Theoretical Population Biology 120 (2018) 62–77

Contents lists available at ScienceDirect

Theoretical Population Biology

journal homepage: www.elsevier.com/locate/tpb

Occupancy time in sets of states for demographic models

Gregory Roth *, Hal Caswell

Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam, Netherlands

ARTICLE INFO

Article history:
Received 6 April 2017

Keywords:
Matrix population models
Absorbing Markov chains
Multistage models
Demography
Occupancy time
Longevity
Southern fulmar

ABSTRACT

As an individual moves through its life cycle, it passes through a series of states (age classes, size classes, reproductive states, spatial locations, health statuses, etc.) before its eventual death. The occupancy time in a state is the time spent in that state over the individual's life. Depending on the life cycle description, the occupancy times describe different demographic variables, for example, lifetime breeding success, lifetime habitat utilisation, or healthy longevity.

Models based on absorbing Markov chains provide a powerful framework for the analysis of occupancy times. Current theory, however, can completely analyse only the occupancy of single states, although the occupancy time in a set of states is often desired. For example, a range of sizes in a size-classified model, an age class in an age×stage model, and a group of locations in a spatial stage model are all sets of states.

We present a new mathematical approach to absorbing Markov chains that extends the analysis of life histories by providing a comprehensive theory for the occupancy of arbitrary sets of states, and for other demographic variables related to these sets (e.g., reaching time, return time). We apply this approach to a matrix population model of the Southern Fulmar (*Fulmarus glacialoides*). The analysis of this model provides interesting insight into the lifetime number of breeding attempts of this species.

Our new approach to absorbing Markov chains, and its implementation in matrix oriented software, makes the analysis of occupancy times more accessible to population ecologists, and directly applicable to any matrix population models.

1. Introduction

The life of an individual is a sequence of events. Birth and death are events common to every individual, but the sequence between birth and death – unique to each individual – consists of a potentially endless list of random events (e.g., surviving, developing, mating, reproducing, growing, dispersing, moving among social or occupational classes, or changing health status). Each event corresponds to a change in the state of the individual, resulting in a stochastic pathway that ends eventually in death. A central role in the analysis of these pathways is played by the concepts of *occupancy time* (the time spent in, or the number of visits to, a state over the individual's lifetime). Occupancy is a property of the stochastic pathway of an individual, and occupancy times define the time spent in each of the possible states during the lifetime. In particular, the longevity of an individual is measured by the sum of all these occupancy times. The interpretation of occupancy times depends on the identity of the transient states and the nature of the absorption. Thus, when the states are health status, occupancy time represents years of life while healthy, not healthy, disabled, etc. When the states are spatial locations, occupancy time represents time spent in different places. When the states are marital status, occupancy times measure the part of the lifetime spent single, married, divorced, remarried, etc. When the states are employment status, or breeding activities, or any other interesting categorisation of individuals, the interpretation follows the same lines. As for absorption, it may be death, in which case occupancy time is a "lifetime" measure in the literal sense. But absorption can be defined as the first entrance to some state or set of states (e.g., occurrence of first breeding, or graduation, or metamorphosis, or hospitalisation, etc.).

Because the pathways are stochastic, occupancy time is a random variable. It is often described by its mean (e.g., life expectancy, expected lifetime reproduction). However, some individuals will live longer and some shorter, than the mean; some will mature later and some earlier than the mean. To characterise this variation, the probability distribution of occupancy time, or at least its moments, must be considered.

Models based on absorbing Markov chains provide a powerful framework for the analysis of occupancy times. An absorbing Markov chain describes the fate of an individual – under the assumption that the future of the individual, given its present, is independent of its past – evolving in a set of states and being eventually absorbed by the death state. The states may refer to

* Corresponding author.
E-mail address: g.roth@uva.nl (G. Roth).

https://doi.org/10.1016/j.tpb.2017.12.007

developmental states, physiological measures, behaviour types, locations, and so on. The set of transition rates between these states – described by a *transition matrix* – defined an absorbing Markov chain. As a population projection matrix describes the fate of a population, the transition matrix describes the fate of individuals in a population, and often is one component of a population projection matrix. The mathematical theory of absorbing Markov chain provides formulae for basic descriptive quantities of the absorbing Markov chain, based on its fundamental matrix (see e.g. Iosifescu (1980) and Kemeny and Snell (1961) for a mathematical perspective, and Caswell (2001) for a demographical perspective). Applied to demographic models, this theory provides simple and direct formulae for the probability distribution, the mean, variance, and all moments of longevity, the distribution of age or stage at death, the survivorship and mortality functions, causes of death, and a variety of measures of life disparity (e.g., Feichtinger, 1971; Cochran and Ellner, 1992; Caswell, 2001, 2006, 2009; Tuljapurkar and Horvitz, 2006; Horvitz and Tuljapurkar, 2008; Van Raalte and Caswell, 2013). Powerful sensitivity analyses are available for all these quantities (Caswell, 2006, 2009, 2011b, 2013).

Current theory, however, can completely analyse only the occupancy time of single states and the occupancy time of the whole state space. Our goal is to extend the analysis of life histories by providing a comprehensive theory for the occupancy of arbitrary sets of states. One type of set is a collection of states deemed biologically relevant for some purpose; we call these *super-states*. For example, a model based on reproductive behaviour might include states describing many details of the success, failure, timing, and number of offspring produced by breeding, but one might want to investigate the super-states created by aggregating these into "successful breeding" and "non-successful breeding" sets. A spatial model might describe habitats along an altitudinal gradient, and one might want to aggregate in order to compare the occupancy of low altitude and high altitude sites. A medical demography study might distinguish a variety of health conditions and treatments, but one might want to compare the occupancy of all states requiring hospitalisation and those not requiring hospitalisation. The utility of super-states will increase as more matrix models are created from the growth and survival kernels of integral projection models (e.g., Ellner et al., 2016). These matrices typically contain hundreds of size classes, no one of which is of particular interest, but sets of which (e.g., all trees large enough to reach the forest canopy) are of great interest.

A second type of sets of states arises in the context of multistate (or megamatrix) or hyperstate models (e.g., Rogers, 1975; Lebreton, 1996; Pascarella and Horvitz, 1998; Tuljapurkar et al., 2003; Roth and Caswell, 2016) in which individuals are classified by two or more criteria (age and stage, stage and location, etc.). One may want to analyse the occupancy of sets of states defined by integrating over one of these criteria; we call these *marginal sets*. For example, in a stage×size-classified model, the marginal set associated with the *juvenile* stage is the set containing the juvenile stage, integrated over all possible sizes.

The extension of occupancy time calculations to sets of states may seem trivial because the occupancy time in a set is the sum of the occupancy times in each state belonging to this set. Therefore, the mean occupancy time in a set is the sum of the means of the occupancy times in each state. However, this observation does not hold for the variance, for any higher moments, or for the probability distribution, because occupancy times in single states are not *independent* from each other. There are few analyses of the occupancy time in set of states, but they only focus on specific aspects of it. For example, Steiner and Tuljapurkar (2012) provide formulae for the mean and variance of the reproductive output using the joint generating function of the single state occupancy times. The reproductive output of an individual is closely related to the occupancy time in the set of reproductive states (both are equal when fertility rates are ones in each reproductive state). Caswell (2011a) provides similar formulae using the theory of Markov chain with reward. The same theory is used by Caswell and Kluge (2015) to calculate the moments of lifetime accumulation of economic variables, which are also closely related to occupancy times. However, these studies do not provide the probability distribution of occupancy time in a set of states. In the mathematical literature, Sericola (2000) provides an iterative formula for the probability distribution of the partial (i.e. up to a fixed time) occupancy time in a set of states, but does not provide a closed formula for the total occupancy time. Here, we present a comprehensive approach to calculate the any moment and the probability distribution of the occupancy time in arbitrary sets of states. Our approach relies on the construction of a *sub Markov chain*, which describes the original Markov chain viewed through a filter that allows one to see only the states in the set of interest. As a consequence, all the statistics of the occupancy time in the set of interest may be calculated with the existing theory of absorbing Markov chain (Iosifescu, 1980), applied to the sub Markov chain.

The construction and the analysis of the sub chain extends the classical theory of absorbing Markov chain by providing not only several measures related to the occupancy of sets of states but also forms a basis for further calculations of measures related to sets of states, including

- The set occupancy time. Depending on the life cycle description, set occupancy times describe different demographic variables (e.g., lifetime breeding attempts in a model of reproductive behaviour, or lifetime habitat utilisation in a spatial model). We provide for the probability distribution, mean, and variance of the occupancy times.
- The correlation between the occupancy times in two different sets. This is an indicator of how the two sets are connected in the life cycle. As a particular case, we provide, for the first time, a formula for the correlation between the occupancy time in a state and the longevity of an individual. Depending on the life cycle description, this formula gives the correlation between different demographic variables and longevity (e.g., lifetime breeding attempts and longevity, lifetime reproduction and longevity, time to maturation and longevity).
- Properties of winners and losers. Relative to a particular target set, a *winner* is an individual that enters the set at least once in its life, and a *loser* is an individual that never enters the set. In a model classifying individuals by their developmental state, the winners might represent those individuals that successfully mature, and the losers those that do not. We provide the probability of becoming a winner, the distribution, mean, and variance of the time required for a winner to reach the set, and the longevity of a loser. After its first success, a winner may leave the set and never return, or it may return at some future time. We obtain the probability that a winner returns, and for those that do return, the probability distribution, mean, and variance of the return time.

Table 1 lists the demographic results to be presented and the equations in which they are derived. All the results are obtained directly from a single matrix, describing the transition probabilities among transient states. This matrix is obtainable from most population projection matrices (Caswell, 2001). Despite the large number of matrices and sometimes complicated expressions that appear in our derivations, our results are easily computed in matrix-oriented software. In the Supplementary Material, we provide the MATLAB code for calculating all of the demographic results listed in Table 1.