2 - 6\mathbf{N}_1 + \mathbf{I}\right) \tag{11.101}$$

$$\partial_{\mathbf{N}_1} \boldsymbol{\eta}_3^\mathsf{T} = 6\boldsymbol{\eta}_1^\mathsf{T}\left(d\mathbf{N}_1\right) + 6\boldsymbol{\eta}_1^\mathsf{T}\mathbf{N}_1\left(d\mathbf{N}_1\right) - 6\boldsymbol{\eta}_1^\mathsf{T}\left(d\mathbf{N}_1\right). \tag{11.102}$$

Applying the vec operator yields

$$\partial_{\boldsymbol{\eta}_1} \boldsymbol{\eta}_3 = \left(6\mathbf{N}_1^2 - 6\mathbf{N}_1 + \mathbf{I}\right)^\mathsf{T} d\boldsymbol{\eta}_1 \tag{11.103}$$

$$\partial_{\mathbf{N}_1} \boldsymbol{\eta}_3 = \left[6\left(\mathbf{N}_1^\mathsf{T} \otimes \boldsymbol{\eta}_1^\mathsf{T}\right) + 6\left(\mathbf{I} \otimes \boldsymbol{\eta}_1^\mathsf{T}\mathbf{N}_1\right) - 6\left(\mathbf{I} \otimes \boldsymbol{\eta}_1^\mathsf{T}\right) \right] d\mathrm{vec}\,\mathbf{N}_1 \tag{11.104}$$

which combine to yield (11.25).

The partial differentials of $\boldsymbol{\eta}_4$ in (11.22) with respect to $\boldsymbol{\eta}_1$ and $\mathbf{N}_1$ are

$$\partial_{\boldsymbol{\eta}_1} \boldsymbol{\eta}_4^\mathsf{T} = d\boldsymbol{\eta}_1^\mathsf{T} \left(24\mathbf{N}_1^3 - 36\mathbf{N}_1^2 + 14\mathbf{N}_1 - \mathbf{I} \right) \tag{11.105}$$

$$\partial_{\mathbf{N}_1} \boldsymbol{\eta}_4^\mathsf{T} = \boldsymbol{\eta}_1^\mathsf{T} \Big[24\,(d\mathbf{N}_1)\,\mathbf{N}_1^2 + 24\mathbf{N}_1\,(d\mathbf{N}_1)\,\mathbf{N}_1 + 24\mathbf{N}_1^2\,(d\mathbf{N}_1)$$

$$- 36\,(d\mathbf{N}_1)\,\mathbf{N}_1 - 36\mathbf{N}_1\,(d\mathbf{N}_1) + 14 d\mathbf{N}_1 \Big]. \tag{11.106}$$

Applying the vec operator to each equation gives

$$\partial_{\boldsymbol{\eta}_1} \boldsymbol{\eta}_4 = \left(24\mathbf{N}_1^3 - 36\mathbf{N}_1^2 + 14\mathbf{N}_1 - \mathbf{I} \right)^\mathsf{T} d\boldsymbol{\eta}_1 \tag{11.107}$$

$$\partial_{\mathbf{N}_1} \boldsymbol{\eta}_4 = \Big\{ 24 \left[\left(\mathbf{N}_1^2 \right)^\mathsf{T} \otimes \boldsymbol{\eta}_1^\mathsf{T} \right] + 24 \left(\mathbf{N}_1^\mathsf{T} \otimes \boldsymbol{\eta}_1^\mathsf{T} \mathbf{N}_1 \right) + 24 \left(\mathbf{I} \otimes \boldsymbol{\eta}_1^\mathsf{T} \mathbf{N}_1^\mathsf{T} \right)$$

$$- 36 \left(\mathbf{N}_1^\mathsf{T} \otimes \boldsymbol{\eta}_1^\mathsf{T} \right) - 36 \left(\mathbf{I} \otimes \boldsymbol{\eta}_1^\mathsf{T} \mathbf{N}_1 \right) + 14 \left(\mathbf{I} \otimes \boldsymbol{\eta}_1^\mathsf{T} \right) \Big\} d\mathrm{vec}\,\mathbf{N}_1 \tag{11.108}$$

which combine to give (11.26).

B   Appendix B: Marine Community Matrix

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

Chapter 12
Sensitivity Analysis of Continuous Markov Chains

Check for updates

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 v_{ij} be the time spent in transient state i by an individual starting in transient state j. Define $\mathbf{N}_k = E\left(v_{ij}^k \right)$ as the matrix whose entries are the kth 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

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

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

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

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

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

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

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

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

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

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

where the square root is taken elementwise. Their derivatives are

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

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

$$d\text{vec}\,CV = \mathcal{D}\,(\text{vec}\,\mathbf{N}_1)^{-1}\,d\text{vec}\,SD$$

$$- \left[(\text{vec}\,SD)^{\mathsf{T}}\,\mathcal{D}\,(\text{vec}\,\mathbf{N}_1)^{-1}\otimes\mathcal{D}\,(\text{vec}\,\mathbf{N}_1)^{-1}\right]$$

$$\times \mathcal{D}\,\left(\text{vec}\,\mathbf{I}_{s^2}\right)\left(\mathbf{1}_{s^2}\otimes\mathbf{I}_{s^2}\right)d\text{vec}\,\mathbf{N}_1 \tag{12.17}$$

(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

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

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

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

Substituting (11.12) for $d\text{vec}\,\mathbf{N}_{\text{dg}}$ 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

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

Apply the vec operator,

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

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

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

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

applying the vec operator,

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

and then using (12.7) and (12.8). The derivative of $SD\left(v_{ij}\right)$ in (12.16) follows from (2.83). The derivative of $CV\left(v_{ij}\right)$ 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 η_j be the time to absorption for an individual currently in transient state j. The vectors of the kth moments of the time to absorption, $\boldsymbol{\eta}_k$, satisfy

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

$$\boldsymbol{\eta}_2^{\mathsf{T}} = (2)\mathbf{1}^{\mathsf{T}}\mathbf{N}_1^2 \tag{12.25}$$

$$\boldsymbol{\eta}_3^{\mathsf{T}} = (6)\mathbf{1}^{\mathsf{T}}\mathbf{N}_1^3 \tag{12.26}$$

$$\boldsymbol{\eta}_4^{\mathsf{T}} = (24)\mathbf{1}^{\mathsf{T}}\mathbf{N}_1^4 \tag{12.27}$$

and in general

$$\boldsymbol{\eta}_k^{\mathsf{T}} = k\boldsymbol{\eta}_{k-1}^{\mathsf{T}}\mathbf{N}_1 \qquad k \geq 2 \tag{12.28}$$

(Iosifescu 1980, Thm. 8.6)

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

$$V(\boldsymbol{\eta}) = \boldsymbol{\eta}_2 - \boldsymbol{\eta}_1 \circ \boldsymbol{\eta}_1 \tag{12.29}$$

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

$$CV(\boldsymbol{\eta}) = \mathcal{D}\left(SD(\boldsymbol{\eta})\right)^{-1}\boldsymbol{\eta}_1 \tag{12.31}$$

with the square root taken elementwise.

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

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

$$d\boldsymbol{\eta}_2 = \left\{2\left[\left(\mathbf{N}_1^{\mathsf{T}}\right)^2 \otimes \boldsymbol{\eta}_1^{\mathsf{T}}\right] + 2\left(\mathbf{N}_1^{\mathsf{T}} \otimes \boldsymbol{\eta}_1^{\mathsf{T}}\mathbf{N}_1\right)\right\} d\mathrm{vec}\,\mathbf{U} \tag{12.33}$$

$$d\boldsymbol{\eta}_3 = \left\{6\left[\left(\mathbf{N}_1^{\mathsf{T}}\right)^3 \otimes \boldsymbol{\eta}_1^{\mathsf{T}}\right] + 6\left[\left(\mathbf{N}_1^{\mathsf{T}}\right)^2 \otimes \boldsymbol{\eta}_1^{\mathsf{T}}\mathbf{N}_1\right]\right.$$
$$\left. + 3\left(\mathbf{N}_1^{\mathsf{T}} \otimes \boldsymbol{\eta}_2^{\mathsf{T}}\mathbf{N}_1\right)\right\} d\mathrm{vec}\,\mathbf{U} \tag{12.34}$$

$$d\boldsymbol{\eta}_4 = \left\{24\left[\left(\mathbf{N}_1^{\mathsf{T}}\right)^4 \otimes \boldsymbol{\eta}_1^{\mathsf{T}}\right] + 24\left[\left(\mathbf{N}_1^{\mathsf{T}}\right)^3 \otimes \boldsymbol{\eta}_1^{\mathsf{T}}\mathbf{N}_1\right]\right.$$
$$\left. +12\left[\left(\mathbf{N}_1^{\mathsf{T}}\right)^2 \otimes \boldsymbol{\eta}_2^{\mathsf{T}}\mathbf{N}_1\right] + 4\left(\mathbf{N}_1^{\mathsf{T}} \otimes \boldsymbol{\eta}_3^{\mathsf{T}}\mathbf{N}_1\right)\right\} d\mathrm{vec}\,\mathbf{U} \tag{12.35}$$

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)

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

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

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

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 $\boldsymbol{\eta}_4$,

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

Applying the vec operator yields

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

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 jth 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 ith 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),

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

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
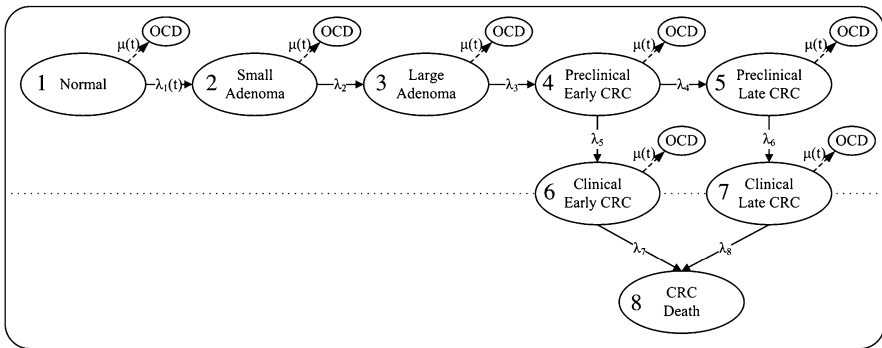
Fig. 12.1 State transition diagram for an absorbing Markov chain model of colorectal cancer (CRC) progression. The model includes 7 transient states based on the stage of development of adenoma (polyps) or cancer, and two absorbing states corresponding to death from CRC and death from other causes (OCD). Transition rates are given by λ_i, and mortality rate from other causes by μ. (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 λ_i are transition rates; μ is the mortality rate from other causes of death. The incidence rate of small adenoma (λ_1) and the mortality rate due to other causes of death (μ) 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 λ_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 λ_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 λ_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 μ 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

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

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

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

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 λ_5 (in column 5 of these matrices), which are positive because λ_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 λ_1, λ_2, and λ_3, the rates of transition from normal to small adenoma, small to large adenoma, and large adenoma to preclinical CRC. The rates λ_2 and λ_3 have large effects on $E(\eta_2)$, and λ_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

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

$$(12.72)$$

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),

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

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 **B** in (12.48), giving the ultimate probability of death from CRC (row 1) or other causes of death (row 2) is

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

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

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

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

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 λ_2, λ_3, and λ_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

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

The sensitivities and elasticities of the probability of contracting clinical CRC are given by the last two rows. These probabilities are highly elastic to λ_1, λ_2 and λ_3. The elasticities to μ 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 λ_i and mortality rates μ_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. Núnez-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.