bioRxiv preprint doi: https://doi.org/10.1101/188276; this version posted September 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Drivers of diversity in individual life courses: Sensitivity of the population entropy of a Markov chain

Ulrich K. Steiner,
Max Planck Odense Center on the Biodemography of Aging, Denmark, and Department of Biology, University of Southern Denmark, Odense, Denmark
usteiner@biology.sdu.dk, +4565502753
Shripad Tuljapurkar,
Department of Biology, Stanford University, USA

September 13, 2017

Abstract

Individuals differ greatly in their life courses, but how such diversity is generated, how it has evolved and how it is maintained is less understood. However, such understanding is crucial to comprehend evolutionary and ecological population dynamics. In structured populations individuals diversify by transitioning through discrete stages that end with death. Such stage transitions can be described by a Markov chain, and the rate of diversification of these stage-sequences individuals experience with increasing age can be quantified by the population entropy of a Markov chain. Here, we derive sensitivities of the population entropy of a Markov chain in order to identify which transitions are generating most diversification among life courses, i.e. what drives diversification. We then use these derived sensitivities to reveal potential selective forces on the dynamics of life courses. We illustrate our exact result for the sensitivity of the entropy using an example on a seabird population that is structured by reproductive stages. In our example on the

bioRxiv preprint doi: https://doi.org/10.1101/188276; this version posted September 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

murres, the most important drivers of diversification of life courses are not linked to the most sensitive transitions influencing population growth, which suggests that observed diversification in life courses are rather neutral than adaptive. However, this interpretation has to approached with caution since we only explored one of many solutions for the constraints among transition probabilities. We believe to only start to understand how individual level dynamics shape population level evolutionary and ecological dynamics, and many more exciting discoveries await us in exploring underlying dynamics of age- and stage-structured populations.

Introduction

In any population we observe a great diversity in life courses among individuals (Tuljapurkar et al., 2009; Steiner and Tuljapurkar, 2012). How such diversity is generated, how it has evolved and how it is maintained is of central interest to us as biodemographers, evolutionary ecologists, and population biologist, because such knowledge furnish understanding of evolutionary and ecological population dynamics (Endler, 1986; Hartl and Clark, 2007). We know that genetic variability, environmental variability and their interaction contribute to individual differences in life courses, as do stochastic events (e.g. demographic stochasticity) (Lande et al., 2003; Finch and Kirkwood, 2000; Melbourne and Hastings, 2008). Population geneticists and quantitative geneticists have provided us with theoretical and empirical understanding about changes in genotype frequencies and trait distributions, such changes in distributions aim at the population level dynamics, but the underlying individual level dynamics that lead to the diversity in individual life courses are less explored (Lande et al., 2003). Besides the lack of understanding drivers of individual level dynamics, we do not know to what degree such individual level dynamics are adaptive, mal-adaptive or neutral

bioRxiv preprint doi: https://doi.org/10.1101/188276; this version posted September 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(Lenormand et al., 2009).

Whatsoever the actual underlying driver of the diversity in life courses might be, in any structured population the diversity of individual life courses is generated by transitions through discrete stages that end with death (Caswell, 2001). Individuals are born into one, or one of several, discrete stage and subequently transition to one of several discrete stages at each observation. Hence, an individual's life course can be be described as a stage sequence, which we call a trajectory (Tuljapurkar et al., 2009). The stages, each individual at each observation attains, can comprise developmental stages including levels of breeding success, morphological stages such as size, behavioral stages such as feeding or mating activity, physiological stages such as condition, gene expression stages such as transcription factor expression, epigenetic stages such as the methylation stage, or geographically defined spatial location.

Say there are s stages, so that the population is described by the frequency with which we observe each stage at each observation, underlying these population level frequencies lay the individual trajectories. Note, that we can generate population stage frequencies with very different underlying individual trajectories. For instance, heterogeneous trajectories with individuals frequently changing among stages can lead to very similar population level frequencies as can a few trajectories with low level of dynamics (Hernandez-Pacheco and Steiner, 2017). Therefore the population stage frequencies, which are also the focus of population genetics and quantitative genetics, does not reveal underlying individual level stage dynamics. Here, we consider only systems whose dynamics – changes in stage frequency – are described by a Markov chain, i.e., there is a probability $p_{ij} \geq 0$ that an individual changes its stage from stage j to stage i, for every possible pair of stages. The notation here is similar to Caswell (2001); Hill et al. (2004).

In many systems initial stage distribution at birth is centered to-

bioRxiv preprint doi: https://doi.org/10.1101/188276; this version posted September 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

wards one or a few stages, e.g. individuals start life in a newborn stage or immature stage if developmental and reproductive stages are tracked. With increasing age, individuals transition through stages described by a Markov chain, individual trajectories diversify. We can quantify the rate of diversification of these trajectories by the population entropy of a Markov chain. This population entropy has breviously been used as a measure of the rate of diversification among life courses, also termed dynamic heterogeneity among alternative individual trajectories (Tuljapurkar et al., 2009; Steiner and Tuljapurkar, 2012) (other related concepts are termed individual heterogeneity (Caswell, 2009)). This quantification of population entropy is based on the transition probabilities p_{ij} of the Markov chain, that determines the diversity in trajectories. Here we aim at revealing the drivers of such diversity by seeking for the sensitivity of the population entropy of a Markov chain, that is, we perturb a set of transition probabilities and see how entropy responds to such perturbation.

The sensitivity of the population entropy reveals to us which transitions are generating most diversification among life courses, but it does not provide any understanding whether such diversification might be under selection, i.e. whether it is adaptive or neutral. To investigate potential adaptive features of such diversification we can compare the sensitivity of the population entropy to the sensitivity with respect to the population growth rate, λ. This latter sensitivity to λ is linked to the evolutionary forces acting on these transition probabilities, because population growth rate quantifies fitness and selection is expected to have acted most strongly on the most sensitive transition probabilities (Caswell, 2001). After calculating both measures of sensitivity (sensitivity to entropy and sensitivity to λ) we can correlate the two to reval potential adaptive drivers of diversification among life courses. If we find a positive correlation diversification should be adaptive, or diversification is rather neutral if we do not see any relationship between the two

bioRxiv preprint doi: https://doi.org/10.1101/188276; this version posted September 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

measures of sensitivities.

Our aim here is two fold, first we derive sensitivities of the population entropy of a Markov chain in order to identify drivers of diversification, and second we use these derived sensitivities to reveal potential selective forces on the dynamics of life histories. We describe such sensitivities for ergodic markov chains, and markov chains with absorbing stages. Classical population projection matrix models that include reproduction (e.g. Lefkovitch or Leslie population matrix models) first need to be transformed into a Markov chain before we can estimate the population entropy. We can achieve this transformation as described by Tuljapurkar (1982) (Appendix). We illustrate our exact result for the sensitivity of the entropy using an example on a seabird population, the Thick-billed murre on Coats Island, Canada (Gaston et al., 1994; Steiner and Gaston, 2005). This population is structured by reproductive stages.

Our results have the virtue that they only require the dominant eigenvalue and corresponding eigenvectors of non-negative matrices – this is numerically straightforward and well-conditioned, unlike the computation of all subdominant eigenvalues. Our approach is therefore applicable to many structured populations.

When the population is ergodic (actually, irreducible and aperiodic) there is a stationary (or equivalently, equilibrium) frequency distribution over the possible stages: a vector $\mathbf{w}$ whose elements w_i are the frequencies of stages $i = 1, \ldots, s$. A stage's equilibrium frequency also equals the fraction of times that an individual is expected to be in that stage, if we make many repeated observations. Entropy $H(\mathbf{P})$ quantifies the diversity in individual

bioRxiv preprint doi: https://doi.org/10.1101/188276; this version posted September 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

trajectories described by the Markov chain:

$$H(\mathbf{P}) = -\sum_{j=1}^{s} w_j \sum_{i=1}^{s} p_{ij} \log p_{ij}, \tag{1}$$

$$= -\mathbf{e}^T \left(\mathbf{P} \circ \log(\mathbf{P})\right) \mathbf{w}. \tag{2}$$

Here $\mathbf{P}$ is a matrix of the Markov chain transition probabilities p_{ij}. The second line above is useful numerically and analytically: the superscript T indicates a transpose; $\mathbf{e}$ is a vector whose entries all equal 1; the Hadamard product ($\circ$) is elementwise so that for matrices $\mathbf{P}, \log(\mathbf{P})$ of equal size with elements $p_{ij}, log(p_{ij})$ respectively the matrix $\mathbf{P} \circ \log(\mathbf{P})$ is of same size and has ij element equal to $p_{ij}\, log(p_{ij})$.

We start with deriving sensitivities for an ergodic chain (irreducible, non-absorbing), by asking what happens if we make a small change in the transition probabilities so that $\mathbf{P}$ becomes $\mathbf{P} + \epsilon\, \mathbf{B}$ (for small positive ϵ). Throughout this paper, we consider only perturbations that leave unchanged the signature of the Markov chain: i.e., whenever $p_{ij} = 0$ we keep $b_{ij} = 0$. Then the entropy must change from $H(\mathbf{P})$ to say $H(\mathbf{P}) + \epsilon\, H_1$. Then H_1 is the sensitivity of the entropy. We obtain here an exact analytical expression for this sensitivity.

Thereafter, we answer the analogous question for a Markov chain that has at least one "absorbing" stage. To see why this is different, suppose death is the absorbing stage so that an individual wanders among the non-absorbing stages until it dies. Conditional on being alive, we expect that there is a quasi-stationary distribution over the non-absorbing stages, if we can find appropriate conditional Markov transition probabilities. Darroch and Seneta (1967) show that we can, providing that absorption takes a long time; see also Matthews (1970). The entropy of this conditional Markov chain measures the dynamic heterogeneity of individual trajectories until death.

bioRxiv preprint doi: https://doi.org/10.1101/188276; this version posted September 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Our contributions are an exact result for the sensitivity of the entropy of an ergodic chain and absorbing Markov chains. Comparing the sensitivities between the two types of markov chains (ergodic and absorbing) from the same system can then be used to evaluate the contribution of survival differences among individuals on the diversity of stage trajectories (Hernandez-Pacheco and Steiner, 2017).

Sensitivity of Entropy: Ergodic Chains

Changing Transition Probabilities

The starting point is a population described by a specified matrix $\mathbf{P}$ of transition probabilities; we assume the chain is irreducible and aperiodic, hence ergodic. An ergodic population is characterized by its asymptotic dynamics being independent of the starting conditions. Here, we are mainly interested in such ergodicity since our focus is on revealing underlying processes, i.e. the drivers of diversity in life courses, than on initial conditions a population starts at. For such ergodic models the stationary frequency is an eigenvector, $\mathbf{P}\,\mathbf{w} = \mathbf{w}$. Transition probabilities out of each stage sum to unity, so $\mathbf{e}^T\,\mathbf{P} = \mathbf{e}^T$. We compute the fundamental matrix, which has also been described as the stage duration matrix (Steiner et al., 2012).

$$\mathbf{Z} = \left[\mathbf{I} - \left(\mathbf{P} - \mathbf{w}\,\mathbf{e}^T\right)\right]^{-1}, \tag{3}$$

where $\mathbf{I}$ is the identity matrix.

Now perturb the transition probabilities to $\mathbf{P}+\epsilon\,\mathbf{B}$, so that transition probability p_{ij} changes to $p_{ij} + \epsilon b_{ij}$. Clearly we must have

$$\mathbf{e}^T\,\mathbf{B} = \mathbf{0}^T, \text{ i.e., columns sum to zero.} \tag{4}$$

bioRxiv preprint doi: https://doi.org/10.1101/188276; this version posted September 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

This means that changes in the transition probabilities are necessarily constrained; some biologically distinct ways of achieving this constraint are discussed by Caswell (2001).

Then following Schweitzer (1968) the stationary frequencies change to $\mathbf{w} + \epsilon\,\mathbf{y} + \epsilon^2\mathbf{y}_2 + O(\epsilon^3)$ where $\mathbf{e}^T\mathbf{y} = \mathbf{e}^T\mathbf{y}_2 = 0$

$$\mathbf{y} = \mathbf{Z}\,\mathbf{B}\,\mathbf{w}, \tag{5}$$

$$y_i = \sum_{k=1}^{s}\sum_{m=1}^{s} Z_{im}\,b_{mk}\,w_k. \tag{6}$$

The more involved expression for $\mathbf{y}_2$ is found in Kato (1966). So the vector $\mathbf{y}$ from equation (5) comprises the time an individual spends in each stage given its current stage (stage duration matrix, $\mathbf{Z}$), the product with the perturbation matrix $\mathbf{B}$ then determines the change in time each individual spends in each stage given its current stage, and finally the multiplication with the stable stage distribution $\mathbf{w}$ quantifies how many individuals (or more precisely what proportion of individuals) are affected by the change in time they spent in each stage, i.e. how many individuals are affected by how much time they spend in each stage due to the perturbation, which is exactly how much change in the stationary stage distribution is caused by the perturbation.

bioRxiv preprint doi: https://doi.org/10.1101/188276; this version posted September 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Sensitivity of Entropy

From equation (1) (and the Appendix) the entropy of the perturbed Markov chain is

$$H(\mathbf{P} + \epsilon\,\mathbf{B}) = H(\mathbf{P}) + \epsilon\, H_1 + \epsilon^2\, H_2 + O(\epsilon^3), \tag{7}$$

$$H_1 = -\sum_{i=1}^{s}\sum_{j=1}^{s} \left[w_j\, b_{ij}\, \log p_{ij} + y_j\, p_{ij}\, \log p_{ij}\right], \tag{8}$$

$$= -\mathbf{e}^T\, \left[\mathbf{B} \circ \log(\mathbf{P})\, \mathbf{w} + \mathbf{P} \circ \log(\mathbf{P})\, \mathbf{y}\right], \tag{9}$$

$$H_2 = -\mathbf{e}^T\, \left[(1/2)\, \mathbf{B} \circ \mathbf{B}\, \mathbf{w} + \mathbf{B} \circ \log(\mathbf{P})\, \mathbf{y} + \mathbf{P} \circ \log(\mathbf{P})\, \mathbf{y}_2 + \mathbf{P} \circ \mathbf{B}\, \mathbf{y}\right]. \tag{10}$$

Here H_1 is the sensitivity to the population entropy we seek. The second-order change in entropy (essentially the second derivative) is H_2.

Special Case: Perturbing a Maximum Entropy chain

A chain with maximum entropy has transition matrix elements $p_{ij} = (1/s)$ where, as before, s is the number of stages (Tuljapurkar et al., 2009). Clearly $\mathbf{w}$ has every element equal to $(1/s)$ and we can write

$$\mathbf{P} = \mathbf{w}\, \mathbf{e}^T. \tag{11}$$

The entropy of this chain is just $H = \log s$ (see also Tuljapurkar et al. (2009)). The chain's fundamental matrix (see (3)) is just $\mathbf{Z} = \mathbf{I}$, which means that when we perturb the chain to $\mathbf{P} + \epsilon\,\mathbf{B}$ the eigenvector $\mathbf{w}$ becomes (see (5)) just $\mathbf{w} + \epsilon\,\mathbf{y}$ with $\mathbf{y} = \mathbf{B}\mathbf{w}$. The second-order perturbation of $\mathbf{w}$ is zero (i.e., $\mathbf{y}_2 = 0$).

The sensitivity of this chain is zero! To see that this is true in our

bioRxiv preprint doi: https://doi.org/10.1101/188276; this version posted September 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

equations, observe that in (9) we have

$$\begin{aligned} \mathbf{B} \circ \log(\mathbf{P}) &= log(1/s)\,\mathbf{B}, \\ \mathbf{P} \circ \log(\mathbf{P})\,\mathbf{y} &= (1/s)\,\log(1/s) \circ \mathbf{E}\,\mathbf{y} = (1/s)\,\log(1/s) \circ \mathbf{E}\,\mathbf{B}\mathbf{w}, \end{aligned} \tag{12}$$

where $\mathbf{E}$ is a matrix with all elements equal 1. Hence both terms in H_1 (9) are proportional to $\mathbf{e}^T\,\mathbf{B}$ – but this has to be zero for any possible perturbation, so $H_1 = 0$. More generally, sensitivity is just a (complicated) derivative of entropy and since we start with maximum entropy it must be true that any derivative of the entropy is zero (that's what defines a maximum).

So what about H_2 in (10)? Note that here $by_2 = 0$, and that the arguments in (12) imply that the only surviving term in (10) is

$$H_2 = -\mathbf{e}^T\,[(1/2)\,\mathbf{B} \circ \mathbf{B}\,\mathbf{w}] = -\frac{1}{2s}\sum_i \sum_j B_{ij}^2. \tag{13}$$

Thus perturbing a maximum entropy chain with transition matrix $\mathbf{P}$ by the constrained matrix $\epsilon\,\mathbf{B}$ always yields a reduced entropy

$$H(\mathbf{P} + \epsilon\,\mathbf{B}) = H(\mathbf{P}) + \epsilon^2\,H_2 = \log s - \frac{\epsilon^2}{2s}\sum_i \sum_j B_{ij}^2,$$

to order ϵ^3.

Sensitivity of Entropy: Chains with Absorbing stages

Transition Probabilities with Absorption

We consider just one absorbing stage – multiple absorbing stages are easily dealt with (Matthews, 1970). Let's say the absorbing stage (think "death")

bioRxiv preprint doi: https://doi.org/10.1101/188276; this version posted September 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

is stage s so that stages 1 to $(s-1)$ are the transient (i.e., "alive") stages. The transition probability matrix must have the form

$$\mathbf{P} = \begin{pmatrix} \mathbf{Q} & \mathbf{0} \\ \boldsymbol{\mu}^T & 1 \end{pmatrix}, \tag{14}$$

with absorption (death) probabilities given by the elements μ_i of vector $\boldsymbol{\mu}$:

$$\mu_i = 1 - \sum_{j=1}^{(s-1)} p_{ji} = 1 - \sum_{j=1}^{(s-1)} q_{ji}.$$

Matrix $\mathbf{Q}$, describes the transition probabilities among the life stages, summing over the columns of $\mathbf{Q}$ gives the survival probability of each stage. Conditional on non-absorption (i.e., being alive), the transition probabilities among the $(s-1)$ transient stages (Darroch and Seneta, 1967) are the entries in the $(s-1) \times (s-1)$ matrix

$$\mathbf{R} = \frac{1}{\rho}\, \widehat{\mathbf{v}}\, \mathbf{Q}\, \widehat{\mathbf{v}}^{-1}, \tag{15}$$

where $0 < \rho < 1$ is the dominant eigenvalue of $\mathbf{Q}$, $\mathbf{v}$ with elements v_i is the corresponding left eigenvector,

$$\mathbf{v}^T\, \mathbf{Q} = \rho\, \mathbf{v}^T,$$

and the diagonal matrix

$$\widehat{\mathbf{v}} = \mathrm{diag}(\mathbf{v}).$$

The ij element of matrix $\mathbf{R}$ is $v_i q_{ij}/(\rho v_j)$; clearly, the columns of $\mathbf{R}$ sum to 1, so this is a Markov matrix, while matrix $\mathbf{Q}$ is not. So what we have done in (15) is to transfrom the transient (absorbing stage transition) matrix $\mathbf{Q}$ to a Markov chain $\mathbf{R}$. Let $\mathbf{w}$ be the right eigenvector of $\mathbf{Q}$ corresponding

bioRxiv preprint doi: https://doi.org/10.1101/188276; this version posted September 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

to its dominant eigenvalue, normalized so that $(\mathbf{v}^T\mathbf{w}) = 1$. The equilibrium frequency distribution of the conditional process governed by $\mathbf{R}$ is given by the products $(w_i v_i), i = 1 \ldots (s-1)$.

We can measure the dynamic heterogeneity of individual trajectories while they are still alive by the entropy of the conditional process (see Appendix),

$$\begin{aligned} H(\mathbf{P}) &= H(\mathbf{Q}), \\ &= -\sum_{j=1}^{(s-1)} w_j\, v_j \sum_{i=1}^{(s-1)} r_{ij}\, \log r_{ij}, \\ &= \log \rho - \frac{1}{\rho} \sum_{i=1}^{(s-1)} \sum_{j=1}^{(s-1)} v_i\, w_j\, q_{ij}\, \log q_{ij}. \end{aligned} \tag{16}$$

Perturbing an Absorbing Chain

We now want the effect on the population entropy of small changes in the transition probabilities of the Markov chain. In (14), consider simply changes in the transient matrix $\mathbf{Q}$ to $\mathbf{Q} + \epsilon\,\mathbf{B}$. It is easy to see how this changes the full matrix $\mathbf{P}$. These changes will alter $\rho, \mathbf{v},$ and $\mathbf{w}$ to $\rho + \epsilon\nu, \mathbf{v} + \epsilon\mathbf{x}, \mathbf{w} + \epsilon\mathbf{y}$, respectively. Here we give explicit formulas to compute these changes and in the next subsection show how these are used to compute the sensitivity of entropy we seek.

Recalling that $(\mathbf{v}^T\mathbf{w}) = 1$, we have the well-known (see e.g., Caswell (2001)) fact that

$$\nu = \mathbf{v}^T\, \mathbf{B}\, \mathbf{w}. \tag{17}$$

bioRxiv preprint doi: https://doi.org/10.1101/188276; this version posted September 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

We define two new matrices:

$$\mathbf{D}_1 = \mathbf{I} - \mathbf{w}\,\mathbf{v}^T, \tag{18}$$

$$\mathbf{Z}_1 = \frac{1}{\rho}\left[\mathbf{I} - \left(\frac{\mathbf{Q}}{\rho} - \mathbf{w}\,\mathbf{v}^T\right)\right]^{-1}. \tag{19}$$

Then we have (see Appendix) the less well-known results,

$$\mathbf{y} = \mathbf{Z}_1\,\mathbf{D}_1\,\mathbf{B}\,\mathbf{w}, \tag{20}$$

$$\mathbf{x}^T = \mathbf{v}^T\,\mathbf{B}\,\mathbf{D}_1\,\mathbf{Z}_1. \tag{21}$$

The interpretation of $\mathbf{y}$ in (20) is similar to the one in equation (5), i.e. how many individuals are affected by how much (more or less) time they spend in each stage due to the perturbation, which equals how much change in the stationary stage distribution is caused by the perturbation, except here (20) this change is based on the absorbing (transient) transition matrix.

Sensitivity of Entropy for an Absorbing Chain

The last step is to compute the difference between the entropy of the perturbed chain ($H(\mathbf{Q})$) and the original chain,

$$H(\mathbf{Q} + \epsilon\,\mathbf{B}) = H(\mathbf{Q}) + \epsilon\,H_1. \tag{22}$$

The sensitivity H_1 is given (see Appendix) by

$$\begin{aligned} H_1 = \nu &\left(\frac{1}{\rho}\right)\,[(1 + \log\rho - H(\mathbf{Q})] \\ &- \left(\frac{1}{\rho}\right)\sum_{i,j=1}^{(s-1)}\left[(x_i w_j + v_i y_j)\,q_{ij}\,\log q_{ij} + v_i w_j\,b_{ij}\,\log q_{ij}\right]. \end{aligned} \tag{23}$$

bioRxiv preprint doi: https://doi.org/10.1101/188276; this version posted September 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Illustrative example sensitivity of entropy: The Thick-billed murre

To illustrate our exact result for the sensitivity of the population entropy of a Markov chain, we first built a stage-structured matrix population model using longitudinal mark resighting data on a highly philopatric and colonial seabird species, the Thick-billed murre *(Uria lomvia)* (Gaston et al., 1994; Steiner and Gaston, 2005). After parameterizing the population projection matrix based on the longitudinal data, we transformed this matrix in order to generate a Markov chain, as described by Tuljapurkar (1982) (Appendix). Here we present the results on population entropy (ergodic chain) of the resulting Markov chain and discuss its implications.

Structured population model of the Thick-billed Murre

To parameterize the stage-structured matrix model, we used data on 1984 individual seabirds, Thick-billed murres, banded between 1981 and 2010, on Coats Island, Nunavut, Canada ($62°30'N$, $83°00'W$). Resightings (band readings) have been made between 1991 and 2011 in the colony over each breeding season. When a bird was resighted its breeding status was recorded as a) immature, I, birds prior to any breeding attempt; b) egg, E, egg laid but not hatched; c) hatch, H, chick hatched but not fledged; d) fledge, F, chick fledged, i.e. chick disappeared >=10 days after hatching; or e) unknown, U, when the breeding outcome was not known. Birds are born into an immature stage (I) and they remain in that stage until they are three years old (only 3 out of the 1128 individuals banded as chicks, i.e. know aged birds, recruited at age two into the breeding cohort). After the third year, individuals can stay as immatures, or transition to and then among one of the other breeding stages, E, H, and F. Since some birds had unknown breeding stages, we cor-

bioRxiv preprint doi: https://doi.org/10.1101/188276; this version posted September 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

rected the estimated survival and transition probabilities among the observed breeding stages (E, H, F) for the unknown events by weighting probabilities according to survival and transition rates (Appendix). Our stage structured matrix projection model included the four stages (I,E,H,F), with stage F being the only stage contributing to reproduction. Since sex determination for Thick-billed murres is challenging, we used data on both sexes for estimating survival, resighting, and transition probabilities (assuming same survival and transitioning among sexes). For the stage structured matrix projection model we assumed 50% of chicks to be female, and we included only females for the fertility of the projection model (Table 1).

The colony on Coats Island is divided in different study plots and resighting effort varied among study plots and years. We therefore accounted for this varying effort among plots and years when we estimated the stage specific survival and transition probabilities, that is we accounted for plot and year specific resighting probabilities (mean= 0.41 ± 0.17 Stdev) but not stage specific resighting probabilities (i.e. we assumed that E,H,F stages are equally likely being resighted). We estimated survival and transition probabilities across all plots. This means, the probability of a bird surviving or transitioning among stages did not depend on the study plot it bred at, but on its current stage. Further detail on estimating resighting, survival and transition probabilities, for which we used program MARK (White and Burnham, 1999), is provided in the Appendix.

The data only included birds that recruited as breeders (or attempted breeders) to the colony, we therefore adjusted the immature survival for the population projection model using a previously described estimate of 40.5% of fledglings survival to age three, the age when many indivdiuals started to recruit as breeders (Gaston et al., 1994). This resulted in an annual immature survival of 0.74. Survival rates of the other states (after correcting for the unknown events) equalled 0.96 for E, 0.87 for H, and 0.91 for F. Table 1

bioRxiv preprint doi: https://doi.org/10.1101/188276; this version posted September 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

shows the population projection matrix, summarizing the stage transition and survival rates (column sums). The corresponding transformed Markov chain is shown in Table 2.

Demographic parameters of the stage structured Thick-billed murre population model

We estimated the population growth rate for the projection model at λ=1.041 (dominant eigenvalue of matrix shown in Table 1), which might be a slight overestimation compared to the observed population growth. The quasi stable stage distribution of the projection model was I=0.33, E=0.25, H=0.07, F=0.36 (scaled corresponding right eigenvalue $\mathbf{w}$) and the corresponding reproductive values are I=1.0, E=2.2, H=2.1, F=2.7 (corresponding left eigenvalue $\mathbf{v}$, scaled for I=1). The sensitivities with respect to λ of the population projection model (Table 1) are given in Table 3. They show that population growth rate is most sensitive to transitions from the immature to the fledging stage, as well as remaining in the fledging stage, the only stage that contributes to fertility. The population shows a high rate of diversification with an entropy, H=0.98% of the maximum entropy for the Markov chain matrix (Table 2).

Integrated sensitivities and selective forces

The sensitivities, as we estimated for instance in Table 3, imply that a realized perturbation in a transition probability is traded-off against survival of that stage. Here we are not interested in reproduction-survival trade-offs, but in changes among stage dynamics without changing stage survival. If we perturb one transition probability, but keeping the survival rate of this stage constant, we have to compensate this perturbation by one or more transitions of this stage. Such constraints in changing the transition

bioRxiv preprint doi: https://doi.org/10.1101/188276; this version posted September 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

probabilities for stage structured models are discussed by Caswell (2001) from a biological perspective. There are many solutions to fulfill these constraints, here, we reduced the transition probability of one matrix parameter by 0.01 and increased at the same time the transition probabilities of the remaining stage parameters by equal amounts as to perturbations $\mathbf{e}^T\,\mathbf{B} = \mathbf{0}^T$, i.e., columns sum of the perturbations equal zero (see also equation (4)). We call these sensitivities integrated sensitivities, because they comprise changes in multiple transition rates and we sum weighted sensitivities according to the perturbations described in $\mathbf{B}$. We estimated such an integrated sensitivity related to a reduction in each transition probability (note we consider only perturbations that leave unchanged the signature of the Markov chain: i.e., whenever $p_{ij} = 0$ we keep $b_{ij} = 0$). Each change in the transition probability changes the entropy and the population growth, but perturbations now having signs, and resulting changes on entropy (diversification) or population growth can be positive or negative. Classical sensitivities, as illustrated for instance in Table 3, hold only positive values (any increase in a transition rate also increases survival and therefore has to increase population growth).

In Table 4 we show results for the integrated sensitivities of population entropy for the murre example. Table 5 shows the corresponding integrated sensitivities with respect to λ. If we reduce the transition rate of remaining as immatures (I to I) by 0.01, and at the same time increase the remaining three transition probabilities (from I to E,H &F) by $0.01/3 = 0.003333$, entropy (rate of diversification) increases by 0.0034, while the population growth rate, λ, increases by 0.00219. A reduction in the probability of fledging a chick in two consecutive years (transition stage F to F) and increasing the probability of transitioning from a fledging event (F) to stage E, or H, and changing fecundity increases entropy most, while reducing the transition between F and H reduces entropy most (Table 4). These inte-

bioRxiv preprint doi: https://doi.org/10.1101/188276; this version posted September 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

grated sensitivites of entropy are distinct from integrated sensitivities with resepct to λ. Reducing the probability of staying in stage I in consecutive years (I to I transition) and increasing the chance of recruiting to the breeding cohort (I to E,H,F transitions) increases population growth rate most strongly. Reducing fecundity (F to I transitions) also leads to a strong increase in population growth when at the same time transitions between F and E, H, F are increased (Table 5). The most negative effect for population growth rates are achieved if transitions between I and F, and F and F are reduced (and at the same time transitions to the other stages are increased). This latter observation is not surpising given that we find the highest classical (non-integrated) sensitivities with respect to λ (Table 3) for the same transitions.

The integrated sensitivities with respect to entropy (Table 4), show which transitions are most critical for generating diversity among life courses, but they do not provide information on whether such diversity might be adaptive or neutral. The integrated sensitivities with respect to λ (Table 5) provide us with information how changes in transitions affect population growth and fitness, and such sensitivities have been used to quantify forces of selection on transition probabilities (Caswell, 2001). The higher the sensitivity with resepect to λ, the stronger selection should have acted on these transition rates. However, these integrated sensitivities with respect to λ (Table 5) do not inform us on diversity among life courses. Therefore, we correlated the two measures of integrated sensitivities to approach the question whether the observed diversity and generating moments measured as the entropy, the rate of diversification among life courses, might be adaptive. As we see in Fig. 1, the two measures of sensitivity are not correlated and hence the rate of diversification and the resulting diversity among life courses might rather be neutral. Such interpretation supports neutral theories of life history evolution (Tuljapurkar et al., 2009; Steiner and Tuljapurkar, 2012),

bioRxiv preprint doi: https://doi.org/10.1101/188276; this version posted September 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

and challenges adaptive theories arguing that variability in life courses is adaptive, an interpretation found in various evolutionary ecological studies (Stearns, 1992).

Conclusions

The sensitivities of the population entropy we derived for different types of Markov chains are widely applicable to population biology and demography, since they only require the dominant eigenvalue and corresponding eigenvectors of non-negative matrices. We have shown how they reveal drivers of diversification in life courses (Table 4), and we can use them in combination with other sensitivities to inform on a larger and controversial debate on potential selective forces acting on the dynamics of life histories. Our example on the Thick-billed Murres illustrates that we only have a poor understanding about apriory predicting drivers of diversification. In our example these drivers are not linked to the most sensitive transitions influencing population growth and hence suggest that observed diversification in life courses and related variability in individual fitness components (survival and reproduction) are rather neutral than adaptive. We have to be cautious about over interpretation of this result, since many solutions for the constraints among transition probabilities exist (Caswell, 2001) and we only have explored one, that seemed to us biologically plausible. We believe novel approaches as our investigation of sensitivities to entropy will proof useful in understanding how individual level dynamics shape population level evolutionary and ecological dynamics and their demographic consequences for the populations. We see such examination only in its infancy and many more exciting discoveries await us in exploring underlying dynamics of age- and stage-structured populations.

bioRxiv preprint doi: https://doi.org/10.1101/188276; this version posted September 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 1: Projection matrix model

	I	E	H	F
I	0.494	0	0	0.5
E	0.22	0.465	0.231	0.161
H	0.006	0.116	0.094	0.088
F	0.02	0.378	0.55	0.657

Table 2: Tranformed Markov chain matrix

	I	E	H	F
I	0.474	0	0	0.178
E	0.462	0.446	0.231	0.125
H	0.012	0.107	0.09	0.066
F	0.446	0.446	0.679	0.63

Table 3: Sensitivity to λ

	I	E	H	F
I	0.165	0	0	0.181
E	0.361	0.277	0.073	0.396
H	0.346	0.266	0.07	0.379
F	0.445	0.342	0.09	0.487

Table 4: Integrated sensitivity to entropy

	I	E	H	F
I	0.0034	0	0	0.00062
E	0.0031	0.00192	-0.00006	-0.0024
H	-0.00461	-0.00352	-0.00093	-0.0058
F	-0.00189	0.0016	0.00099	0.00758

bioRxiv preprint doi: https://doi.org/10.1101/188276; this version posted September 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 5: Integrated sensitivity to λ

	I	E	H	F
I	0.00219	0	0	0.0024
E	-0.00043	0.00026	0.00007	-0.00047
H	-0.00022	0.00044	0.00012	-0.00024
F	-0.00154	-0.0007	-0.00019	-0.00169

A Appendix

A.1 The Ergodic Case

The perturbation matrix $\mathbf{B}$ satisfies (1) of the main text. Writing

$$\mathbf{D} = \mathbf{P} - \mathbf{w}\,\mathbf{e}^T,$$

see also that

$$\mathbf{e}^T\,\mathbf{Z} = \mathbf{e}^T\,\left[\mathbf{I} + \mathbf{D} + \mathbf{D}^2 + \ldots\right] = \mathbf{e}^T,$$

so finally, from (5),

$$\mathbf{e}^T\,\mathbf{y} = \mathbf{e}^T\,\mathbf{Z}\,\mathbf{B}\,\mathbf{w} = \mathbf{e}^T\,\mathbf{B}\,\mathbf{w} = 0.$$

The perturbation of the entropy in (1) uses the expansion

$$p\log(p+\epsilon b) = p\log p + p\,\epsilon(b/p) + O(\epsilon^2) = p\log p + \epsilon\,b + \epsilon^2\,(b^2/p) + O(\epsilon^2).$$

Keeping terms to $O(\epsilon)$ yield three terms (omitting the summations over i and j),

$$w_j b_{ij} + w_j\,b_{ij}\,\log p_{ij} + y_j\,p_{ij}\,\log p_{ij}.$$

Recall that $\sum_i b_{ij} = 0$ for every j to see that the first term is zero, leaving us with equation (8).

bioRxiv preprint doi: https://doi.org/10.1101/188276; this version posted September 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

A.2 Conditional Entropy

A.2.1 Simplifying the Entropy

The entropy is defined by the middle line of (16). Insert (15) to obtain

$$-\frac{1}{\rho}\sum_{i=1}^{(s-1)}\sum_{j=1}^{(s-1)} w_j\, v_i\, q_{ij}\, \left[\log q_{ij} + \log v_i - \log v_j - \log \rho\right].$$

Now use the facts $\sum_j q_{ij} w_j = \rho w_i$, $\sum_i v_i q_{ij} = \rho v_j$ to see that the two middle terms cancel, and to see that the last term (with sums) is just $\log \rho$. This yields the last line of equation (16).

A.2.2 Perturbing Eigenvectors

We derive (20); proceed similarly to get (21). Now the perturbed right eigenvector of $\mathbf{Q}$ satisfies the usual equation

$$(\mathbf{Q} + \epsilon\, \mathbf{B})\,(\mathbf{w} + \epsilon \mathbf{y}) = (\rho + \epsilon\nu)\,(\mathbf{w} + \epsilon \mathbf{y}).$$

The order ϵ terms here are:

$$\mathbf{Q}\mathbf{y} + \mathbf{B}\mathbf{w} = (\nu \mathbf{w} + \rho^T \mathbf{y}). \tag{A-24}$$

Now note that $\mathbf{w}\,\mathbf{v}^T$ is a matrix that projects any vector onto $\mathbf{w}$. When we perturb the matrix $\mathbf{Q}$, the change $\mathbf{y}$ must be orthogonal to $\mathbf{w}$ (otherwise we are just making a proportional change in every matrix element). Hence we must have

$$\mathbf{D}_1\mathbf{y} = (\mathbf{I} - \mathbf{w}\,\mathbf{v}^T)\mathbf{y} = \mathbf{y}.$$

Also

$$\mathbf{D}_1\mathbf{Q} = (\mathbf{I} - \mathbf{w}\,\mathbf{v}^T)\mathbf{Q} = \mathbf{Q} - \mathbf{w}\,\mathbf{v}^T\mathbf{Q} = \mathbf{Q} - \rho(\mathbf{w}\,\mathbf{v}^T).$$

bioRxiv preprint doi: https://doi.org/10.1101/188276; this version posted September 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Using these facts, multiply all terms of (A-24) by matrix $\mathbf{D}1$ to get, first,

$$\mathbf{D}_1\mathbf{Q}\mathbf{y} + \mathbf{D}_1\mathbf{B}\mathbf{w} = \rho\mathbf{y},$$

and then

$$\mathbf{D}_1\mathbf{B}\mathbf{w} = \left(\rho - \left[\mathbf{Q} - \rho\left(\mathbf{w}\,\mathbf{v}^T\right)\right]\right)\mathbf{y}.$$

Using the inverse of the matrix on the right (guaranteed to exist because ρ is the dominant eigenvalue) leads to (20).

A.2.3 Sensitivity of Entropy

We examine separately the two terms of (16) and find perturbations to order ϵ. The first term changes to

$$\log(\rho + \epsilon\nu) = \log(\rho) + \epsilon\left(\frac{\nu}{\rho}\right).$$

The second term of (16) has the form

$$\frac{1}{\rho}\,F,$$

say, where F stands for the double sum.

Now (much as in Section A.1) the perturbation of the double sum in (16) is

$$F_1 = \sum_{i,j=1}^{(s-1)} \left[(x_i w_j + v_i y_j)\, q_{ij}\, \log q_{ij} + v_i w_j\, b_{ij}\, \log q_{ij}\right].$$

Thus the effect of the perturbation on the second term of (16) is to produce

$$\frac{1}{(\rho + \epsilon\nu)}\,(F + \epsilon F_1) = \frac{1}{\rho}\,F + \epsilon\left[\frac{F_1}{\rho} - \nu\frac{F}{\rho^2}\right].$$

bioRxiv preprint doi: https://doi.org/10.1101/188276; this version posted September 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

So the total perturbation is

$$\left(\frac{\nu}{\rho}\right)\left[1+\frac{F}{\rho}\right]-\frac{F_1}{\rho}.$$

Using (16) to express F/ρ in terms of the entropy $H(\mathbf{Q})$ yields (23).

A.3 Transforming projection matrix to Markov chain

To transform a population projection model into a Markov chain, we follow Tuljapurkar's approach (Tuljapurkar, 1982). Note, Tuljapurkar's projection matrix describes transitions from rwo to column, whereas our matrix $\mathbf{P}$ describes transitions from columns to rows, hence the transformation for our matrix is as follows:

$$\mathbf{P}_M = \frac{1}{\lambda}\mathbf{W}^{-1}\mathbf{P}_P\mathbf{W}$$

with $\mathbf{P}_M$ being the Markov chain (Table 2), $\mathbf{P}_P$ being the population projection matrix (Table 1), λ being the population growth rate (dominant eigenvalue of $\mathbf{P}_P$), and $\mathbf{W}$ being a matrix of zeros except for the diagonal elements of (w_i), which are the normalized stable stage distribution values (normalized right eigenvector corresponding to dominant eigenvalue of matrix $\mathbf{P}_P$). $\mathbf{W}^{-1}$ is the inverse of matrix $\mathbf{W}$.

A.4 The thick-billed murre, population projection model

We used data from a total of 1984 individuals, of which 1128 individuals where banded as chicks (immatures), and 856 were banded as adults (left censored). In the breeding colony on Coats Island, these birds were observed over a breeding season and many resightings of an individual bird were made each year. Birds are highly philopatric to their breeding sites which makes it relatively easy to record the breeding outcome for a given year (Steiner and

bioRxiv preprint doi: https://doi.org/10.1101/188276; this version posted September 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Gaston, 2005). We used 5956 records of annual breeding outcomes of which 1313 were egg failures, E, 518 hatching failures, H, 3031 fledging successes, F, and 1094 unknown events, U. Since birds are highly philopatric to their breeding site we could assing each bird to a breeding plot. For a few birds that switched a breeding plot within their lifetime, we assigned them to the breeding plot they spent most time breeding at.

We only included data from six study plots (D, J, K, N, Q, S) that had longitudinal data on a larger number of individuals. For the 1128 immature individuals that were banded as chicks in the colony and then later recruited as breeders, we assumed that they would stay as immatures for the first three years, before they would be allowed to start transitioning to and among the breeding stages (E, H, F, U). Only three of these 1128 birds recruited at age two into the breeding cohort, for these three birds we considered their observed breeding status at age three. Once a bird left the immature stage it was not allowed to transition back to the immature stage. Entering the immature stage from a breeding stage (E, H, F, U) was only possible as a newborn, that is through fertility (Table 1).

When we estimated the stage specific survival and transition rates using program MARK (White and Burnham, 1999), we accounted for differences in resighting effort among years and plots. Banding of chicks started in 1981 but band reading only began in 1991, so all resighting probabilities for all plots prior to 1991 were set to 0. Similarly no resightings were made for plot D in 2001; for plot J prior to 1995, and in 2000, 2005-2008, 2010, 2011; for plot K in 2001, 2003-2006, and 2011; for plot N in 2001, 2003-2006, and 2011; for plot S in 2000-2002, 2004, 2006, and 2011. In those years for these plots resighting probabilities for the breeding stages (E, H, F, U) were set to 0. For plot Q we had resighting records for each year between 1991 and 2011 and estimated plot specific resighting rates for each year. We did not estimate stage specific resighting probabilities, but only plot and year specific

bioRxiv preprint doi: https://doi.org/10.1101/188276; this version posted September 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

resighting probabilities, since the resighting probability should not depend on the breeding stage (recall we have many observation of each individual within a breeding season).

When we estimated the stage specific transition parameters, using program MARK, we used a multinomial logit function to assure that the transition rates of a given stage sum to 1. This estimation of the survival and transition probabilities included unknown breeding outcomes, U. To account for these unknown breeding outcomes we corrected the survival and transition probabilities of the known breeding stages (E, H, F). We did this by first estimating the fractions of the known breeding outcomes (1313 E, 518 H, 3031 F; i.e. 0.27% E, 0.11% H; and 0.62% F). The expected number of unknown events and their associated survival rates compared to the known events was then taken into account to correct the survival rates of the known stages.

Transition rates to the unknown stage were added to the transition rates of the known stages (E, H, F). We did this by taking the estimated transition probability of a given stage to the unknown stage, and increased each stage transition of the observed stages by its relative weight. This correction was done for each stage (I, E, H, F) and provided the four by four matrix that contributes to Table(1).

Suvival estimates of the immature stage, I, based on the MARK model was very close to 1 (if we forced it to be exactly one we had convergence issues). Such a high survival rate is expected since only birds entered the data base if they were recorded as breeders (or atempted breeders), i.e. they all needed to survive the immature stage. In order to get a more realistic population projection model, we reduced annual immature survival to 0.74 which leads to a survival between fledging and age three of 40.5 %; a survival rate reported by Gaston et al. (1994) for this population.

Murres lay a single egg and do not have multiple broods, for that

bioRxiv preprint doi: https://doi.org/10.1101/188276; this version posted September 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

any successfull fledgling (stage F event) contributed to fertility. We did only consider female fledglings, assuming that 50% of all fledglings are females. So our resulting population projection model can be seen as a one sex (female) model even though we used male and female observations for estimating survival, transition and resighting probabilities. Other than a slight delay in onset of breeding for males, transition and survival rates have been estimated to be very similar in this species (Gaston et al., 1994). If we only had used data from known females the amount of data would have been much lower and parameter estimations less accurate.

bioRxiv preprint doi: https://doi.org/10.1101/188276; this version posted September 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

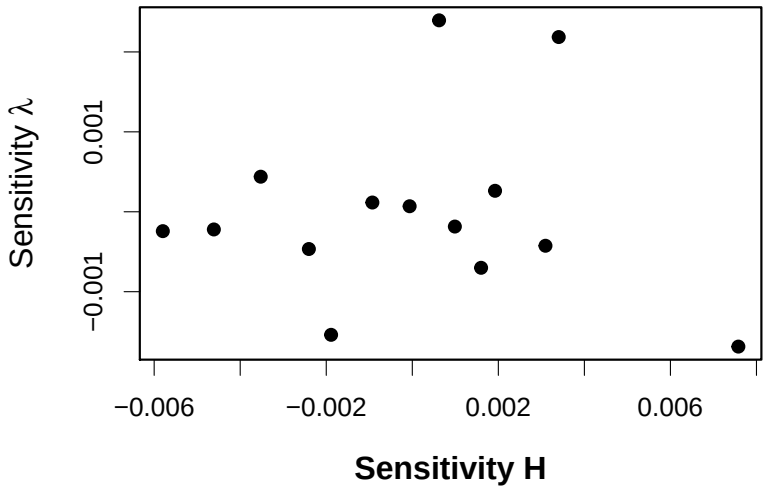

Figure 1: Correlation between sensitivity of entropy and sensitivity of λ

bioRxiv preprint doi: https://doi.org/10.1101/188276; this version posted September 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

References

Caswell, H. (2001). *Matrix population models: construction, analysis, and interpretation*, volume 2nd. Sinauer Associates.

Caswell, H. (2009). Stage, age and individual stochasticity in demography. *Oikos*, 118(12):1763–1782.

Darroch, J. N. and Seneta, E. (1967). On Quasi-Stationary Distributions in Absorbing Continuous-Time Finite Markov Chains. *Journal of Applied Probability*, 4(1):192.

Endler, J. A. (1986). *Natural selection in the wild*, volume 21 of *Monographs in Population Biology 21*. Princeton University Press.

Finch, C. and Kirkwood, T. B. (2000). *Chance, Development, and Aging*. Oxford University Press, Oxford.

Gaston, A. J., de Forest, L. N., Donaldson, G., and Noble, D. G. (1994). Population Parameters of Thick-Billed Murres at Coats Island, Northwest Territories, Canada. *The Condor*, 96(4):935–948.

Hartl, D. J. and Clark, A. (2007). *Principles of population genetics*. Sinauer, Sunderland.

Hernandez-Pacheco, R. and Steiner, U. K. (2017). Drivers of diversification in individual life courses. *bioRxiv*.

Hill, M. F., Witman, J. D., and Caswell, H. (2004). Markov chain analysis of succession in a rocky subtidal community. *The American naturalist*, 164(2):E46–61.

Kato, T. (1966). Perturbation theory for linear operators. *Grundlehren der mathematischen Wissenschaften*, 132.

Lande, R., Engen, S., and Saether, B. (2003). *Stochastic population dynamics in ecology and conservation*.

Lenormand, T., Roze, D., and Rousset, F. (2009). Stochasticity in evolution. *Trends in ecology & evolution*, 24(3):157–165.

bioRxiv preprint doi: https://doi.org/10.1101/188276; this version posted September 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Matthews, J. P. (1970). A Central Limit Theorem for Absorbing Markov Chains. *Biometrika*, 57(1):129.

Melbourne, B. a. and Hastings, A. (2008). Extinction risk depends strongly on factors contributing to stochasticity. *Nature*, 454(7200):100–3.

Schweitzer, P. J. (1968). Perturbation Theory and Finite Markov Chains. *Journal of Applied Probability*, 5(2):401.

Stearns, S. C. (1992). *The evolution of life-histories*. Oxford University Press, Oxford.

Steiner, U. K. and Gaston, A. (2005). Reproductive consequences of natal dispersal in a highly philopatric seabird. *Behavioral Ecology*, 16(3):634–639.

Steiner, U. K. and Tuljapurkar, S. (2012). Neutral theory for life histories and individual variability in fitness components. *Proceedings of the National Academy of Sciences of the United States of America*, 109(12):4684–9.

Steiner, U. K., Tuljapurkar, S., Coulson, T., and Horvitz, C. (2012). Trading stages: life expectancies in structured populations. *Experimental gerontology*, 47(10):773–81.

Tuljapurkar, S., Steiner, U. K., and Orzack, S. H. (2009). Dynamic heterogeneity in life histories. *Ecology letters*, 12(1):93–106.

Tuljapurkar, S. D. (1982). Why use population entropy? It determines the rate of convergence. *Journal of Mathematical Biology*, 13(3):325–337.

White, G. C. and Burnham, K. P. (1999). Program MARK: survival estimation from populations of marked animals. *Bird Study*, 46(sup1):S120–S139.