Migration in uncertain times

Harman Jaggi[1], David Steinsaltz[2], and Shripad Tuljapurkar[1]

[1]Department of Biology, Stanford University, Stanford, CA 94305-5020, USA

[2]Department of Statistics, University of Oxford, Oxford OX1 2HB, United Kingdom

Abstract

Understanding the effect of fluctuations on populations is crucial in the context of increasing habitat fragmentation, climate change, and biological invasions, among others. Migration in response to environmental disturbances enables populations to escape unfavorable conditions, benefit from new environments and thereby ride out fluctuations in variable environments. Would populations disperse if there were no uncertainty? Karlin showed in 1982 that when sub-populations experience distinct but fixed growth rates at different sites, greater mixing of populations will lower the overall growth rate relative to the most favorable site. Here we ask, when might environmental variability favor migration over no-migration? Specifically, in random environments, would a small amount of migration increase the overall long-run growth rate relative to the zero migration case? We use mathematical analysis and simulations to show how long-run growth rate is affected by migration rate. Our results show that when fitness (dis)advantages fluctuate over time across sites, migration may allow populations to benefit from variability. When there is one best site with highest expected growth rate, the effect of migration on long-run growth rate depends on the difference in expected growth between sites, scaled by the variance of the difference (for two sites). When variance is large, there is a substantial probability of an inferior site experiencing occasional bursts of higher growth than its average. Thus, a high variance can compensate for a large difference in average growth rates between sites. With multiple sites and large fluctuations, we explore the interplay between the length of the shortest cycle linking the best site with the second best, the average differences in growth rates between sites, and the size of fluctuations. Our findings have implications for conservation biology: even when there are superior sites in a sea of poor habitats, variability and habitat quality across space may be key to determining the importance of migration.

1 Introduction

All populations experience changing environmental conditions, both temporally and spatially. The effect of variable environments on populations is of considerable interest in the context of increasing habitat fragmentation, global climate change, and biological invasions, among others (Lande and Orzack 1988; Travis and Dytham 1999; Schreiber 2010; Oldfather et al. 2021). Understanding these effects is important as disturbances can not only depress population numbers but as fluctuations become more severe, they can cause extinctions (Lewontin and Cohen 1969; May 1971). Therefore, populations must develop strategies and adaptations that enable them to survive in a landscape of change.

Andrewartha, Birch, et al. (1954) formulated the concept of "spreading of risk" as a way to survive and reproduce in a spatially variable environment so that any one failure will not be decisively harmful. This concept recognizes the importance of migration[1] as an escape from unfavourable conditions, and adaptation to changing environments by exploiting new sites. Previous empirical and theoretical work emphasized the role of migration in rapidly-changing habitats by allowing populations to keep pace with the variation in space and time of the habitat (Southwood 1962; Boer 1968; Tuljapurkar 1990c; Denno et al. 1991). Other studies focused on the importance of migration as a stabilizing factor in a heterogeneous environment (Gadgil 1971; Roff 1974a,b); allowing for genetic cohesion of a species across space (Ronce 2007; Oldfather et al. 2021); promoting persistence in metapopulation models via recolonization of empty patches following extinction (Levins 1970; Hanski 1998; Hansson et al. 2014). However, migration may be expensive, because migrating individuals are vulnerable to risks of death, and may die without leaving any offspring, and so can reduce overall population growth rate (Karlin 1982; McNamara and Dall 2011; Altenberg 2012).

Would populations disperse in constant environments? Karlin (1982) developed an important theorem (discussed in the next section) to answer the question for sub-populations undergoing

[1]Note we use the general definition of migration as defined by Baker et al. (1978) as "the act of moving from one spatial unit to another" and use migration and dispersal interchangeably.

natural selection and migration. Karlin showed for populations experiencing different but fixed growth rates at different sites, greater mixing of populations will lower the long-term growth rate of the metapopulation (the spectral radius of the migration matrix) relative to the most favorable site. In this case, the overall population growth rate will be determined by the best site. This selection for reduced dispersal (under constant selection regimes), also known as the Reduction Principle was described in a genetic context by Feldman et al. (1980) and Feldman and Liberman (1986), and extended to the evolution of mutation rates, dispersal, sexual reproduction, and even cultural transmission (Altenberg 2012; Altenberg et al. 2017).

Karlin's theorem (Karlin 1982) assumed different but deterministic growth rates, and is relevant if there is no uncertainty. There is important later work that has incorporated stochastic environments by accounting for temporal variation in carrying capacities (Levin et al. 1984; Cohen and Levin 1991) or growth rates (Wiener and Tuljapurkar 1994; Khaladi et al. 2012). Cohen and Levin (1991) analyzed density-dependent models and find the evolutionarily stable strategy (nonzero optimal rate of dispersal) increases with variability (increased variance in carrying capacity). Wiener and Tuljapurkar (1994) focus on a more general model that emphasized the role of fluctuations without crowding effects. They show that population growth in the long run increases when there are random variations across space (migration) and/or across time (diapause). More importantly, they note through simulations and approximation that the fitness advantage gained by migrants is greatest near the limit of zero migration (no-mixing). That is, the long-run growth rate is non-differentiable in the limit of no-mixing and corresponds to a state of population instability. This behaviour, known as a singularity is also discussed in a book chapter (Tuljapurkar 1994) on frontiers in mathematical biology.

A general and precise answer as to how stochastic growth rate behaves analytically in the limit of no mixing has been difficult. This is because population structure and the stochastic growth rate change very rapidly in the neighborhood of no-mixing or singularity. Similar singularity arises in products of random matrices used in a model of stochastic structured population (Benaïm and Schreiber 2009) applied to biennial plant model (Figure 3 (a)), and in the models of disordered

matter (Derrida and Hilhorst 1983). Benaïm and Schreiber (2009) analyze metapopulations and prove that the singularity is continuous with respect to migration rate. In the absence of an analytical description of behaviour near singularity, previous studies have resorted to approximations and/or numerical analysis (Wiener and Tuljapurkar 1994; Tuljapurkar and Istock 1993; Schreiber 2012). Such approximations limit the scope of results as they rely on assumptions, such as temporal perturbations to dispersal terms being small relative to their temporal averages (Tuljapurkar et al. 2003). Altenberg (2012) examined the evolution of dispersal in variable environments and discusses why the analysis is challenging when models consider environment-specific growth rates as a random variables.

Here we ask: in a randomly varying environment, when might migration be favoured over no-migration? Specifically, does randomness in site-specific growth rates enable populations to increase their overall growth rate by migration to an otherwise unfavorable site? A possible mechanism is that when there are independent fluctuations in growth rates across sites, populations may exploit this environmental variability to their benefit. With randomly varying growth rates, there are surely instances when an otherwise unfavorable site can have higher growth rates, and migrating individuals may be able to benefit from this variability to increase their growth rate. We hypothesize that migration should evolve in simple models that account for fluctuating environments.

Our model considers a single best site (site 1 with highest expected growth rate) amidst other sites with low growth rates on average and examines the asymptotic behaviour of long-run growth rate, both analytically and through simulations, in the limit of zero migration. We consider a density-independent model for two- or more sites connected by some migration rate, ε. The growth of the population is modelled by a product of random matrices and the long-run time average of the product converges to a stochastic or long-run growth rate, $a(\varepsilon)$ (Tuljapurkar 1990a). Disturbances in environment are characterized by the variance σ^2 of the difference in growth rates at time t.

Our analytical and simulation results consider a migration rate, ε and show that increase in stochas-

tic growth rate due to migration, behaves as a power of the migration rate as ε^{β}. We use simple assumptions to prove that β (the power of migration rate ε) depends on two factors, and is best understood when we consider two sites. For a two-site case (site 1 and site 2), the power of ε depends on the difference $\tilde{\mu}$ in expected growth rate of site 2 from the best site 1, and on the variance σ^2 of the difference in site-specific growth rates at time t between site 1 and site 2. The relation is as follows: ε to the power of $(\tilde{\mu}/\sigma^2)$ (scaled by a constant). Increased fluctuations imply large values of σ^2. Since the probability of migration, ε lies between 0 and 1, smaller values of the power of ε (due to higher fluctuations) will lead to an overall increase in long-run growth rate. Thus the evolution of migration will be favoured when the variance of difference of growth rates between sites is large enough to overcome a fixed difference in expected growth rates and migration can be advantageous in spite of migration costs on the order of ε. However, when sites become more different, that is the difference between average growth rates is high, migration may not evolve despite large fluctuations.

The results are extended for multiple sites by introducing differences of any site from the best site and minimizing the ratio $\tilde{\mu}/\sigma^2$. This is analogous to the theorem for two sites in the sense that the theorem establishes sequences of possible paths (and their probabilities) for which the less favorable site does better than the favorable site. However, when there are more than two sites, there can be many different paths. The paths may involve no direct transition between the best site and other sites. Hence for more than two sites, we define a new variable: κ_j. κ_j is defined as the smallest length of a cycle in migration graph $\mathcal{M}$ that starts and ends at site 1 (best site), and passes through site j. For two sites, κ is always 2 since the only possible transitions are from site 1 to site 2 and site 2 to site 1. For multiple sites, the power of migration rate ε behaves as minima over ε to the power of $\kappa_j(\tilde{\mu}_j/\sigma_j^2)$.

Our findings provide a precise analysis that is applicable in the limit of zero migration and as a consequence to the question of evolution of migration in heterogeneous environments. The results also provide analytical insight into singular behaviour and conditions for departures from the reduction principle, thereby answering some of the challenges described in Altenberg (2012).

Note that we present the paper assuming differences in growth rates at time t are Gaussian, but the theorem holds for sub-Gaussian rate differences; the necessary modifications are described in the Appendix. This makes for more realistic analysis since growth rates are bounded and allows us to examine the sensitivity of the stochastic growth rate on the upper bound. The outline of the paper is as follows- the next section summarizes Karlin's theorem, followed by our model and theorem for the two-site case. We then discuss the simulation results and intuition. Further, we describe the general theorem for any number of sites and follow it up with simulations for the four-site case. Finally, we discuss the relevance of our results in the Discussion section.

1.1 Why disperse? Karlin's theorem

Suppose there are populations in two patches and the growth rate of population at site 1 is always greater than the population growth rate at site 2, with the growth rates fixed in time. Would dispersal evolve in this setting?

In the case where growth and survival (individual growth rates on each site) are the same every year the answer is simple: migration slows population growth and hence does not evolve. Qualitatively, any dispersal from site 1 (with greater growth rate) to site 2 (less growth rate) would decrease the overall growth rate of the total population making dispersal between sites not beneficial.

Although biologically intuitive, this result turns out to be hard to prove mathematically. Karlin (1982) combined growth and movement to prove that the asymptotic growth rates of such a system decreases with migration rate. The proof uses the Donsker–Varadhan formula for principal eigenvalues and describes the long-term behavior of coupled linear systems. Though originally described for discrete time and discrete space, Karlin's result has later been extended to linear differential equations (Altenberg 2010). The theorem has been used in the mathematical generalization of the reduction principle, greater mixing reduces growth (Altenberg 2012; Altenberg et al. 2017). The theorem holds not only for migration in physical space but also for recombination and mutation in the genetic space as extended by Altenberg and Feldman (1987) and Altenberg

et al. (2017). In this section, we explain Karlin's work and its relationship to our work. The result holds for multiple patches and the next section discusses the general many-patch case. However, we often consider 2-patch model for simpler interpretation.

In the migration context, suppose ε is the probability of dispersal (or mutation, mixing, etc) between sites. The growth rates of populations are described by a diagonal matrix, $\mathbf{R}$ where each element represents site-specific growth rate. Assuming no individuals die or reproduce during migration, the probabilities of moving from or staying at any site sum to 1. Hence, dispersal is described by a matrix, $\mathbf{M}$ whose entries are nonnegative and whose rows sum to 1. Such a matrix whose rows (or columns) are probability vectors (summing to one) is called a stochastic matrix. So the migration matrix $\mathbf{M}$ is a stochastic matrix. Since individuals can get from the one site to any other site in a finite number of steps, the migration matrix $\mathbf{M}$ is an irreducible matrix. The overall growth rate of the total population is given by the magnitude of the largest eigenvalue ($|\lambda|$) of the product of the dispersal matrix $\mathbf{M}$ and the growth matrix $\mathbf{R}$. This eigenvalue is known as the spectral radius ($\mathbf{s}$).

Karlin's theorem, interpreted biologically, states: in a system undergoing exponential growth or decay in different sites (at fixed different rates), the greater the level of mixing of population between sites, the lower the long-term growth rate of the total population.

1.2 Restatement of Karlin's theorem 5.2

Suppose $\mathbf{N}(t)$ is a vector denoting the population size or the frequency of a rare allele at each site. Then the population dynamics can be described as a discrete time linear population model

$$\mathbf{N}(t+1) = [(1-\varepsilon)\mathbf{I} + \varepsilon\mathbf{M}]\ \mathbf{R}\ \mathbf{N}(t). \tag{1}$$

where $\mathbf{M}$ is the stochastic dispersal matrix, $\mathbf{R}$ is the diagonal matrix of growth rates at each site and $\mathbf{I}$ is the identity matrix. The dimension of the matrices correspond to the number of sites. In the case of no migration ($\varepsilon = 0$), equation (1) simplifies to equation $\mathbf{N}(t+1) = \mathbf{R}\ \mathbf{N}(t)$. In this

case, the overall growth is determined by the rates prevailing at the most favorable site.

When there is migration ($\varepsilon \neq 0$), the overall growth rate is governed by the spectral radius $\mathbf{s}$ (dominant eigenvalue) of the product of $[(1-\varepsilon)\mathbf{I} + \varepsilon\mathbf{M}]$ and diagonal matrix $\mathbf{R}$. The spectral radius $\mathbf{s}\left(\mathbf{M}(\varepsilon)\mathbf{R}\right)$ measures the growth rate of the total population. In general, if we consider a family of stochastic dispersal matrice

$$\mathbf{M}(\varepsilon) = (1-\varepsilon)\mathbf{I} + \varepsilon\mathbf{M}$$

for any ε between 0 and 1. Then for any diagonal matrix $\mathbf{R} = \mathrm{diag}\,(r_i)$ with $r_i > 0$ for all i,

$$\mathbf{s}\left(\mathbf{M}(\varepsilon)\mathbf{R}\right) \text{ is strictly decreasing in } \varepsilon$$

provided that $\mathbf{R}$ is not a multiple of $\mathbf{I}$, i.e, $\mathbf{R} \neq \alpha\mathbf{I}$ for any $\alpha \in \mathbf{R}$. Thus, Karlin proved that the spectral radius, $s(\mathbf{M}(\varepsilon)\mathbf{R}))$ is a decreasing function of the migration rate ε and dispersal should not evolve if one site is better than other sites. Of course, migration between equivalent sites will be neutral.

Hence dispersal cannot evolve in a model with fixed growth rates, i.e., with matrix $\mathbf{R}$ being time-independent. This assumption, though, is not particularly realistic. Contrary to the assumption of Karlin's model, population growth rates depend on environmental factors that change over multiple timescales. A first step toward accounting for this variation is to treat stepwise growth rates as an autonomous random process. In such a case, the matrix $\mathbf{R_t}$ specifying growth-rates is environment specific (or stochastic) or time-dependent, equation (1) is replaced by the messier equation

$$\mathbf{N}(t+1) = \mathbf{M}\,\mathbf{R_t}\,\mathbf{M}\,\mathbf{R_{t-1}} \cdots \mathbf{M}\,\mathbf{R_1}\,\mathbf{N}(0). \tag{2}$$

Analysis of such a system is challenging for many reasons (Roerdink 1988; Benaïm and Schreiber 2009; Altenberg 2012). For structured populations, the stochastic growth rate is notoriously difficult to compute. As the matrix products do not commute, the order of taking products may make

a difference to the long-run behaviour. General convergence results for products of stochastic matrices do imply that the population 'forgets its past'. That is, it loses any dependence on initial conditions, as has been extensively discussed in the context of the birth and survival process in demography (Cohen 1979; Tuljapurkar 1990b). In addition, we look beyond the domain of classical probability, which primarily concerned itself with the likelihoods of typical events, to the theory of large deviations. The field of large deviations examines how behaviour of a random system diverges from its ordinary/average behaviour and focuses on extreme events that emerge from multiple rare occurrences. In some settings the vanishingly unlikely events may dominate the aggregate of an ensemble, and so produce qualitatively different behaviour. Here it is higher long-term population growth than any ancestry (or sequence of migration decisions) would be likely to produce.

A similar argument holds in the case of diapause. Diapause is a delay in biological development, often at an early stage. Many plant species produce seeds that do not germinate immediately, but remain dormant for extended times forming seed banks (Tuljapurkar 1989, 1990c). The paper by Steinsaltz and Tuljapurkar (2015) considers a two-year life-cycle model of a species, where the first year is spent growing from a seed to an adult, and the second year is spent producing seeds. They asked the question- what will be the effect on growth rates if a small fraction of the seeds remain dormant for one year? In the case where growth, survival, and seed production remain constant (fixed) from year to year the answer is similar to the migration case, that is, delaying individual growth slows down population growth. However, they prove that when environmental conditions are variable then even a tiny amount of delay can yield a substantial boost to population growth. This is shown in Steinsaltz and Tuljapurkar 2018a to be equivalent to the migration model described here, where the same maximum average log growth rate is achieved at two different sites as opposed to one unique site as presented in this paper. The increase in stochastic growth rate in that case behaves as $1/\log(\varepsilon^{-1})$ as ε goes to zero.

In the next section, we describe the model and what the analyses is trying to achieve via theorems and simulations, while also providing intuitive interpretations.

2 The model

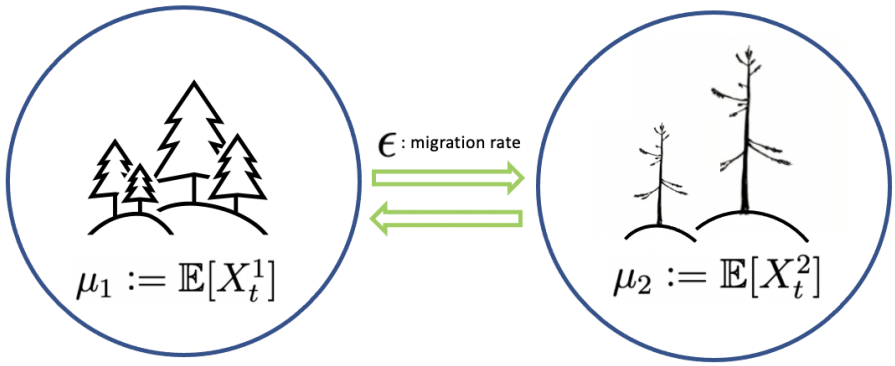


Figure 1: Site 1 (superior site) has a higher expected growth rate (μ_1) than the expected growth rate at site 1 (μ_1).

We now describe the model for two sites. Consider two sites in which populations have log growth rates $\mathbf{X}_t^1$ and $\mathbf{X}_t^2$ during time period t at site 1 and 2 respectively. This means the population at site 1 and site 2 grows at a rate $e^{\mathbf{X}_t^1}$ and $e^{\mathbf{X}_t^2}$, respectively. The growth rate $\mathbf{X}_t = (\mathbf{X}_t^1, \mathbf{X}_t^2)$ during t are assumed independent, but the components within each $\mathbf{X}_t$ corresponding to growth rates at different sites at the same time may be dependent.

We assume the population at each site is density independent, and there is no st(age) structure to the population (i.e., a scalar population in each site). The expected value of growth rates at each site are described as $\mu_1 := \mathbb{E}[\mathbf{X}_t^1]$ and $\mu_2 := \mathbb{E}[\mathbf{X}_t^2]$. We assume that μ_1 is strictly larger than other sites such that the difference between average growth rates between site 1 and site 2 is always positive, that is, $\tilde{\mu} := \mu_1 - \mu_2 > 0$.

Since the growth rates vary in time we now define the fluctuations at each site. We assume the difference of growth rates between site 1 and site 2 at each time t is distributed normally with variance σ^2. That is, $\widetilde{\mathbf{X}}_t := \mathbf{X}_t^2 - \mathbf{X}_t^1$, and $\widetilde{\mathbf{X}}_t$ is Gaussian with variance σ^2. Throughout the paper, we return to these two parameters: $\tilde{\mu}$- the difference between expected growth rate of best site (site 1) from any other site (site 2 in the two site case), and σ^2: the variance of difference of growth rates between site 1 and site 2. Please note (discussed in Appendix in detail), the argu-

ments and theorems hold even with sub-gaussian assumption for $\tilde{\mathbf{X}}_t$. ε measures the probability of migration between the two sites. The probability of staying at respective sites is then $(1-\varepsilon)$. The populations at the two sites are governed by the system of equations

$$\begin{aligned}\mathbf{N}^1_{t+1} &= (1-\varepsilon)\,\mathrm{e}^{\mathbf{X}^1_t}\mathbf{N}^1_t + \ \varepsilon\,\mathrm{e}^{\mathbf{X}^2_t}\mathbf{N}^2_t\,,\\ \mathbf{N}^2_{t+1} &= \varepsilon\,\mathrm{e}^{\mathbf{X}^1_t}\mathbf{N}^1_t + \ (1-\varepsilon)\,\mathrm{e}^{\mathbf{X}^2_t}\mathbf{N}^2_t\,.\end{aligned}$$

In matrix form the equations are written as

$$\begin{pmatrix}\mathbf{N}^1_{t+1}\\ \mathbf{N}^2_{t+1}\end{pmatrix} = \left(\mathbf{I}+\varepsilon\begin{pmatrix}-1 & 1\\ 1 & 1\end{pmatrix}\right)\begin{pmatrix}\mathrm{e}^{\mathbf{X}^t_1} & 0\\ 0 & \mathrm{e}^{\mathbf{X}^t_2}\end{pmatrix}\begin{pmatrix}\mathbf{N}^1_t\\ \mathbf{N}^2_t\end{pmatrix}.$$

We define

$$\mathbf{D}_t(\varepsilon) := (\mathbf{I}+\varepsilon\mathbf{A})\,\mathrm{e}^{\left(\mathrm{diag}(\mathbf{X}_t)\right)}$$

where $\mathbf{A} = \begin{pmatrix}-1 & 1\\ 1 & -1\end{pmatrix}$ is a 2×2 dispersal matrix such that off-diagonal elements represent migration rates in a time interval $(t, t+1)$ between the sites and $\mathbf{X}_t = \begin{pmatrix}\mathbf{X}^t_1 & 0\\ 0 & \mathbf{X}^t_2\end{pmatrix}$ such that $\mathbf{X}_t, \mathbf{X}_{t-1}, \cdots, \mathbf{X}_1$ denotes the i.i.d. sequence of 2×2 diagonal matrices, representing population growth rates across sites in a succession of times. We follow the convention from the matrix population model literature, that transition rates from state i to state j are found in matrix entry (j,i). Population distributions are thus naturally column vectors, and are updated from time $t-1$ to time t by left multiplication.

2.1 Connection to Karlin's theorem

Recall that the collection of migration matrices is given by $\mathbf{M}(\varepsilon)\mathbf{R} = \big(\mathbf{I}+\varepsilon(\mathbf{M}-\mathbf{I})\big)\,\mathbf{R}$. This can be rewritten as $(\mathbf{I}+\varepsilon\mathbf{A})\,\mathbf{R}$, where $\mathbf{A} = (\mathbf{M}-\mathbf{I})$. Since $\mathbf{M}$ is irreducible stochastic, $\mathbf{A}$ is irreducible. In our model, we work with $(\mathbf{I}+\varepsilon\mathbf{A})\,\mathbf{R}$. Karlin's theorem states that the spectral radius of $\mathbf{R}+\varepsilon\mathbf{A}\mathbf{R}$ is strictly decreasing in ε for ε between 0 and 1 if $\mathbf{R}$ is a fixed matrix of site-specific growth rates, not varying in time. Our results extend to the case where $\mathbf{R_t}$ varies randomly over time. In this way we incorporate environmental randomness into our model:

$$\mathbf{N}(t+1) = (\mathbf{I}+\varepsilon\mathbf{A})\;\mathbf{R_t}\cdots(\mathbf{I}+\varepsilon\mathbf{A})\,\mathbf{R_1}\;\mathbf{N}(0).$$

(Note that subtracting $\mathbf{I}$ from $\mathbf{M}$ is equivalent to saying there is no direct penalty for migration: Exactly 1 individual is lost from site i for every 1 that migrates to site j. The effect of increasing these diagonal "penalty" elements is mathematically straightforward, but will be ignored in the mathematical analysis to simplify the notation. It will be discussed briefly in the simulations of section 3.1.3.)

We now consider several important questions. What is the asymptotic behaviour of the product in equation (2)? Should dispersal evolve when population growth rates vary in temporally variable environments? When environmental conditions vary randomly from year to year, will a species benefit from the ability to disperse to an alternative site, even when the conditions at the alternative site are (on average) worse?

2.2 Long-run stochastic growth rate

There is no closed formula for the long-term stochastic growth rate, hence we will focus on understanding its general behavior for ε close to 0. Owing to the multiplicative nature of population growth (survival and reproduction), in a stochastic environment, the expected long-term growth rate grows as the expectation of log of the matrices and converges to a quantity a, calculated as $a = \log\lambda_s = \lim_{t\to\infty}(1/T)\log\big(\mathbf{N}(t)/\mathbf{N}(0)\big)$, where $\mathbf{N}(t) = \mathbf{N}_1(t)+\mathbf{N}_2(t)+\cdots$ is the total popu-

lation size at time t (Tuljapurkar 1990a). The stochastic growth rate is best understood as a geometric mean growth rate for a site, where the realized cumulative growth at the site is due to long temporal sequence of habitats (Tuljapurkar et al. 2003).

As the population evolves over time, the long-run growth rate or stochastic growth rate is defined as as the product of the matrices:

$$a(\varepsilon) := \lim_{T\to\infty} \frac{1}{T} \log(\mathbf{D}_T(\varepsilon) \cdot \mathbf{D}_{T-1}(\varepsilon) \cdot \cdots \cdot \mathbf{D}_1(\varepsilon)).$$

For $\varepsilon > 0$ the i.i.d. sequence $\mathbf{D}_t(\varepsilon) = (\mathbf{I} + \varepsilon\mathbf{A})\,\mathbf{R_t} = (\mathbf{I} + \varepsilon\mathbf{A})\,\mathrm{e}^{\left(\mathrm{diag}(\mathbf{X}_t)\right)}$ satisfies the conditions for the existence of a stochastic growth rate independent of starting condition (Cohen 1979). When the stochastic growth rate a is negative, then the total population abundance goes to zero with probability one. When the rate is positive, there is a unique positive stationary distribution. This stochastic growth rate can be calculated from numerical simulations of many time steps of a population in many independent sites.

For the case with no migration, $\varepsilon = 0$ and all the off-diagonal elements of $\mathbf{A}_t$ are 0. The product for the $\varepsilon = 0$ case is given by:

$$\mathbf{D}_t(0)\mathbf{D}_{t-1}(0)\cdots\mathbf{D}_1(0) = (\mathbf{I} + 0 * \mathbf{A})\,\mathbf{R}_t\mathbf{R}_{t-1}\cdots\mathbf{R}_0 = \begin{pmatrix} \mathrm{e}^{\mathbf{X}_1^t + \mathbf{X}_1^{t-1} + \cdots + \mathbf{X}_1^0} & 0 \\ 0 & \mathrm{e}^{\mathbf{X}_2^t + \mathbf{X}_2^{t-1} + \cdots + \mathbf{X}_2^0} \end{pmatrix}.$$

By the Law of Large Numbers, the stochastic growth rate for the no-migration case is:

$$a(0) = \lim_{T\to\infty} \frac{1}{T} \log(\mathbf{D}_{T-1}(0) \cdot \mathbf{D}_{T-2}(0) \cdot \cdots \cdot \mathbf{D}_0(0)) = \mu_1.$$

Consistent with Karlin's Theorem, the growth rate is dominated by the average rate prevailing at the best site, in this case, μ_1. Having defined $a(\varepsilon)$ and $a(0)$, we now turn to the theorem that governs the asymptotic behaviour of $a(\varepsilon) - a(0)$ in the next section.

3 Theorem for the two-site case

For migration to evolve, the stochastic growth rate $a(\varepsilon)$ must increase as a function of migration rate relative to the growth rate with no migration, $a(\varepsilon = 0)$ or $a(0)$. As a result, the difference between $a(\varepsilon)$ and $a(0)$ must be positive so the population can benefit from migration in the long-run.

Before stating the theorem, we discuss the probabilistic reasoning for how the theorem works. Since the environment is fluctuating, there is always a fraction of times when the site with lower growth rate is doing better than the best site. The theorem establishes sequences of possible excursions (and their probabilities) for which the less favorable site does better than the favorable site. Here, excursions are defined as the set of paths in the migration graph that start and end at the best site 1, with no intervening returns to site 1. Once the excursions are established, we sum over their probabilities such that cumulative probability is large, that is approaches 1. The mathematical details for the proof are described in the Appendix.

As described in the model section above, we consider two sites with expected growth rates as μ_1 and μ_2. We are interested in the difference in expected growth rates between two sites, $\tilde{\mu} = \mu_1 - \mu_2$ and the variance σ^2 describing the fluctuations. The theorem below for the two-site case describes the lower and upper bound for the difference in $a(\varepsilon) - a(0)$.

Theorem 1. *The difference between stochastic growth rates, $a(\varepsilon)$ and $a(0)$ is bounded below and above (scaled by a constant) as follows (Steinsaltz and Tuljapurkar 2018b):*

$$\varepsilon^{4\tilde{\mu}/\sigma^2} \leq a(\varepsilon) - a(0) \leq \varepsilon^{\left(\frac{4\tilde{\mu}/\sigma^2}{2\tilde{\mu}/\sigma^2+1}\right)}$$

where $\tilde{\mu}$ and σ^2 are the difference in expected growth rates between two sites and the variance in difference between growth rates, respectively. When $\rho = \tilde{\mu}/\sigma^2$ is smaller than $\frac{1}{4}$, small deviations from zero migration definitely become advantageous (although the theorem does not exclude the possibility that this transition occurs at an even higher level, up to $\frac{1}{2}$). Since we assume the unique maximum average growth rate is at site 1, the maximum growth for the perturbed pro-

cess arises from rare excursions away from 1; in particular, from those that include the (not necessarily unique) site that decrease (minimize) ρ. This is because for a fixed $\tilde{\mu}$, as fluctuations σ^2 increase, the power ρ decreases. Since the probability of migration ε lies between 0 and 1, the value of ε^{β} also goes up (due to high fluctuations) leading to an increase in the stochastic growth rate.

Thus the evolution of migration will be favoured when the variance of difference of growth rates between sites is large enough to overcome a fixed difference in expected growth rates and migration can be advantageous in spite of migration costs on the order of ε. For the ratio $\frac{\tilde{\mu}}{\sigma^2}$ to be small, the sites need to be similar such that on average, the difference between μ_1 and μ_2 is small. In summary, for a fixed difference $\tilde{\mu}$, when the variance σ^2 is large, despite any migration costs merely proportional to the total migration rate, migration will be advantageous. In the next section, we examine the effects of $\tilde{\mu}$ and σ^2 through simulations.

3.1 Simulation results

3.1.1 Effect of varying fluctuations on stochastic growth rate

We first find that when the average growth rates at two sites are similar i.e, when the difference of mean growth rates is small ($\tilde{\mu} = 0.1$ in Figure 2), the population with a small positive migration rate will benefit from the variability between sites (described by σ^2) relative to the zero migration case. Increasing the variability σ^2 of the difference between $\widetilde{X}_t^{(1)} = X_t^{(1)} - X_t^{(0)}$ increases the benefit from migration. This may enable populations to grow faster than would have been possible at the single best site. Our finding suggests that a species will benefit from the ability to wander between two equally good territories where environmental conditions vary randomly between them from year to year.

Figure 2 describes stochastic growth rate a as a function of migration rate ε for different levels of variability, characterized by σ^2. For low values of σ^2, the gain from migration is not sufficient to cause an overall increase in the long-term growth rate. However, for higher values of σ^2, introduction of even small level of migration leads to a sharp increase in stochastic growth rate. The

sharp increase with the introduction of migration (neighborhood of $\varepsilon = 0$) is even more evident in Figure 3 where y-axis is the log of the stochastic growth rate $a(\varepsilon)$. Therefore, given enough variability (a high enough value of σ), migration rate ε can increase $a(\varepsilon)$ relative to $a(0)$. The simulations hold even when we consider more than 2 sites and will be discussed in the next sections.

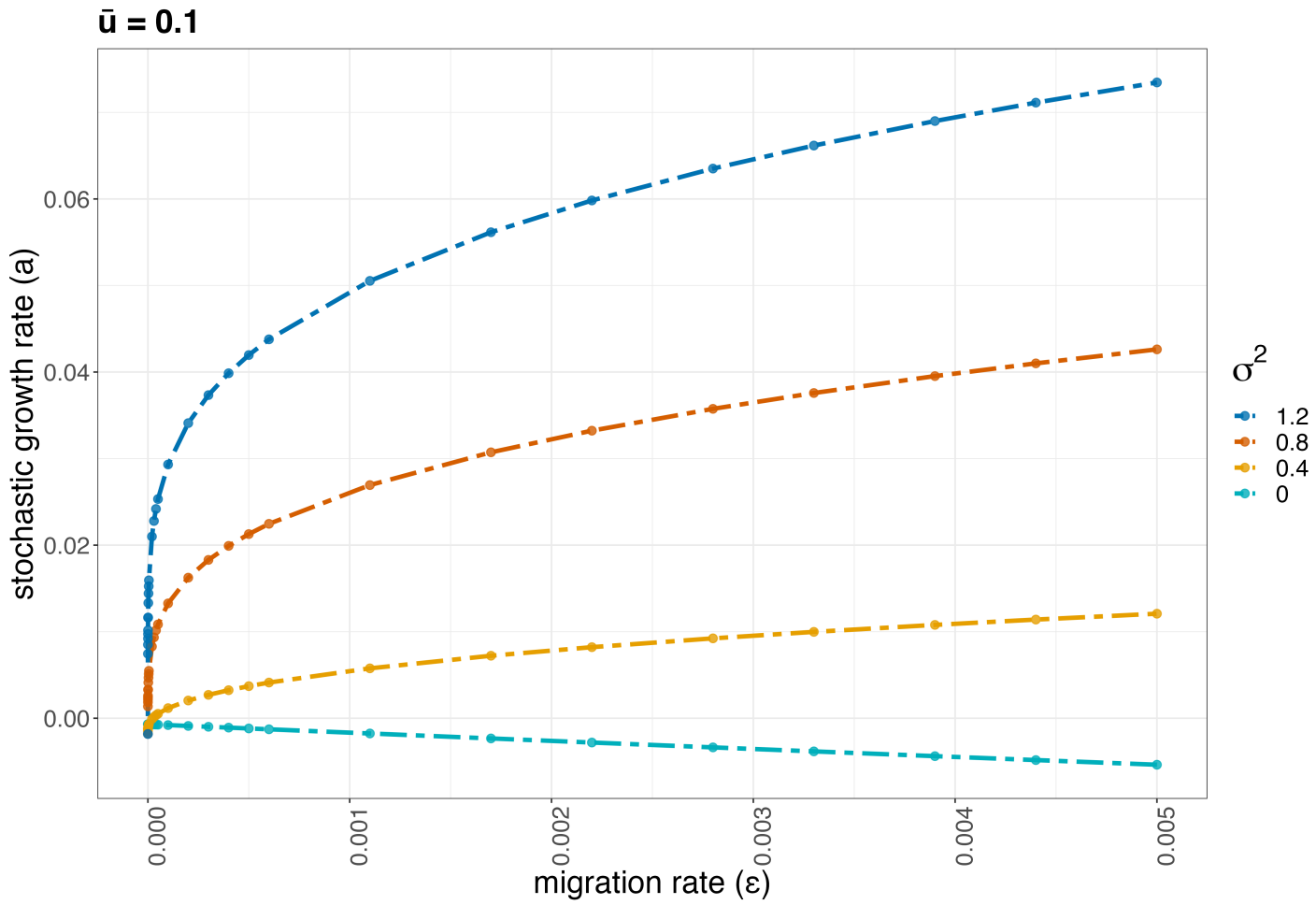


Figure 2: Stochastic growth rate as a function of migration rate. Different colors indicate different values of variability, described by σ^2. Higher values of σ^2 correspond to increase in stochastic growth rate for a fixed value of difference between average site-specific growth rates ($\tilde{\mu} = 0.1$). Here, $a(0) is 0$.

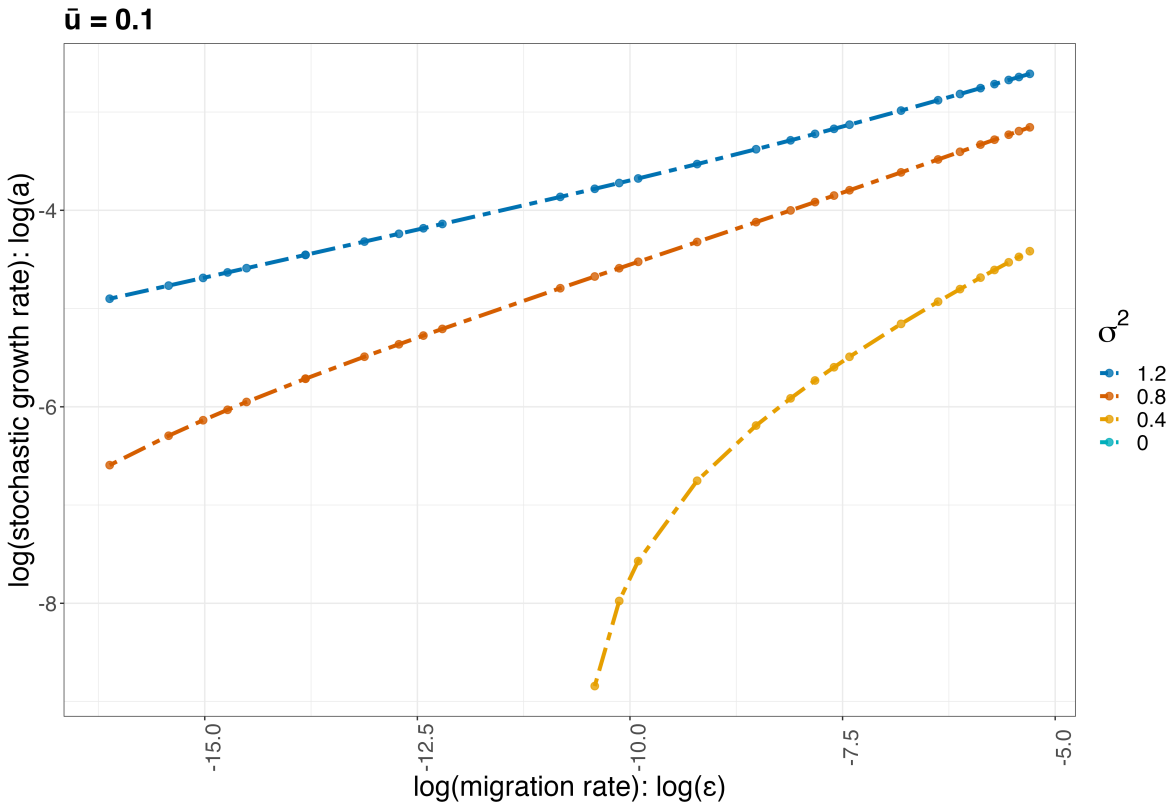


Figure 3: The y-axis represents logarithm of the stochastic growth rates. On the x-axis is the migration rate. Difference in expected growth rates between sites is set to $\tilde{\mu} = \mu_1 - \mu_2 = 0.1$. Colors indicate values of variability, σ^2.

3.1.2 Effect of varying difference in mean growth rates between sites

Next we consider the effect of changing the gap $\tilde{\mu} = \mu_1 - \mu_2$ between the average growth rates at sites 1 and 2. As predicted by the theorem, we find that the larger the difference between the growth rate between the best site and at the second site ($\tilde{\mu}_1$ increased from 0 to 0.45), the smaller the benefit from migration, measured in terms of increase in long-term growth rate.

Biologically, a greater difference $\tilde{\mu}_1$ will make the rare events when growth at the inferior site exceeds the growth at the best site even rarer, while a greater variance (σ) will make such exceptional deviations more likely. Thus, high variance may compensate for a large difference in mean growth rate to an extent.

Figure 4 depicts the effect of changing $\tilde{\mu}$ on stochastic growth rate a as a function of ε. With increasing $\tilde{\mu}$, as the second-best site becomes more inferior to the best site, the benefit from migration decreases significantly. Note that the first panel on the top-left describes the case where both sites have same expected growth rate and hence $\tilde{\mu} = 0$. In this case, with variability σ close to 0, the stochastic growth rate is the average of the growth rates between two sites (Blue line). In the last panel, the stochastic growth rate does not increase with increase in fluctuations and there is no benefit from migration as the sites are very different on average.

We plot the combined effect of changing site quality $\tilde{\mu}$ and variability σ on the difference in stochastic rates (with and without migration) $a(\varepsilon) - a(0)$ in Figures 5 and 6. The migration rate ε is fixed at 0.1 in Figure 5, while Figure 6 presents the effects of differing migration rates. To emphasize the role of fluctuations on increasing the difference $a(\varepsilon) - a(0)$ (legend "a_diff" in Figure 5), the limits on the x-axis are above 0.5. It is evident from the contours that for a fixed difference in expected growth rates between sites $\tilde{\mu}$, with enough variability σ, populations overcome the effect of migrating to a site with habitat quality.

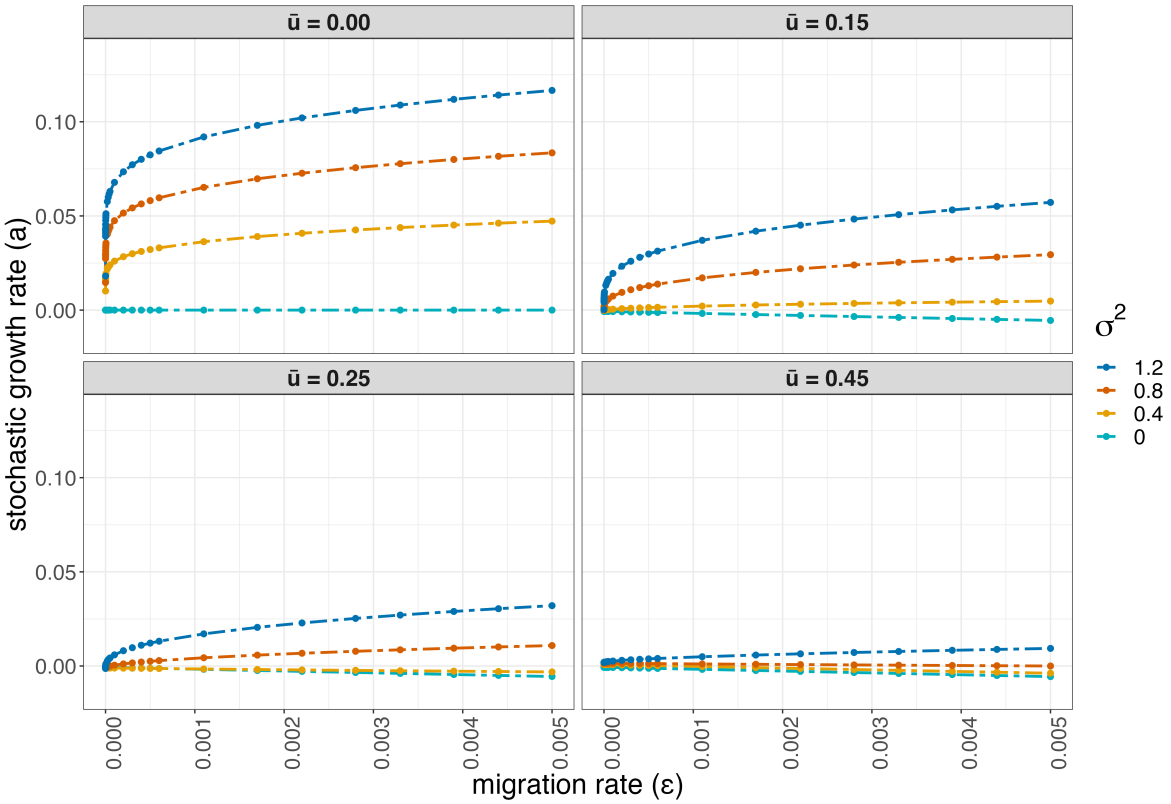


Figure 4: Stochastic growth rate as a function of migration rate. On the x-axis is the migration rate, on the y-axis is the stochastic growth rate. Each panel corresponds to a fixed value of difference in expected growth rates between two sites, $\tilde{\mu} = \mu_0 - \mu_1$. Different colors indicate values of variability, σ^2.

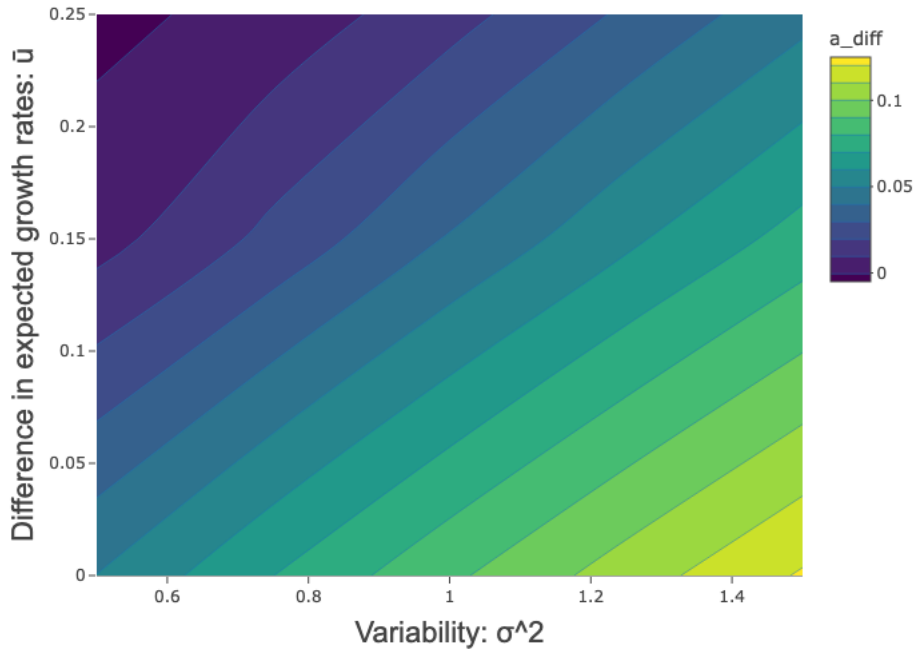


Figure 5: The plot depicts the combined effect of $\tilde{\mu}$ and σ^2 on the difference in stochastic growth rate, with and without migration, $a(\varepsilon) - a(0)$. On the y-axis is the difference in expected growth rates $\tilde{\mu}$. On the x-axis is the variability σ^2. We fix the migration rate ε to 5e-03. Each color in the legend corresponds to values of $a(\varepsilon) - a(0)$

.

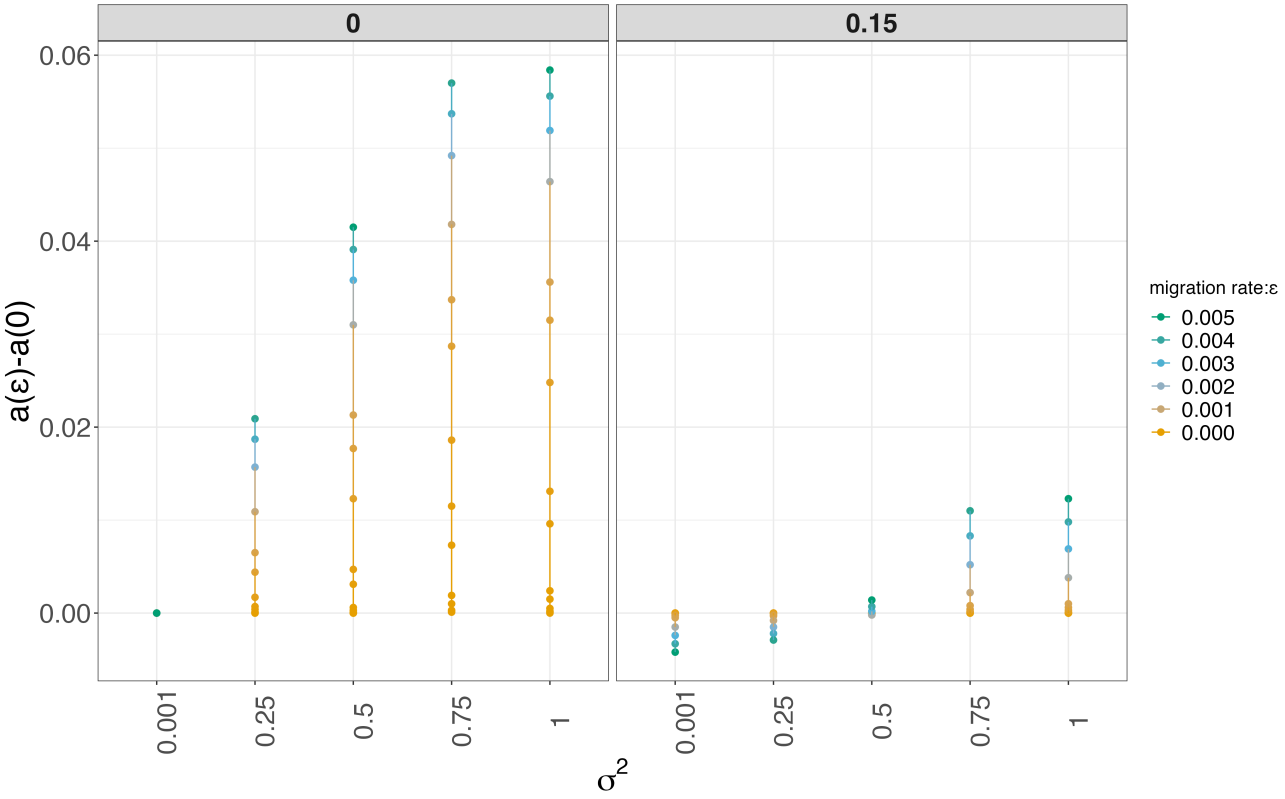


Figure 6: The plot depicts the combined effect of $\tilde{\mu}$ and σ on the difference in stochastic growth rate, with and without migration, $a(\varepsilon) - a(0)$. On the y-axis is the difference in expected growth rates $\tilde{\mu}$. On the x-axis is the variability σ^2. We do not fix the migration rate in this plot such that each color in the legend corresponds to values of migration rate ε

.

3.1.3 Effect of varying migration penalty

We also examined the effect of increasing migration penalty or costs on the long-run growth rate. The costs are increased by changing the diagonal elements in the migration matrix. This imposes an additional linear cost (as a function of migration rate ε) to migration, which will of course make itself felt when ε is large. For small ε the relevance depends essentially on the ratio $\tilde{\mu}/\sigma^2$ that determines the exponent of the power law that describes the migration benefit. When this exponent is < 1 the benefit will dominate, and growth will be increased for sufficiently small positive ε, regardless of the migration penalty. When it is ≥ 1 the migration penalty will dominate. Figure 7 shows three panels with increasing levels of migration penalty at fixed $\tilde{\mu}$ and σ.

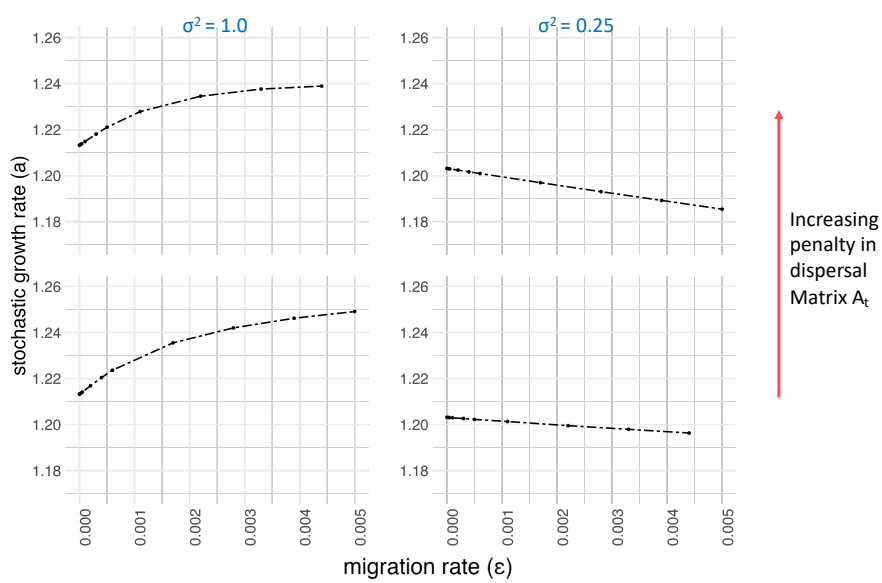


Figure 7: The figure has migration rate on the x-axis and the stochastic growth rate on the y-axis. The two plots in the left panel correspond to high variability ($\sigma^2 = 1$) while the right panel corresponds to low variability ($\sigma^2 = 0.25$). The top and bottom plots are arranged in decreasing values of migration penalty for a fixed $\tilde{\mu}$.

3.2 Intuition for the two-site case

Theorem 1 gives the lower and upper bound for the difference in stochastic growth rates (with and without migration), $a(\varepsilon) - a(0)$. Since the lower bound is positive and increasing as a function of ε, migration increases the overall growth rate on the long run. The bounds depend on the ratio $\rho = \tilde{\mu}/\sigma^2$. We discuss an intuition for the ratio. A similar ratio is discussed in Levin et al. (1984) based on the mean and variance of the lognormal distribution.

In the case of no stochasticity, $X^1 > X^2$, and dispersal should not evolve as the population achieves a higher growth rate in site 0. When some individuals disperse during a time interval from site 0 to site 1, the disperser's growth rate relative to a non-disperser is $Y_t = X_t^2 - X_t^1$. The geometric growth rate, per unit time, is changed by a factor $\exp(Y_t)$. Now the expected growth rate, $\mathbb{E}(Y_t) = \mu_2 - \mu_1 = -\mu < 0$ and $\mathrm{Var}(Y_t) = \sigma^2 > 0$. Since we are assuming X_t^i are normal, the distribution of Y_t is lognormal and

$$\log \mathbb{E}(\exp(Y_t)) = -\mu + \frac{1}{2}\sigma^2.$$

So the expected relative growth rate is greater than 1 if the ratio

$$\frac{\mu}{\sigma^2} < \frac{1}{2}.$$

This is one approach to understand why the bounds depend on ratio of μ/σ^2, and also why $\frac{1}{2}$ is an important threshold. However, the actual relative rate depends on the time period T over which a migrating ancestry experiences this relative rate, a concept that is formalised in section 5 as the duration of an excursion.

4 Theorem for more than two sites

We now consider multiple sites $0, 1, \ldots, d-1$ with log growth rates $X_t^{(0)}, X_t^{(1)}, \ldots, X_t^{(d-1)}$ during time period t. The expected growth rates at each site are μ_0, μ_1, $\cdots$, μ_{d_1}, respectively. Since we

are interested in the difference in expected growth rates between of any site j from the unique best site 1, we define $\tilde{\mu}_j = \mu_0 - \mu_j$. As discussed in earlier sections, the random variables $\widetilde{X}_t^{(i)} := X_t^{(i)} - X_t^{(0)}$ are assumed Gaussian with variance σ_i^2 and describe the fluctuations of the difference in growth rates at time t. The requirement that the log growth rates have Gaussian distribution is for simplicity; in the Appendix this assumption is relaxed to allow $\widetilde{X}_t^{(i)} := X_t^{(i)} - X_t^{(0)}$ to be sub-Gaussian with variance factor τ_j.

The probabilistic reasoning for the theorem for multiple sites is analogous to that for the theorem for two sites. The theorem establishes sequences of possible excursions (and their probabilities) for which the less favorable site does better than the favorable site. However, when there are more than two sites, there can be many different paths. The paths may involve no direct transition between the best site and the second-best site. Hence for more than two sites, we define a new variable: κ_j. κ_j is defined as the smallest length of a cycle in migration graph $\mathcal{M}$ that starts and ends at site 0, and passes through site j. For two sites, κ is always 2 since the possible transitions can be from site 0 to site 1 and site 1 to site 0. We incorporate κ by modifying the paths between sites, reflected in the dispersal matrix A_t.

Establishing and summing probabilities for excursions are more involved when there are more than two sites and the mathematical details are described in the Appendix. We now state the theorem for more than two sites. As discussed before, for migration to evolve, the difference between $a(\varepsilon)$ and $a(0)$ must be positive so the population can benefit from migration in the long-run.

Theorem 2. *Let $\rho_j = \frac{\tilde{\mu}_j}{\sigma_j^2}$ and κ_j be the smallest length of a cycle in migration graph $\mathcal{M}$ that starts and ends at site 0, and passes through site j. Define $\underline{\eta} = \min_j 2\kappa_j\rho_j$, $\bar{\eta} = \min_j 2\kappa_j\rho_j/(1+2\rho_j)$. Then for any $\delta > 0$, if $\underline{\eta} < 1$ and for $\varepsilon > 0$ sufficiently small:*

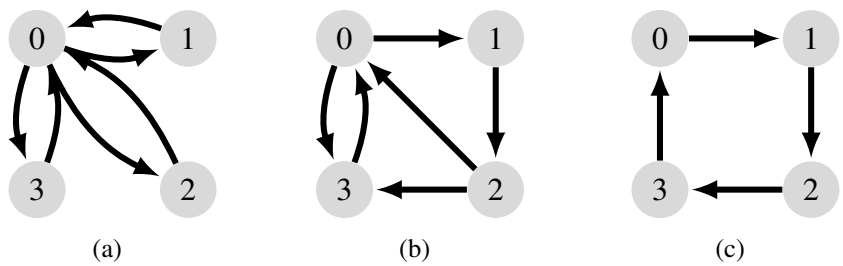
$$\varepsilon^{\underline{\eta}+\delta} \leq a(\varepsilon) - a(0) \leq \varepsilon^{\bar{\eta}-\delta}$$

Note that the lower bound is dominated by the higher power of ε. As σ_j^2 increases, so does the power (since ε is smaller than 1), yielding conditions that make small deviations from the case

of zero migration advantageous despite the costs that are merely proportional on the order of ε. Note that κ features in the two-site case as well but since $\kappa = 2$ is fixed for the two site case, it is absorbed as a constant.

4.1 Example for four-sites

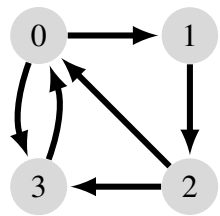
Consider the case for 4 sites- populations dispersing away from site 0 can return in a cycle of length 2, 3 or 4. Since we assume there is no net gain or loss of individuals, the row sums in migration matrix $A_t(j,i)$ is 0. In Figure 8, the migration graphs and corresponding matrices are shown. Each graph corresponds to different length for κ_j.



$$\begin{array}{c|c} & \begin{matrix} 0 & 1 & 2 & 3 \end{matrix} \\ \begin{matrix} 0 \\ 1 \\ 2 \\ 3 \end{matrix} & \begin{pmatrix} -1 & 1 & 0 & 0 \\ 1 & -1.2 & 0.1 & 0.1 \\ 1 & 0 & -1.1 & 0.1 \\ 1 & 0 & 0 & -1 \end{pmatrix} \end{array} \qquad \begin{array}{c|c} & \begin{matrix} 0 & 1 & 2 & 3 \end{matrix} \\ \begin{matrix} 0 \\ 1 \\ 2 \\ 3 \end{matrix} & \begin{pmatrix} -1 & 0.5 & 0 & 0.5 \\ 0 & -1 & 1 & 0 \\ 0.5 & 0 & -1 & 0.5 \\ 1 & 0 & 0 & -1 \end{pmatrix} \end{array} \qquad \begin{array}{c|c} & \begin{matrix} 0 & 1 & 2 & 3 \end{matrix} \\ \begin{matrix} 0 \\ 1 \\ 2 \\ 3 \end{matrix} & \begin{pmatrix} -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \\ 1 & 0 & 0 & -1 \end{pmatrix} \end{array}$$

Figure 8: The migration graph $\mathcal{M}$ for four-sites and their respective migration matrices. For graph (a): κ_j is 2 for all sites j, for graph (b): κ_1 is 3 for site 1, and for graph (c): κ_j is 4 for all sites j.

For migration graph $\mathcal{M}$ in Figure 9, we can calculate $\rho = \tilde{\mu}_j/\sigma_j^2$ for the four-site graph. From the theorem, it is clear that to increase $a(\varepsilon) - a(0)$, $\kappa\rho$ must be minimized. We see that $\kappa\rho$ is minimized at site 1, despite the fact that it is not in the shortest cycle, nor does it have the smallest mean difference in log growth rate from the best site 0, yielding a lower bound on the sensitivity to migration on the order of $\varepsilon^{0.6}$. The upper bound $\varepsilon^{0.41}$ comes from migration to site 3, which has the lowest mean log growth rate, but compensates for this with a shorter cycle length.



(a) Migration graph

j	$\mathbb{E}[X_t^{(j)}]$	σ_j^2	ρ_j	κ_j	$\kappa_j\rho_j$	$\frac{\kappa_j\rho_j}{1+2\rho_j}$
0	1					
1	0.4	3	0.2	3	0.6	0.43
2	0.7	0.6	0.5	3	1.5	0.75
3	0.3	2	0.35	2	0.7	0.41

(b) Selecting the second site

Figure 9: Migration graph

4.2 Simulation results for four sites

4.2.1 Effect of varying fluctuations and difference in mean growth rates

For four sites (labelled 0, 1, 2, and 3), we assume site 0 has highest average growth rate μ_0. We define the difference for sites from the best site as follows: $\tilde{\mu}_0 = \mu_0 - \mu_1$, $\tilde{\mu}_1 = \mu_0 - \mu_2$, and $\tilde{\mu}_2 = \mu_0 - \mu_3$ and vary the fluctuations σ^2. For the simulations we consider two cases: first, we consider the vector $(\tilde{\mu}_0, \tilde{\mu}_1, \tilde{\mu}_2) = (0.05, 0.10, 0.15)$. This corresponds to a case when the three sites are similar to the best site on average. Next, we consider $(\tilde{\mu}_0, \tilde{\mu}_1, \tilde{\mu}_2)$ is $(0.35, 0.40, 0.45)$ so that the three sites in this case are the most different from the best site. Please note that these numbers are chosen for an illustrative purpose, and other combinations work.

As shown in Figure 10, similar to the results for two-sites, we find that increasing fluctuations increases the gain in stochastic growth rate relative to no migration. However, as other sites become different on average from the best site, migration does not yield an increase in stochastic growth rate and can be detrimental despite high fluctuations. Hence populations can not benefit from variability in this case (last panel Figure 10).

We consistently find that the smallest kappa yields the highest increase in stochastic growth rate

as discussed in the next section.

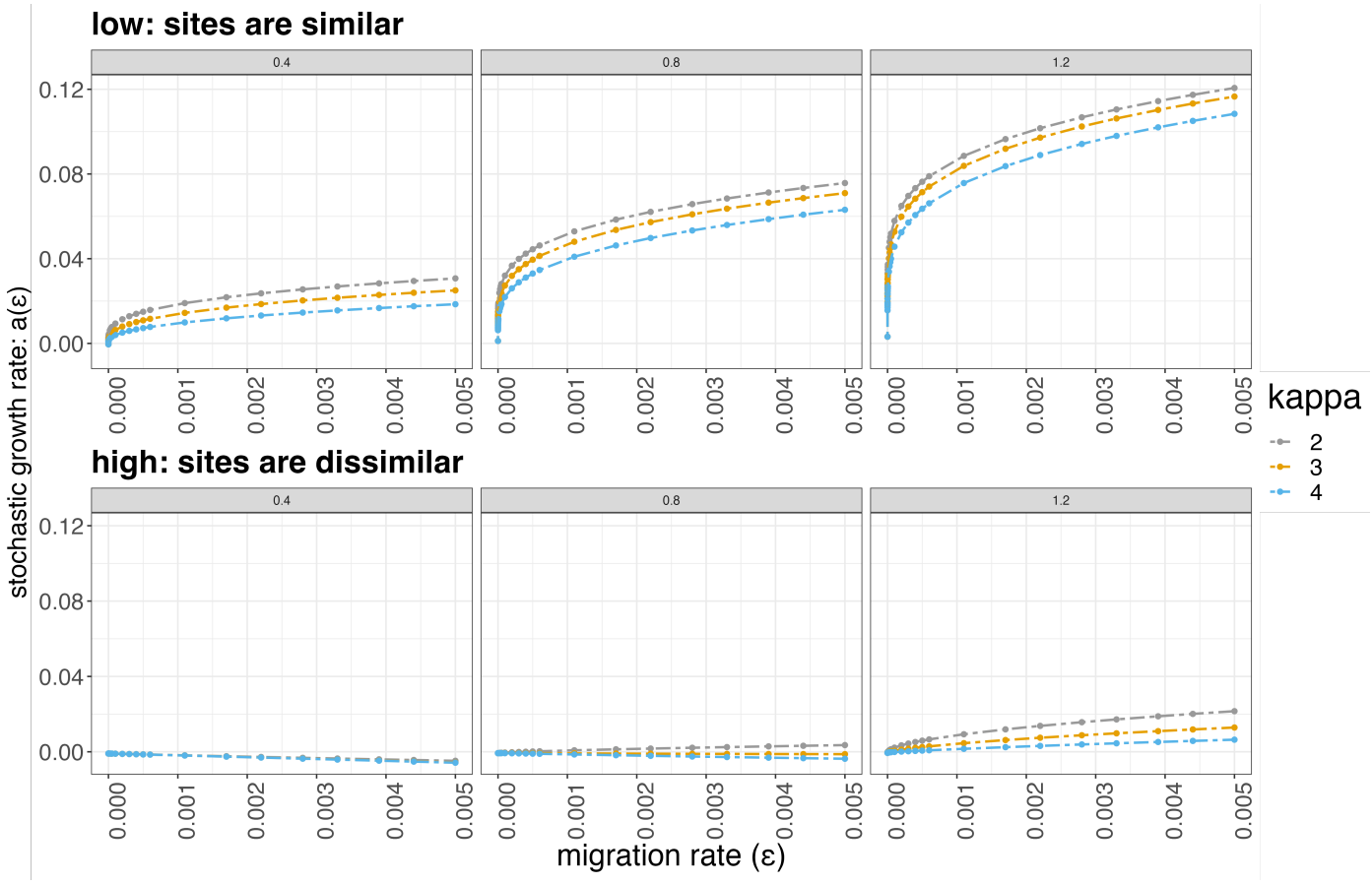


Figure 10: The x-axis is migration rate and y-axis is the stochastic growth rate: $a(\varepsilon)$. The top panel corresponds to values for the vector $(\tilde{\mu}_0, \tilde{\mu}_1, \tilde{\mu}_2)$ such that the sites are similar. The panel on the bottom corresponds to values such that sites are dissimilar. Each color corresponds to a value of κ (for four sites) since an excursion can be made in steps of 2, 3 and 4.

4.2.2 Effect of varying path length between sites

We now examine how the length of κ affects the long-term growth rate. Recall that κ is defined to be the length of shortest cycle in the migration graph that starts and ends at 0, with no intervening returns to 0.

In Figure 10, each color corresponds to a value of κ for four sites since an excursion can be made in lengths of 2, 3 and 4. From the theorem, we expected the shortest excursion with $\kappa = 2$ to lead to the maximum gain in stochastic growth rate, and that is what we find in Figure 10. The longest

excursion with $\kappa = 4$ leads to the lowest increase in stochastic growth rate. Although populations sample more variability by increasing the number of possible paths and sites, it is important to return to the best site in the shortest sequence possible to yield gain in overall growth rate. That is, increase in growth rate is determined by being able to gain access to the best site and spending more time around the best site (lower κ).

In addition, as evident from the last two panels in Figure 10, the length of κ matters when the fluctuations are high and when the sites are more similar to the best site. As depicted in bottom two panels of Figure 11, when the difference between sites is increased (high μ), the stochastic growth rates for different path lengths remain similar. Hence, path length matters only when the sites are similar as shown in top two panels of Figure 11.

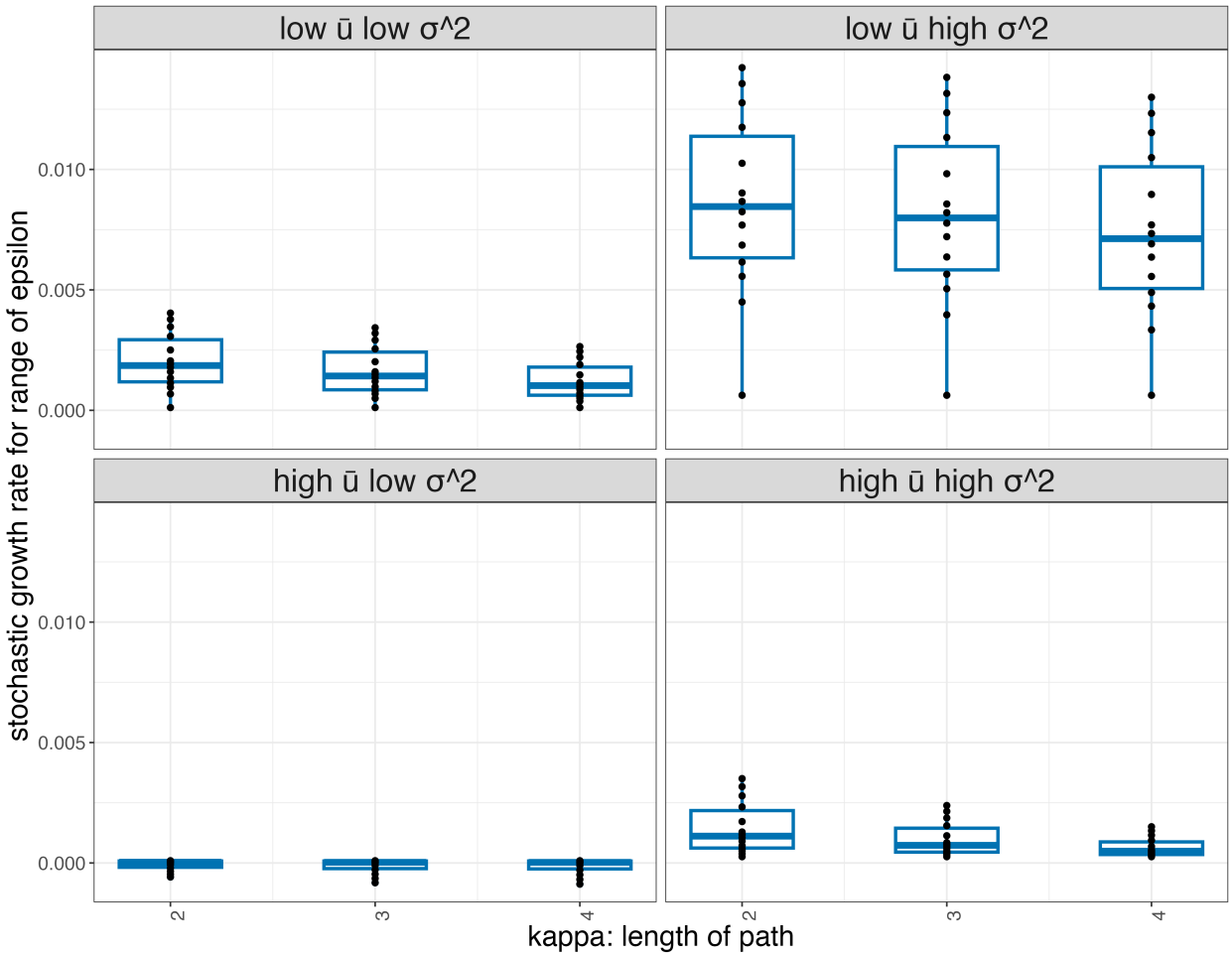


Figure 11: The effect of changing topology on growth rates in a 4-site model. The x-axis is the length of the shortest path to the best site (site 0) in steps of 2, 3 and 4. The y-axis are the values of stochastic growth rate for a range of ε. Each panel corresponds to a fixed level of σ^2, and varies in low and high difference in expected growth rates from best site.

5 Discussion

Theoretical work on the evolution of dispersal has a long history as illustrated by our many citations (Hamilton and May 1977; Levin et al. 1984; McPeek and Holt 1992; Hanski and Simberloff 1997; Travis and Dytham 1999; Altenberg et al. 2017). We add to this literature by showing that the increase in long-run growth rate, $(a(\varepsilon) - a(0))$ with migration at rate ε approximately behaves like the power of ε and under plausible conditions, for the two-site case, the power depends on the ratio $\tilde{\mu}/\sigma^2$. Here, $\tilde{\mu}$ is the difference in expected growth rates and σ^2 is the variance of the difference of site-specific growth rates at time t. Our results provide a precise analysis that is applicable in the limit of zero migration and as a consequence to the question of evolution of migration in heterogeneous environments. This is important because developing an explicit formula for stochastic growth rate has proved to be challenging in variable environments. In addition, our findings provide conditions for departures from the reduction principle in changing environments thereby providing a solution to the one of the questions posed by Altenberg (2012).

In the simulations we examine a wide range of fluctuations and the differences in average growth rates between sites. We find as the sites become more dissimilar, the increase in long-run growth rate due to introduction of migration becomes less pronounced. Environmental fluctuations compensate (to an extent) for the fixed difference in average growth rates. Biologically, this means that habitat quality and fluctuations are both important determinants of migration evolution and population persistence. We also examine the impact that the length of shortest path, κ has on long-run growth rate by changing how each site is connected to the 'best' site. We find that when variability σ^2 is low, excursions of any length are unlikely to overcome the average deficit on growth (refer to Figure 11). When variability σ^2 is high, excursions may overcome average differences when the length of κ is small and will yield the maximum increase in stochastic growth rate in that case. However there are recent studies for seabirds (Dufour et al. 2021; Kentie et al. 2023) that show that a longer migration route does not necessarily have carry-over effects on subsequent reproduction and apparent survival for little auk *(Alle alle)* and lesser black-backed gulls *(Larus fuscus)*, respectively. We also did the analysis for positively and negatively correlated fluc-

tuations and our key results remain unchanged.

In variable environments, migration will be favored by natural selection, in the sense that introducing a small amount of migration will increase the growth rate of the population relative to the zero-migration case. More interestingly, as fluctuations in growth rates (σ^2) increase, introduction of even small migration goes a long way in increasing the overall growth rate in the long run which is consistent with earlier work that found similar singular behaviour (Levin 1976; Motro 1982; Derrida and Hilhorst 1983; Wiener and Tuljapurkar 1994). Thus heterogeneity favors the evolution of mechanisms as dispersal allows escape in space from locally unfavorable conditions and, on the average, exploitation of locations that are temporarily more favorable.

Our findings may have implications in the wake of global habitat degradation and climate change as dispersal can help populations ride-out fluctuations and increase their fitness in the long run. Earlier work hypothesized the importance of migration in rapidly-changing environments. Southwood (1962) intuited that the rate of migration should be highest in species inhabiting the most temporary habitats and found that migratory movement in terrestrial Arthropoda taxa is positively correlated with impermanence of the habitat. Den Boer (1968) emphasized that the chance of survival of a population may even increase, as dispersal may make it possible to cope with the variation in space and time of the habitat. In his models, den Boer argued that migration may increase long-run growth rate when there is independent or weakly-correlated stochastic variation in growth among sites. Numerous empirical studies have found that dispersal is more common in rapidly-changing habitats; e.g., Denno et al. (1991) studied 35 species of plant hoppers living in habitats of varying persistence, and found that winged morphs were more common in ephemeral habitats while flightlessness was favoured in persistent habitats. Many other examples are given by Tuljapurkar (1990c). These empirical and theoretical arguments have been extended by analyses showing that spread in space (dispersal) is important in variable environments (Cohen 1966; Wiener and Tuljapurkar 1994; Tuljapurkar 1990c; Tuljapurkar and Istock 1993).

There are several directions for future research. There is a dearth of empirical studies to test or parameterize theoretical models (Ronce 2007). In fragmented landscapes, there are high costs of

dispersal leading to decrease in dispersal evolution. Cheptou et al. (2008) studied seed dispersal in the weed *Crepis sancta* in an urban environment and showed that a high cost of dispersal after recent fragmentation causes rapid evolution toward lower dispersal. A decrease of dispersal at higher fragmentation levels has been observed in other various animal or plant species (Bonte et al. 2006; Schtickzelle et al. 2006). Whether life-history traits and population structure or environmental factors are more influential for the evolution of dispersal is another area for further research. One of the challenges is to elucidate how environmental factors influence the dispersal decision of individuals with distinct phenotypes. In their work on butterflies, Legrand et al. (2015) suggest that the environmental context, and especially weather conditions and habitat quality, prevails over social factors and individual phenotypes in butterflies' decision to disperse. The scale and frequency of environmental fluctuations is not considered explicitly and can control the evolution of dispersal strategy (Nakamaru et al. 2007). In their model, Cadet et al. (2003) show that demographic stochasticity may provide enough variability between and within habitats to favor the evolution of dispersal, even if the environment itself is homogeneous in time and space. In models with relatively constant spatial variation in landscape quality, different dispersal rates evolve in narrow corridors and on the boundaries of good habitats than in the centers of habitats, with potentially important implications for conservation (Travis and Dytham 1999). In addition, Travis (2001) find that when spatial and temporal variability exhibits red rather than white noise (increasing rather than constant variance with time or distance), there is reduced selection for dispersal.

References

Altenberg, Lee (2010). "Karlin theory on growth and mixing extended to linear differential equations". *arXiv preprint arXiv:1006.3147*.

— (2012). "The evolution of dispersal in random environments and the principle of partial control". *Ecological Monographs* 82.3, pp. 297–333.

Altenberg, Lee and Marcus W Feldman (1987). “Selection, generalized transmission and the evolution of modifier genes. I. The reduction principle”. *Genetics* 117.3, pp. 559–572.

Altenberg, Lee, Uri Liberman, and Marcus W Feldman (2017). “Unified reduction principle for the evolution of mutation, migration, and recombination”. *Proceedings of the National Academy of Sciences* 114.12, E2392–E2400.

Andrewartha, Herbert George, L Charles Birch, et al. (1954). *The distribution and abundance of animals.* Edn 1. University of Chicago press.

Baker, Robin et al. (1978). *Evolutionary ecology of animal migration*. Holmes & Meier Publishers.

Benaïm, Michel and Sebastian J Schreiber (2009). “Persistence of structured populations in random environments”. *Theoretical population biology* 76.1, pp. 19–34.

Boer, PJ den (1968). “Spreading of risk and stabilization of animal numbers”. *Acta biotheoretica* 18.1, pp. 165–194.

Bonte, Dries, Jeroen Vanden Borre, Luc Lens, and Jean-Pierre Maelfait (2006). “Geographical variation in wolf spider dispersal behaviour is related to landscape structure”. *Animal behaviour* 72.3, pp. 655–662.

Boucheron, Stéphane, Gábor Lugosi, and Pascal Massart (2013). *Concentration inequalities: A nonasymptotic theory of independence*. Oxford University Press.

Buldygin, V. V. and Yu. V. Kozachenko (2000). *Metric characterization of random variables and random processes*. Vol. 188. American Mathematical Soc.

Cadet, Claire, Régis Ferrière, Johan AJ Metz, and Minus Van Baalen (2003). “The evolution of dispersal under demographic stochasticity”. *The American Naturalist* 162.4, pp. 427–441.

Cheptou, P-O, O Carrue, S Rouifed, and A Cantarel (2008). “Rapid evolution of seed dispersal in an urban environment in the weed Crepis sancta”. *Proceedings of the National Academy of Sciences* 105.10, pp. 3796–3799.

Cohen, Dan (1966). “Optimizing reproduction in a randomly varying environment”. *Journal of theoretical biology* 12.1, pp. 119–129.

Cohen, Dan and Simon A Levin (1991). “Dispersal in patchy environments: the effects of temporal and spatial structure”. *Theoretical Population Biology* 39.1, pp. 63–99.

Cohen, Joel E (1979). “Ergodic theorems in demography”. *Bulletin of the American Mathematical Society* 1.2, pp. 275–295.

Dembo, Amir and Ofer Zeitouni (2009). “LDP for Finite Dimensional Spaces”. *Large deviations techniques and applications*. Springer, pp. 11–70.

Den Boer, Pieter J (1968). “Spreading of risk and stabilization of animal numbers”. *Acta biotheoretica* 18.1, pp. 165–194.

Denno, Robert F, George K Roderick, Karen L Olmstead, and Hartmut G Dobel (1991). “Density-related migration in planthoppers (Homoptera: Delphacidae): the role of habitat persistence”. *The American Naturalist* 138.6, pp. 1513–1541.

Derrida, Bernard and HJ715727 Hilhorst (1983). “Singular behaviour of certain infinite products of random 2×2 matrices”. *Journal of Physics A: Mathematical and General* 16.12, p. 2641.

Dufour, Paul et al. (2021). “A two-fold increase in migration distance does not have breeding consequences in a long-distance migratory seabird with high flight costs”. *Marine Ecology Progress Series* 676, pp. 117–126.

Feldman, Marcus W, Freddy B Christiansen, and Lisa D Brooks (1980). “Evolution of recombination in a constant environment”. *Proceedings of the National Academy of Sciences* 77.8, pp. 4838–4841.

Feldman, Marcus W and Uri Liberman (1986). “An evolutionary reduction principle for genetic modifiers”. *Proceedings of the National Academy of Sciences* 83.13, pp. 4824–4827.

Gadgil, Madhav (1971). “Dispersal: population consequences and evolution”. *Ecology* 52.2, pp. 253–261.

Hamilton, William D and Robert M May (1977). “Dispersal in stable habitats”. *Nature* 269.5629, pp. 578–581.

Hanski, Ilkka (1998). “Metapopulation dynamics”. *Nature* 396.6706, pp. 41–49.

Hanski, Ilkka and Daniel Simberloff (1997). “The metapopulation approach, its history, conceptual domain, and application to conservation”. *Metapopulation biology*, pp. 5–26.

Hansson, Lars-Anders, Susanne Akesson, and Susanne Åkesson (2014). *Animal movement across scales*. Oxford University Press.

Karlin, Samuel (1982). “Classifications of selection-migration structures and conditions for a protected polymorphism”. *Evol. Biol* 14.61, p. 204.

Kentie, Rosemarie, J Morgan Brown, Kees CJ Camphuysen, and Judy Shamoun-Baranes (2023). “Distance doesn’t matter: migration strategy in a seabird has no effect on survival or reproduction”. *Proceedings of the Royal Society B* 290.1997, p. 20222408.

Khaladi, Mohamed, Jean-Dominique Lebreton, and Abdelaziz Khermjioui (2012). “The evolution of dispersal in random environment”. *Acta biotheoretica* 60.1, pp. 155–165.

Lande, Russell and Steven Hecht Orzack (1988). “Extinction dynamics of age-structured populations in a fluctuating environment”. *Proceedings of the National Academy of Sciences* 85.19, pp. 7418–7421.

Legrand, Delphine et al. (2015). “Ranking the ecological causes of dispersal in a butterfly”. *Ecography* 38.8, pp. 822–831.

Levin, Simon A (1976). “Population dynamic models in heterogeneous environments”. *Annual review of ecology and systematics*, pp. 287–310.

Levin, Simon A, Dan Cohen, and Alan Hastings (1984). “Dispersal strategies in patchy environments”. *Theoretical population biology* 26.2, pp. 165–191.

Levins, Richard (1970). “Extinction”. *Some mathematical questions in biology*.

Lewontin, Richard C and Daniel Cohen (1969). “On population growth in a randomly varying environment”. *Proceedings of the National Academy of sciences* 62.4, pp. 1056–1060.

May, Robert M (1971). “Stability in model ecosystems”. *Proc. Ecol. Soc. Aust*. Vol. 6, pp. 18–56.

McNamara, John M and Sasha RX Dall (2011). “The evolution of unconditional strategies via the ‘multiplier effect’”. *Ecology letters* 14.3, pp. 237–243.

McPeek, Mark A and Robert D Holt (1992). “The evolution of dispersal in spatially and temporally varying environments”. *The American Naturalist* 140.6, pp. 1010–1027.

Motro, Uzi (1982). “Optimal rates of dispersal I. Haploid populations”. *Theoretical Population Biology* 21.3, pp. 394–411.

Nakamaru, Mayuko, Yayoi Beppu, and Kazuki Tsuji (2007). “Does disturbance favor dispersal? An analysis of ant migration using the colony-based lattice model”. *Journal of Theoretical Biology* 248.2, pp. 288–300.

Oldfather, Meagan F, Courtney L Van Den Elzen, Patrick M Heffernan, and Nancy C Emery (2021). “Dispersal evolution in temporally variable environments: implications for plant range dynamics”. *American journal of botany* 108.9, pp. 1584–1594.

Roerdink, JBTM (1988). “The biennial life strategy in a random environment”. *Journal of Mathematical Biology* 26.2, pp. 199–215.

Roff, DA (1974a). “Spatial heterogeneity and the persistence of populations”. *Oecologia* 15.3, pp. 245–258.

— (1974b). “The analysis of a population model demonstrating the importance of dispersal in a heterogeneous environment”. *Oecologia* 15.3, pp. 259–275.

Ronce, Ophélie (2007). “How does it feel to be like a rolling stone? Ten questions about dispersal evolution”. *Annu. Rev. Ecol. Evol. Syst.* 38, pp. 231–253.

Schreiber, Sebastian J (2012). “The evolution of patch selection in stochastic environments”. *The American Naturalist* 180.1, pp. 17–34.

— (2010). “Interactive effects of temporal correlations, spatial heterogeneity and dispersal on population persistence”. *Proceedings of the Royal Society B: Biological Sciences* 277.1689, pp. 1907–1914. DOI: 10.1098/rspb.2009.2006. eprint: http://rspb.royalsocietypublishing.org/content/277/1689/1907.full.pdf+html. URL: http://rspb.royalsocietypublishing.org/content/277/1689/1907.abstract.

Schtickzelle, Nicolas, Gwénaëlle Mennechez, and Michel Baguette (2006). “Dispersal depression with habitat fragmentation in the bog fritillary butterfly”. *Ecology* 87.4, pp. 1057–1065.

Southwood, TRE (1962). “Migration of terrestrial arthropods in relation to habitat”. *Biological reviews* 37.2, pp. 171–211.

Steinsaltz, David and Shripad Tuljapurkar (2015). “Stochastic growth rates for life histories with rare migration or diapause”. *arXiv preprint arXiv:1505.00116*.

— (2018a). “Stability of fixed life histories to perturbation by rare diapause”. *arXiv preprint arXiv:1809.04027*.

— (2018b). “Stochastic Models for Structured Populations”. Ed. by A. R. Srinivasa Rao and C R Rao. Vol. 40. Handbook of Statistics. Elsevier. Chap. 4, pp. 133–155. DOI: https://doi.org/10.1016/bs.host.2018.11.005.

Travis, Justin MJ (2001). “The color of noise and the evolution of dispersal”. *Ecological Research* 16.1, pp. 157–163.

Travis, Justin MJ and Calvin Dytham (1999). “Habitat persistence, habitat availability and the evolution of dispersal”. *Proceedings of the Royal Society of London. Series B: Biological Sciences* 266.1420, pp. 723–728.

Tuljapurkar, S. (1990a). *Population dynamics in variable environments*. New York: Springer-Verlag, Berlin.

Tuljapurkar, S. D. (1990b). *Population dynamics in variable environments*. Lecture notes in biomathematics, 85. New York: Springer-Verlag.

Tuljapurkar, Shripad (1989). “An uncertain life: demography in random environments”. *Theoretical population biology* 35.3, pp. 227–294.

— (1990c). “Delayed reproduction and fitness in variable environments.” *Proceedings of the National Academy of Sciences* 87.3, pp. 1139–1143.

— (1994). “Frontiers in Mathematical Biology”. Ed. by Simon A. Levin. Berlin, Heidelberg: Springer Berlin Heidelberg. Chap. Stochastic Demography and Life Histories.

Tuljapurkar, Shripad, Carol C Horvitz, and John B Pascarella (2003). “The many growth rates and elasticities of populations in random environments”. *The American Naturalist* 162.4, pp. 489–502.

Tuljapurkar, SHRIPAD and CONRAD Istock (1993). “Environmental uncertainty and variable diapause”. *Theoretical Population Biology* 43.3, pp. 251–280.

Wiener, P. and S. Tuljapurkar (1994). “Migration in variable environments: exploring life-history evolution using structured population models.” *J Theor Biol* 166.1, pp. 75–90.

Appendix

Notation and basic assumptions

- Let there be d sites such that populations $0,\ldots,d-1$ have log growth rates $X_t^{(i)}$ during time period t at site i. The vectors $X_t = (X_t^{(0)},\ldots,X_t^{(d-1)})$ are assumed independent, but the components within each X_t — corresponding to growth rates at different sites at the same time — may be dependent.

- We write $\mu_i := \mathbb{E}[X_t^{(i)}]$, and assume that μ_0 is strictly the largest of these. Writing $\tilde{\mu}_i := \mu_0 - \mu_i$, these are all positive. We write $\mathbf{X}$ for the complete collection of all $X_t^{(j)}$ for $j = 0,1,\ldots,d-1$, $0 \le t < \infty$.

- The random variables $\widetilde{X}_t^{(i)} := X_t^{(i)} - X_t^{(0)}$ are assumed sub-Gaussian with variance factor τ_i, defined below in (3). When they are Gaussian, as we have assumed for the simulations, τ_i is then simply the variance.

- Suppose $D_1, D_2, \ldots$ is an i.i.d. sequence of $d \times d$ diagonal matrices. The diagonal elements of D_t are the log growth rates at each site at time t such that $X_t^{(i)}$ all have finite mean μ_i and finite variance.

- We let A_t be an i.i.d. sequence of nonnegative $d \times d$ matrices with zeros on the diagonal, and we assume that the pairs (D_t, A_t) are all independent. We define the migration graph $\mathscr{M}$ to be a directed graph whose vertices are the sites $\{0,\ldots,d-1\}$, possessing an edge (i,j) if $\operatorname{ess\,inf} \mathbb{E}\left[\log A_t(j,i) \mid D_t\right] > -\infty$. That is, we do not assume that A_t and D_t are in-

dependent, but for the migration steps we consider to be possible we assume that there is a lower bound to how close $A_t(j,i)$ can come to 0 that is independent of all D_t. It follows trivially that when (i,j) is an edge $\mathbb{E}\left[\log A_t(j,i) \mid \mathscr{D}\right] > -\infty$ where $\mathscr{D}$ is the σ-algebra generated by $\{D_t, 1 \leq t < \infty\}$. We assume that $\mathscr{M}$ is connected.

We define the *migration graph* $\mathcal{M}$ to be a simple and irreducible directed graph whose vertices are the sites $\{0,\ldots,d-1\}$, representing the transitions that have nonzero probability. We let A_t be an i.i.d. sequence of nonnegative $d \times d$ matrices, representing migration rates in time-interval t. Conventionally migration rates from site i to site j are placed in the (i,j) position of the matrix —corresponding to representing populations by row vectors, and updating by right-multiplication. Because we wish to be consistent with the life-history-stage interpretation, and thence with the conventions that follow from Leslie matrices in the population ecology literature we adopt here the opposite convention, with migration rates from site i to site j in $\varepsilon A_t(j,i)$. Here ε parametrises the overall rate of migration. Population distributions are thus naturally column vectors, and are updated from time $t-1$ to time t by left multiplication. Node i is not connected by a directed edge to node j — written $i \nrightarrow j$ — precisely when $\mathbb{P}\{A_t(j,i) = 0\} = 1$.

When $i \to j$ (i is directly linked to j in $\mathcal{M}$) we assume $\mathbb{E}[(\log A_t(j,i))_+]$ exists and is finite, though $A_t(j,i)$ may still have nonzero probability of being 0. We write $\mu_A :=$ $\max_{i,j} \mathbb{E}[(\log A_t(j,i))_+]$.

The diagonal elements of A_t represent costs of migration. If we are thinking of these simply as individuals lost from the originating site to out-migration then it would be reasonable to require that the total of all migration from a site $\sum_j A_t(j,i)$ is 0, or negative if some individuals are lost in the process of migration, but this is not a formal requirement. We will assume that $A_t(i,i)$ is almost surely bounded below, and we will henceforth always be assuming that ε is sufficiently small that $1 + \varepsilon A_t(i,i) > 0$ for all i with probability 1. We write $\Delta^i := -\mathbb{E}[A_t(i,i)]$, which we assume also to be finite.

The random variables $\widetilde{X}_t^{(i)} := X_t^{(i)} - X_t^{(0)}$ are assumed sub-Gaussian with variance factor τ_i. Recall that a random variable Z is said to be *sub-Gaussian* if it has finite *variance factor* (the term applied in (Boucheron et al. 2013)) $\tau = \tau(Z)$, defined by

$$\tau := \inf\left\{c \geq 0 : \log \mathbb{E}\left[\mathrm{e}^{\lambda(Z-\mathbb{E}[Z])}\right] \leq \frac{c\lambda^2}{2} \forall \lambda \in \mathbf{R}\right\}. \tag{3}$$

($\sqrt{\tau(Z)}$ is called the *sub-Gaussian standard* in (Buldygin and Kozachenko 2000).) This may be thought of as an upper bound on the scale of the tails, and it is this that determines the upper bound on the sensitivity of the stochastic growth rate to migration. Note that the Legendre transform for the moment generating function $I_Z(z) := \sup_\lambda \{\lambda z - \log \mathbb{E}[\mathrm{e}^{\lambda(Z-\mathbb{E}[Z])}]\}$ satisfies $I_Z(z) \geq z^2/2\tau(Z)$ for all z. Applying Markov's inequality implies the standard sub-Gaussian concentration inequality

$$\mathbb{P}\left\{Z \geq \mathbb{E}[Z] + z\right\} \leq \mathrm{e}^{-z^2/2\tau(Z)}. \tag{4}$$

Like variances, sub-Gaussian variance factors are also sub-additive for sums of independent random variables, as stated in (Buldygin and Kozachenko 2000) (Lemma 1.7):

Lemma 3. *If $Z_1, \ldots, Z_k$ are independent random sub-Gaussian random variables with variance factors $\tau_1, \ldots, \tau_k$. Then $Z_1 + \cdots + Z_k$ is sub-Gaussian with variance factor less than or equal to $\tau_1 + \cdots + \tau_k$.*

To describe the lower bounds we define a lower subvariance $\underline{\tau}(Z)$ by

$$\underline{\tau}(Z) := \sup\left\{c \geq 0 : \log \mathbb{E}\left[\mathrm{e}^{\lambda(Z-\mathbb{E}[Z])}\right] \geq \frac{c^2\lambda^2}{2} \forall \lambda \in \mathbf{R}\right\}. \tag{5}$$

Then for all z we have $I_Z(z) \leq z^2/2\tau(Z)$. If Z is Gaussian with variance σ^2 then of course

$\tau(Z) = \underline{\tau}(Z) = \sigma$. We define also

$$\rho_j := \frac{\tilde{\mu}_j}{\tau_j} \qquad \text{and} \qquad \underline{\rho}_j := \frac{\tilde{\mu}_j}{\underline{\tau}_j}. \tag{6}$$

The calculation of ρ is illustrated in Figure 12.

We define the *migration graph* $\mathcal{M}$ to be a simple and irreducible directed graph whose vertices are the sites $\{0,\ldots,d-1\}$, representing the transitions that have nonzero probability. We let A_t be an i.i.d. sequence of nonnegative $d \times d$ matrices with zeros on the diagonal, representing migration rates in time-interval t. Conventionally migration rates from site i to site j are placed in the (i,j) position of the matrix — corresponding to representing populations by row vectors, and updating by right-multiplication. Because we wish to be consistent with the life-history-stage interpretation adopted in (Steinsaltz and Tuljapurkar 2018a), and thence with the conventions that follow from Leslie matrices in the population ecology literature we adopt here the opposite convention, with migration rates from site i to site j in $\varepsilon A_t(j,i)$. Here ε parametrises the overall rate of migration. Population distributions are thus naturally column vectors, and are updated from time $t-1$ to time t by left multiplication. Node i is not connected by a directed edge to node j — written $i \nrightarrow j$ — precisely when $\mathbb{P}\{A_t(j,i) = 0\} = 1$. When $i \to j$ (i is directly linked to j in $\mathcal{M}$) we assume $\mathbb{E}[(\log A_t(j,i))_+]$ exists and is finite, though $A_t(j,i)$ may still have nonzero probability of being 0. We write $\mu_A := \max_{i,j} \mathbb{E}[(\log A_t(j,i))_+]$.

While we assume that the collection of pairs $(X_t, A_t)_{t=0}^{\infty}$ is jointly independent (and identically distributed), we do not assume for a given t that A_t and X_t are independent. We assume that the random variables $\log A_t(i,j)$ for $i \neq j$ — and hence the combinations $\widetilde{X}_t^{(j)} + \log A_t(i,j)$ — are all sub-Gaussian, and we write

$$\tau_A := \max_{j \in \{0,\ldots,d-1\}} \frac{\tilde{\mu}_j}{\tau_j} \max_{i \neq \ell} \left[\tau\left(\widetilde{X}_t^{(i)} + \Big(\log A_t(\ell,i)\Big)_+ \right) - \tau_i \right]. \tag{7}$$

Finally, we have random penalties $\Delta_t^0, \ldots, \Delta_t^{d-1}$ that also act proportionately to ε. (We think of Δ as the growth penalty for migration, which makes it natural to suppose the entries will be negative, but this is not essential.) We assume that these penalties are sub-Gaussian, and write τ_Δ for the maximum variance factor.

We define

$$D_t(\varepsilon) := \exp\left(\operatorname{diag}(X_t + \varepsilon\Delta_t)\right)(I_d + \varepsilon A_t).$$

For $\varepsilon > 0$ the i.i.d. sequence $D_t(\varepsilon)$ satisfies the conditions for the existence of a stochastic growth rate independent of starting condition.(Cohen 1979) That is, if we define the partial products

$$R_T(\varepsilon) := D_{T-1}(\varepsilon) \cdot D_{T-2}(\varepsilon) \cdot \cdots \cdot D_0(\varepsilon)$$

then

$$a(\varepsilon) := \lim_{T\to\infty} T^{-1} \log R_T(\varepsilon)_{ij}$$

are well defined deterministic quantities, in the sense that the limit exists almost surely, is almost-surely constant, and is the same for any $0 \le i, j \le d-1$.

Of course, $R_T(0)$ has a different structure: The off-diagonal terms are all 0, while on the diagonal, by the Strong Law of Large Numbers,

$$\lim_{T\to\infty} T^{-1} \log R_T(0)_{ii} = \mu_i.$$

Remark 4. *The effect of migration on population growth is affected by the structure of the migration graph — the contribution of a site to the growth rate is weakened if multiple migration steps are required to reach it and to return to 0 — but the impact of the behavior of the site itself is determined by the "cost" in reduced expected log growth, scaled by the variance of the difference. It is clear that a greater difference $\tilde{\mu}_j$ will make the rare events when growth at the inferior site happens to exceed growth at site 0 even rarer, while a greater variance (or gener-*

alized variance τ_j) will make such exceptional deviations more likely. Thus, high variance may compensate for a large difference in mean log growth. It may seem surprising, though, that the mean difference is scaled by the variance, rather than by the standard deviation. The exact form that this ratio takes comes out of the proof of Theorem 6, and we do not know of a simpler intuitive explanation. It is easy to see, though, that ρ needs to be scaled by a larger function than $\sqrt{\tau}$ (the standard deviation) in the denominator. In the case of Gaussian log increments, scaling by the standard deviation would mean that doubling all the log increments should leave the migration effect unchanged. In fact, this transformation will leave the frequency of sequences of times whose cumulative growth at site j exceeds that at site 0. But when such an exceedance occurs, it will be twice as large. Thus changing the mean difference and the standard deviation by the same factor is expected to **increase** *the effect of migration.*

The effect of the penalty Δ

We be concerned primarily with analyzing the case $A(j,j) \equiv 0$. The crucial point is that the effect of the diagonal terms in A is always nearly linear in ε, while the increase of a near 0 is often superlinear, growing as ε^{β}. If the power β is strictly less than 1, the rapid increase in a near 0 will be qualitatively unaffected by a linear term for ε sufficiently small. Even when the linear term is negative (as we will generally be assuming it to be), the growth rate a will still be increasing on a small interval of $\varepsilon > 0$.

On the other hand, as discussed in section 5, in some cases we cannot exclude the possibility that the growth rate when $A(0,0) \equiv 0$ is qualitatively like ε^{β} with $\beta \geq 1$. If $\beta > 1$ and $\Delta^0 < 0$ then a will be decreasing near 0; if $\beta = 1$ then a more sensitive analysis would be required.

Since the upper and lower bounds on the appropriate power of ε in Theorem 6 are distinct, with the lower bound on the growth rate (the upper bound on the power of ε) being sometimes larger than 1, the current results will not always permit us to ascertain whether the growth rate increases or decreases for small increases in ε.

The following bound follows directly from Hölder's inequality, with $p = 1 + \varepsilon\sqrt{\tau_\Delta/\tau_i}$ and $q =$

$1+\sqrt{\tau_i/\tau_\Delta}/\varepsilon$:

Lemma 5. *The random variables $X_t^{(i)}+\varepsilon A_t(i,i)$ are sub-Gaussian, with variance factor bounded by $(\sqrt{\tau_i}+\varepsilon\sqrt{\tau_\Delta})^2$.*

Main result

If $d=2$ then $a(\varepsilon)-a(0)$ is bounded above by a constant times $\varepsilon^{4\tilde{\mu}_1/(2\tilde{\mu}_1+\tau_1)}$, and lower bound $\varepsilon^{4\tilde{\mu}_1/\tau_1}$. For $d>2$ this becomes slightly more complicated for two reasons: First, the effect of migration will tend to be dominated by the site that has the second fastest growth; second, the increment to growth will be smaller if direct transition between the two leading sites is impossible. For this purpose, for each $1\le j\le d-1$ we define κ_j to be the smallest length of a cycle in $\mathcal{M}$ that starts and ends at 0, and passes through j. (Thus $2d-2\ge\kappa_j\ge 2$, and κ_j is equal to 2 when $0\to j$ and $j\to 0$.)

Theorem 6. *Under the above assumptions, let $\underline{\eta}=\min_j 2\kappa_j\underline{\rho}_j$, and $\bar{\eta}=\min_j 2\kappa_j\rho_j/(1+2\rho_j)$. Then for any $\delta>0$*

$$\varepsilon^{\underline{\eta}+\delta}-\varepsilon\Delta_t^0\le a(\varepsilon)-a(0)\le\varepsilon^{\bar{\eta}-\delta}-\varepsilon\Delta_t^0 \tag{8}$$

for $\varepsilon>0$ sufficiently small. If $A_t(i,i)\ge A_t(0,0)$ almost surely for all i then we may strengthen the upper bound to

$$a(\varepsilon)-a(0)\le\varepsilon^{\bar{\eta}}(\log\varepsilon^{-1})^c-\varepsilon\Delta_t^0, \tag{9}$$

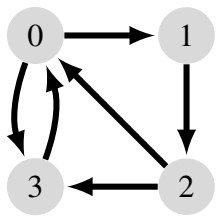
for some positive constant c depending only on $\mathcal{M}$.

for all $\varepsilon>0$ sufficiently small.

Remark 7. *In general, the upper bound is dominated by the lower power of ε, and the lower bound is dominated by the higher power of ε. Thus*

- *If $\underline{\eta}<1$ then the linear penalty is irrelevant (asymptotically), and for ε sufficiently small*

$$\varepsilon^{\underline{\eta}+\delta}\le a(\varepsilon)-a(0)\le\varepsilon^{\bar{\eta}-\delta}; \tag{10}$$



(a) The migration graph $\mathcal{M}$

$$\begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{cccc} 0 & 1 & 2 & 3 \end{array} \\ \begin{pmatrix} 1 & -0.5 & 0.7 & 0.5 \\ -0.5 & 1 & 0 & 0 \\ 0.7 & 0 & 1 & -0.3 \\ 0.5 & 0 & -0.3 & 2 \end{pmatrix} \end{array}$$

(b) $\mathrm{Cov}(X_t^{(i)}, X_t^{(j)})$

j	$\mathbb{E}[X_t^{(j)}]$	τ_j	ρ_j	κ_j	$\kappa_j\rho_j$	$\frac{\kappa_j\rho_j}{1+2\rho_j}$
0	1					
1	0.4	3	0.2	3	0.6	0.43
2	0.7	0.6	0.5	3	1.5	0.75
3	0.3	2	0.35	2	0.7	0.41

(c) Selecting the second site

Figure 12: Calculating ρ on a four-site graph. For simplicity, we assume the variables X_t are all Gaussian, so that $\tau^2 = \underline{\tau}^2$ is the variance, and $\rho = \underline{\rho}$. We see that $\kappa\rho$ is minimized at site 1, despite the fact that it is not in the shortest cycle, nor does it have the smallest mean difference in log growth rate from the best site 0, yielding a lower bound on the sensitivity to migration on the order of $\varepsilon^{0.6}$. The upper bound $\varepsilon^{0.41}$ comes from migration to site 3, which has the lowest mean log growth rate, but compensates for this with a shorter cycle length.

- *If $\bar{\eta} < 1 \leq \underline{\eta}$ then for ε sufficiently small*

$$a(\varepsilon) - a(0) \leq \varepsilon^{\bar{\eta}-\delta}; \tag{11}$$

- *If $\bar{\eta} > 1$ then a is differentiable at 0, with $a'(0) = -\Delta^0$.*

Remark 8. *Note that $\underline{\tau}_j = 0$, hence $\underline{\rho}_j = \infty$, when the distribution of $e^{\widetilde{X}_t^{(j)}}$ is not heavy-tailed — for example, very natural choices such as gamma-distributed growth rates on the diagonal — making the lower bound on the left-hand side vacuous, but it remains an open question whether zero subvariance implies that the approach of $a(\varepsilon)$ to 0 is faster than polynomial in ε.*

Excursion decompositions

Since we are assuming the unique maximum average growth rate is at site 0, the maximum growth for the perturbed process will arise from rare excursions away from 0; in particular, from those that include the (not necessarily unique) site that minimises $\rho\kappa$ in (6).

Define $\mathscr{E}$ to be the set — called the *excursions* — of cycles in the migration graph that start and end at 0, with no intervening returns to 0. For an excursion $\mathbf{e}$ we write $|\mathbf{e}|$ for the length of the cycle minus 2 — that is, the number of time steps spent away from 0.

As is standard in statistical mechanics, the change in stochastic growth rate will ultimately be determined by the tradeoff between the penalty for migration and for spending time away from the best site, and the entropy gain from increasing the number of possible paths as the number of allowed transitions grows. To calculate this tradeoff quantitatively we need a notation that will allow us to enumerate the paths — defined by a sequence of excursions — that will be draw similar penalties. For this purpose the essential features will be the number of excursions in the path, total length of the excursions, and the number of transitions between sites.

For a given excursion $\mathbf{e}$, comprising sites $0 = \mathbf{e}_0, \mathbf{e}_1, \ldots, \mathbf{e}_{|\mathbf{e}|-1}, \mathbf{e}_{|\mathbf{e}|}, \mathbf{e}_{|\mathbf{e}|+1} = 0$ we define

$$\Xi(\mathbf{e}) := \left\{ 0 \leq t \leq |\mathbf{e}| + 1 \,:\, \mathbf{e}_t \neq \mathbf{e}_{t+1} \right\}$$

Interpretation: $\Xi(\mathbf{e})$ denotes the set of change points in an excursion i.e, it takes into account of all those times where a transition is made from one site to another and therefore $e_t \neq e_{t+1}$. Notice that $t_1 \in \Xi(\mathbf{e})$ since $e_{t_1} = 0 \neq e_{t_1+1}$. Using the same idea, note that $t_1 + |\hat{\mathbf{e}}| \in \Xi(\mathbf{e})$ since $e_{t_1+|\hat{\mathbf{e}}|} \neq e_{t_1+|\hat{\mathbf{e}}|+1} = 0 = e_{t_2}$. Trivially, 0 and $|\hat{\mathbf{e}}|$ are always in $\Xi(\mathbf{e})$. The total number of transitions made from one site to the next, i.e, number of $\Xi(\mathbf{e})$ is denoted by m.

$$\kappa(\mathbf{e}) := \max\left\{ \kappa_j \,:\, j \in \mathbf{e} \right\};$$

Interpretation: Here κ_j=min $\{$no. of edges of $P|P$ is a closed path passing through 0 and $j\}$.

That is, κ_j is the length of the smallest cycle that passes through vertex j and starts and ends at 0. Please note that this length is the number of edges in the closed path and not the time-length. Now $\kappa_j \leq m$ because the number of transitions for the smallest cycle can not be strictly greater than the possible transitions between those two sites. Therefore, $\kappa_j \leq m \; \forall \; j$. This implies $\kappa(\mathbf{e}) \leq \#\Xi(\mathbf{e}) = m$ since $\kappa(\mathbf{e})$ is the maximum of all κ_j. The other way to prove this is: $\kappa_j \leq \text{length}(\hat{e})$. But $\text{length}(\hat{e}) = m$ (proof trivial but messy), so $\kappa_j \leq m$ and $\kappa(\mathbf{e}) \leq m$. We call $\kappa(\mathbf{e})$ the diameter of $\hat{e}$.

$$\rho(\mathbf{e}) := \min\left\{\rho_j \,:\, j \in \mathbf{e}\right\}.$$

Interpretation: This is the loss in expected log growth at the site that might be thought of as the "target" of the excursion — the site where the loss is the smallest, scaled by the variance, which turns out to be the appropriate combination for measuring the influence of the fluctuations at the non-optimal site j, given that an excursion reaches that site.

Note that 0 and $|\mathbf{e}|$ are always in $\Xi(\mathbf{e})$, and the definition of κ_j implies that $\kappa(\mathbf{e}) \leq \#\Xi(\mathbf{e})$. We will refer to $\kappa(\mathbf{e})$ as the *diameter* of $\mathbf{e}$.

We write $\hat{\mathscr{E}}_T$ for the collection of sequences of excursions that can be fit into time $\{1, \ldots, T\}$. That is, an element $\hat{\mathbf{e}} \in \hat{\mathscr{E}}_T$ has an *excursion count* $k(\hat{\mathbf{e}})$, such that each $i \in \{1, \ldots, k(\hat{\mathbf{e}})\}$ there is a pair $(t_i, \hat{\mathbf{e}}_i)$ with $t_i \in \{2, \ldots, T-1\}$ and $\hat{\mathbf{e}}_i \in \mathscr{E}$ satisfying

$$t_i + |\hat{\mathbf{e}}_i| < t_{i+1},$$
$$t_{k(\hat{\mathbf{e}})} + |\hat{\mathbf{e}}_{k(\hat{\mathbf{e}})}| \leq T.$$

We write the total length of an excursion sequence as

$$\|\hat{\mathbf{e}}\| := \sum_{i=1}^{k(\hat{\mathbf{e}})} |\hat{\mathbf{e}}_i|.$$

We also write $\hat{\mathscr{E}}_{T;k,n,m}$ for the subset of $\hat{\mathscr{E}}_T$ comprising excursion sequences whose excursion

count is k, whose total length is n, and the sum of whose change-point counts $\#\Xi(\hat{\mathbf{e}}_i)$ is m. That is, $m = \#\Xi(\hat{\mathbf{e}})$, where $\Xi(\hat{\mathbf{e}}) := \bigcup_{i=1}^{k} \{t_i + t : t \in \Xi(\hat{\mathbf{e}}_i)\}$. The *null excursion sequence* is the element of $\hat{\mathscr{E}}_T$ with $k(\hat{\mathbf{e}}) = \|\hat{\mathbf{e}}\| = 0$.

Lemma 9.

$$\log \binom{n}{j} \leq j + j \log \frac{n}{j}, \tag{12}$$

which holds for all positive integers n and $0 \leq j \leq n$*, following the convention* $0 \cdot \log 0 = 0 \cdot \log \infty = 0$

Proof. Since $j \geq 0$,

$$e^j = \sum_m \frac{j^m}{m!} > \frac{j^j}{j!}, \text{ so } (1/j!) < \frac{e^j}{j^j}.$$

Next,

$$\binom{n}{j} = \frac{n(n-1)\dots(n-j+1)}{j!}, \tag{13}$$

$$< \frac{n^j e^j}{j^j}, \tag{14}$$

$$\log \binom{n}{j} < j + j \log n - j \log j, \tag{15}$$

$$< j + j \log \frac{n}{j}. \tag{16}$$

□

Lemma 10. *For 2 sites,* $d = 2$*, each excursion has exactly 2 changes in site (0 to 1 and back), so the diameter of each excursion (and hence the average diameter of excursions) is exactly* $\beta = 2$*. An excursion sequence in a time* T *is specified only by the values* k, n*. For each excursion in such a sequence, we have 2 changes in site and so the total number of changes is* $m = 2k$*. From the bounds on* $\binom{T}{k}$ *and* $\binom{n}{k}$*, we have*

$$\log \#\hat{\mathscr{E}}_{T;k,n} \leq 2k + k \log \frac{n}{k} + k \log \frac{T}{k} \leq 2k + k \log \frac{n}{k} + T \log \frac{T}{k} \tag{17}$$

Proof. An element of $\hat{\mathscr{E}}_{T;k,n}$ may be determined by the following choices:

a. Choose k points out of T where the excursions begin, yielding no more than $\binom{T}{k}$ possibilities;

b. Choose k numbers for the lengths of the excursions that add up to n, yielding no more than $\binom{n}{k}$ possibilities;

The first inequality follows from combining these pieces with the crude bound in (12). □

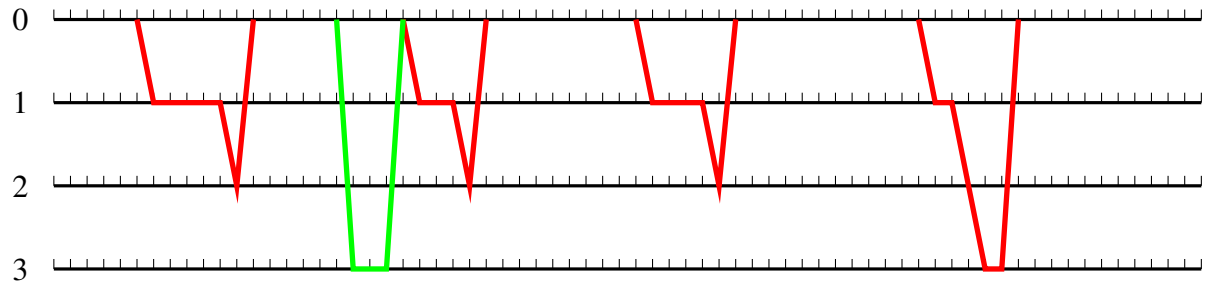


Figure 13: An excursion sequence for $T = 70$ comprising $k = 5$ excursions. This is based on the migration graph example from Figure 12. Three excursions (red) have diameter 3, and one (green) has diameter 2. Note that one timepoint ($t = 21$) is included in two different excursions. The red excursions all have $\rho(\hat{\mathbf{e}}) = 0.2$, and the green excursion has $\rho(\hat{\mathbf{e}}) = 0.35$. The lengths are 6, 3, 4, 5, 5, giving the sequence a total length $n = 23$. The change-point counts are 3, 2, 3, 3, 4, summing to $m = 15$.

The $(0,0)$ entry of the product R_T will be a sum of terms that are enumerated by elements of $\hat{\mathscr{E}}_T$, corresponding to paths through the sites. We define new random variables as a function of the realizations of $\mathbf{X}$ and of $\mathbf{A}$ (the collection of all matrices A)

$$\alpha_t(i,j) := \widetilde{X}_t^{(j)} + \log \frac{1+\varepsilon A_t(j,j)}{1+\varepsilon A_t(0,0)} + \begin{cases} \log \varepsilon + \log A_t(j,i) & \text{if } i \neq j, \\ 0 & \text{if } i = j. \end{cases} \tag{18}$$

Given an excursion $\mathbf{e}$ and a starting time $t_0 \in \{0, \ldots, T - |\mathbf{e}| - 1\}$ we define the random variables

$$\mathbf{e}[t_0; \mathbf{X}, \mathbf{A}] := \sum_{0 \le t \le |\mathbf{e}|} \alpha_{t+t_0}(\mathbf{e}_t, \mathbf{e}_{t+1}) \tag{19}$$

Of course, this sum may be $-\infty$, if it includes a transition at which the corresponding entry of A is 0. But our assumptions imply that it is finite with nonzero probability if $\mathbf{e} \in \mathscr{E}$. Given an

excursion sequence $\hat{\mathbf{e}} = \Big((t_i, \hat{\mathbf{e}}_i)\Big)_{i=1}^{k} \in \hat{\mathscr{E}}_T$, we define

$$\hat{\mathbf{e}}[\mathbf{X}, \mathbf{A}] := \sum_{i=1}^{k} \hat{\mathbf{e}}_i[t_i; \mathbf{X}, \mathbf{A}]. \tag{20}$$

The quantity we are trying to approximate is

$$a(\varepsilon) - a(0) = \lim_{T\to\infty} T^{-1}\left(\log R_T(\varepsilon)_{0,0} - \sum_{t=1}^{T} X_t^{(0)}\right). \tag{21}$$

Lemma 11.

$$\log R_T(0,0) = \sum_{t=1}^{T} X_t^{(0)} + \varepsilon \sum_{t=1}^{T} \log\big(1 + \varepsilon A_t(0,0)\big) + \log\left(1 + \sum_{\hat{\mathbf{e}} \in \hat{\mathscr{E}}_T \setminus \{\hat{\mathbf{e}}^0\}} \mathrm{e}^{\hat{\mathbf{e}}[\mathbf{X},\mathbf{A}]}\right), \tag{22}$$

where $\hat{\mathbf{e}}^0$ is the null excursion sequence.

Proof. We have, by definition,

$$R_T(0,0) = \sum_{(x_0,\ldots,x_T)} \prod_{t=1}^{T} D_t(\varepsilon)\,(x_t, x_{t-1})$$

where the summation is over $(x_0, \ldots, x_T) \in \{0, \ldots, d-1\}^{T+1}$ with $x_0 = x_T = 0$. We may restrict the summation to $(T+1)$ -tuples such that $D_t(\varepsilon)\,(x_t, x_{t-1}) > 0$, which will only be true when (x_{t-1}, x_t) is an edge of $\mathscr{M}$. Such sequences of states map one-to-one onto excursion sequences. By the definition of matrix products,

$$R_T(0,0) = \sum_{k=0}^{T} \sum_{\hat{e} \in \hat{\mathscr{E}}_{T,k}} \prod_{i=1}^{k+1} \left(\prod_{t=t_{i-1}+|\hat{e}_{i-1}|+1}^{t_i} D_t(\varepsilon)(0,0) \right) \cdot \prod_{i=1}^{k} \left(\prod_{t=1}^{|\hat{e}_i|} D_{t+t_i}(\varepsilon)\left((\hat{\mathbf{e}}_i)_{t+1}, (\hat{\mathbf{e}}_i)_t\right) \right), \tag{23}$$

where $t_0 = 0$ and $t_{k+1} = T$. The logarithm of the summand in the above expression is

$$\sum_{i=1}^{k+1} \sum_{t=t_{i-1}+|e_{i-1}|+1}^{t_i} \log D_t(\varepsilon)(0,0) + \sum_{i=1}^{k} \sum_{t=1}^{|e_i|} \log \frac{D_{t+t_i}(\varepsilon)\left((\hat{\mathbf{e}}_i)_{t+1}, (\hat{\mathbf{e}}_i)_t\right) D_t(\varepsilon)(0,0)}{D_t(\varepsilon)(0,0)}$$
$$= \sum_{t=0}^{T} D_t(\varepsilon)(0,0) + \sum_{i=1}^{k} \sum_{t=1}^{|e_i|} \log \frac{D_{t+t_i}(\varepsilon)\left((\hat{\mathbf{e}}_i)_{t+1}, (\hat{\mathbf{e}}_i)_t\right)}{D_t(\varepsilon)(0,0)}.$$

We have $D_t(\varepsilon)(0,0) = \mathrm{e}^{X_t^{(0)}}(1+\varepsilon A_t(0,0))$. Taking the logarithm of the expression, we get for the expression above

$$\sum_{t=1}^{T} \left(X_t^{(0)} + \log(1+\varepsilon A_t(0,0))\right) - \sum_{i=1}^{k} \sum_{t=1}^{|\hat{\mathrm{e}}_i|} \log \frac{D_{t+t_i}(\varepsilon)\left((\hat{\mathbf{e}}_i)_t, (\hat{\mathbf{e}}_i)_{t+1}\right)}{D_{t+t_i}(\varepsilon)(0,0)}$$
$$= \sum_{t=0}^{T} \left(X_t^{(0)} + \log\Big(1+\varepsilon A_t(0,0)\Big)\right) + \sum_{i=1}^{k} \sum_{t=1}^{|\hat{\mathrm{e}}_i|} \alpha_{t+t_i}(\varepsilon)\left((\hat{\mathbf{e}}_i)_t, (\hat{\mathbf{e}}_i)_{t+1}\right),$$

since

$$\log \frac{D_t(\varepsilon)(i,j)}{D_t(\varepsilon)(0,0)} = \alpha_t(i,j) := \widetilde{X}_t^{(\mathbf{e}_t)} + \log \frac{1+\varepsilon A_t(\mathbf{e}_t,\mathbf{e}_t)}{1+\varepsilon A_t(0,0)} + \begin{cases} \log\varepsilon + \log A_t(j,i) & \text{if } i \neq j, \\ 0 & \text{if } i = j. \end{cases}$$

The result then follows from splitting off the null excursion and invoking the definition of random variable $\hat{\mathbf{e}}[X,A]$. □

We have then

$$
\begin{aligned}
&\log R_T(0,0) - \sum_{t=1}^{T} \log\big(1+\varepsilon A_t(0,0)\big) - \sum_{t=1}^{T} X_t^{(0)} \\
&\qquad = \log\left(1 + \sum_{\hat{\mathrm{e}}\in\hat{\mathscr{E}}_T\setminus\{\hat{\mathrm{e}}^0\}} \mathrm{e}^{\hat{\mathrm{e}}[\mathrm{X},\mathrm{A}]}\right) \\
&\qquad = \log\left(\sum_{\hat{\mathrm{e}}\in\hat{\mathscr{E}}_T} \mathrm{e}^{\hat{\mathrm{e}}[\mathrm{X},\mathrm{A}]}\right) \\
&\qquad = \log\left(\sum_{1\le k,n,m\le T} \sum_{\hat{\mathrm{e}}\in\hat{\mathscr{E}}_{T;k,n,m}} \mathrm{e}^{\hat{\mathrm{e}}[\mathrm{X},\mathrm{A}]}\right) \\
&\qquad \le \log\left(\sum_{1\le k,n,m\le T} \#\hat{\mathscr{E}}_{T;k,n,m} \mathrm{e}^{\max_{\hat{\mathrm{e}}\in\hat{\mathscr{E}}_T} \hat{\mathrm{e}}[\mathrm{X},\mathrm{A}]}\right) \\
&\qquad \le \log T^3 \max_{1\le k,n,m\le T} \#\hat{\mathscr{E}}_{T;k,n,m} \mathrm{e}^{\max_{\hat{\mathrm{e}}\in\hat{\mathscr{E}}_T} \hat{\mathrm{e}}[\mathrm{X},\mathrm{A}]} \\
&\qquad = \log T^3 + \max_{1\le k,n,m\le T} \log\left(\#\hat{\mathscr{E}}_{T;k,n,m} \mathrm{e}^{\max_{\hat{\mathrm{e}}\in\hat{\mathscr{E}}_T} \hat{\mathrm{e}}[\mathrm{X},\mathrm{A}]}\right) \\
&\qquad = 3\log T + \max_{1\le k,n,m\le T} \left(\log(\#\hat{\mathscr{E}}_{T;k,n,m}) + \max_{\hat{\mathrm{e}}\in\hat{\mathscr{E}}_T} \hat{\mathrm{e}}[\mathrm{X},\mathrm{A}]\right).
\end{aligned}
\tag{24}
$$

Thus we have the bounds

$$
a(\varepsilon) - a(0) \ge \liminf_{T\to\infty} T^{-1} \max_{\hat{\mathbf{e}}\in\hat{\mathscr{E}}_T} \hat{\mathbf{e}}[\mathbf{X},\mathbf{A}] - \varepsilon\Delta^0 + O(\varepsilon^2), \tag{25}
$$

and

$$
\begin{aligned}
&a(\varepsilon) - a(0) \\
&\quad \le \limsup_{T\to\infty} T^{-1} \max_{1\le k,n,m\le T} \left(\log \#\hat{\mathscr{E}}_{T;k,n,m} + \max_{\hat{\mathbf{e}}\in\hat{\mathscr{E}}_{T;k,n,m}} \hat{\mathbf{e}}[\mathbf{X},\mathbf{A}]\right) - \varepsilon\Delta^0 + O(\varepsilon^2).
\end{aligned}
\tag{26}
$$

Derivation of the upper bound

We prove the upper bound in (8). Replace $A_t(j,i)$ by $A_t(j,i) \vee 1$ for any $i \to j$, so we can take

$$\mu_A = 0.$$

We also begin by assuming $A_t(i,i) \equiv 0$. (We will remove this restriction at the end.)

Here we have only 2 sites, call them 0 and 1. We refer to $\tilde{\mu}_1, \tau_1, \kappa_1, \rho_1$ simply as $\tilde{\mu}, \tau, \kappa, \rho$. Both κ_1, ρ_1 are the only values and so are the minimum values. Recall that here $\kappa = 2$ is the minimum number of changes in each excursion.

Again, for 2 sites, define

$$\zeta := \frac{k}{T}, \quad \nu := \frac{n}{k}, \quad \beta := \frac{m}{k}.$$

Here $\beta = 2 = \kappa$ is the diameter of every excursion and is also the average diameter of excursion. So we can equivalently write

$$k = \zeta T, n = \nu k = \nu \zeta T, m = 2k = 2\zeta T. \tag{27}$$

Here note that ζ is the rate of excursions per unit time; ν is the average length of excursions. Note that $\nu \geq 2$ since each excursion must have length at least 2. Also, $T \geq n$, yielding the constraints $1/\zeta \geq \nu \geq \beta = \kappa = 2$.

For 2 states, the essential computation is modified as below from the general case.

Lemma 12. *Setting $c = (2+\delta), \eta = 4\rho/(1+2\rho)$ for any $\delta > 0$ and*

$$z = \varepsilon^{\eta} \cdot (\log \varepsilon^{-1})^c,$$

we have

$$\limsup_{T\to\infty} T^{-1} \log \mathbb{P}\Big\{ \max_{\hat{\mathbf{e}} \in \hat{\mathscr{E}}_{T;k,n}} \Big(\hat{\mathbf{e}}[\mathbf{X}, \mathbf{A}] + \log \# \hat{\mathscr{E}}_{T;k,n} \Big) \geq zT \Big\} < 0 \tag{28}$$

for all $\varepsilon > 0$ sufficiently small.

Proof. Use (27) to see that the bound (17) may be written as

$$\log \#\hat{\mathscr{E}}_{T;k,n} \leq \zeta T \left(\log \nu - \log \zeta + 2\right). \tag{29}$$

Suppose now we fix some element $\hat{\mathbf{e}}$ of $\hat{\mathscr{E}}_{T;k,n}$, and list all the states of all the excursions in order as $j_1, \ldots, j_n$, and the associated times as $t_1 < t_2 < \cdots < t_n$. Remember that here $j_i = 1$ for every i. Therefore we have

$$\mathbb{E}\Big[\hat{\mathbf{e}}[\mathbf{X},\mathbf{A}]\Big] \leq 2k \log \varepsilon - \sum_{i=1}^{n} \tilde{\mu} = k\Big(2\mu_A + 2\log \varepsilon - \nu\tilde{\mu}\Big).$$

and the random variable $Y := \hat{\mathbf{e}}[\mathbf{X},\mathbf{A}] - \mathbb{E}[\hat{\mathbf{e}}[\mathbf{X},\mathbf{A}]]$ is sub-Gaussian with variance factor bounded by

$$\sum_{i=1}^{n} \tau_{j_i} + m\tau_A = n\tau + 2k\tau_A = k(\nu\tau + 2\tau_A).$$

Recall that the constant τ_A has been defined in (7) so that the variance factor of $\widetilde{X}_t^{(i)} + \log A_t(\ell, i)$ for any $i \neq 1$ will be less than or equal to τ_A/ρ. By (4), for any $x, z > 0$,

$$\begin{aligned}\mathbb{P}\Big\{\Big(\hat{\mathbf{e}}[\mathbf{X},\mathbf{A}] + x\Big) \geq zT\Big\} &\leq \mathbb{P}\Big\{Y \geq \sum_{i=1}^{n} \tilde{\mu}_{j_i} - m\mu_A + m\log \varepsilon^{-1} + zT - x\Big\} \\ &\leq \exp\Bigg\{-\frac{1}{2}\left(\sum_{i=1}^{n} \tau_{j_i} + m\frac{\tau_A}{\rho_1}\right)^{-1}\left(\sum_{i=1}^{n} \tilde{\mu}_{j_i} - m\mu_A + m\log \varepsilon^{-1} + zT - x\right)^2\Bigg\}.\end{aligned} \tag{30}$$

For the 2 sites, first define

$$S = \nu\tilde{\mu} + 2\tau_A, \text{ so } \nu\tilde{\mu} = (S - 2\tau_A). \tag{31}$$

Next, recalling that $\rho = (\tilde{\mu}/\tau)$,

$$
\begin{aligned}
(\sum_{i=1}^{n} \tau_{j_i} + m\frac{\tau_A}{\rho}) &= k(\nu\tau + 2\frac{\tau_A}{\rho}), \\
&= \frac{\zeta}{\rho} T(\nu\tilde{\mu} + 2\tau_A) \\
&= \frac{\zeta T}{\rho} S
\end{aligned}
$$

Also

$$
\begin{aligned}
\sum_{i=1}^{n} \tilde{\mu}_{j_i} + m\log\varepsilon^{-1} + zT - x &= n\tilde{\mu} + 2k\log\varepsilon^{-1} + zT - x, \\
&= \zeta T\left(\nu\tilde{\mu} + 2\log\varepsilon^{-1} + (z/\zeta) + \Big(x/(\zeta T)\Big)\right) \\
&= \zeta T\left(S - 2\tau_A + 2\log\varepsilon^{-1} + (z/\zeta) + \Big(x/(\zeta T)\Big)\right).
\end{aligned}
$$

Hence for our 2 sites we can use (30) to get

$$
\log\mathbb{P}\Big\{\Big(\hat{\mathbf{e}}[\mathbf{X},\mathbf{A}] + x\Big) \geq zT\Big\} \leq -\frac{\rho}{2} T B_1, \tag{32}
$$

where

$$
B_1 := \zeta S^{-1} \cdot \left(S - 2\tau_A + 2\log\varepsilon^{-1} + (z/\zeta) + \Big(x/(\zeta T)\Big)\right)^2, \tag{33}
$$

Taking $x = \log\#\hat{\mathscr{E}}_{T;k,n}$ and substituting (29), we get

$$
\begin{aligned}
\log\mathbb{P}\Big\{\max_{\hat{\mathbf{e}}\in\hat{\mathscr{E}}_{T;k,n}} \Big(\hat{\mathbf{e}}[\mathbf{X},\mathbf{A}] + \log\#\hat{\mathscr{E}}_{T;k,n}\Big) \geq zT\Big\} \\
&\leq \log\#\hat{\mathscr{E}}_{T;k,n} + \max_{\hat{\mathbf{e}}\in\hat{\mathscr{E}}_{T;k,n}} \log\mathbb{P}\Big\{\Big(\hat{\mathbf{e}}[\mathbf{X},\mathbf{A}] + \log\#\hat{\mathscr{E}}_{T;k,n}\Big) \geq zT\Big\} \\
&\leq -\frac{\rho}{2}\cdot T \inf_{S\geq 2(\tau_A+\tilde{\mu}_A)} \inf_{u\geq z} \frac{z}{u}\Theta(S,\beta,u)
\end{aligned}
$$

where $u = z/\zeta$.

Writing (29) and (31) to write

$$
\begin{aligned}
B_2 &:= (x/\zeta T) \\
&= (\log \nu - \log \zeta + 2), \\
&= (\log(S - 2\tau_A) - \log \tilde{\mu} - \log z + \log u + 2), \text{ and} \\
B_3 &:= \left(S - 2\tau_A + 2\log \varepsilon^{-1} + (z/\zeta) + \Big(x/(\zeta T) \Big) \right), \\
&= \Big(S - 2\tau_A + 2\log \varepsilon^{-1} + u + \log(S - 2\tau_A) - \log \tilde{\mu} - \log z + \log u + 2 \Big),
\end{aligned}
$$

we then have

$$\Theta = \Theta(S, \beta, u) := S^{-1} B_2^2 + B_3.$$

Here we have taken advantage of the fact that $\log \zeta < 0$.

We will now show that there is a positive constant ε_0 (expressible in terms only of $\delta, \rho_1, \kappa_1, \tilde{\mu}_*, d, \tau_A, \mu_A$) such that for any fixed $\varepsilon \in (0, \varepsilon_0)$,

$$z = \varepsilon^{2\kappa_1 \rho_1/(1+2\rho_1)} (\log \varepsilon^{-1})^{\kappa_1 \vee 3 + \delta},$$

and any $S \geq \kappa_1(\tau_A + \tilde{\mu}_*)$, $\beta \geq \kappa_1$, and $u \geq 0$,

$$\Theta(S, \beta, u) \geq \frac{2}{\rho_1} + u.$$

The result then follows immediately, since then for all T,

$$T^{-1} \log \mathbb{P}\Big\{ \max_{\hat{\mathbf{e}} \in \hat{\mathscr{E}}_{T;k,n,m}} \Big(\hat{\mathbf{e}}[\mathbf{X}, \mathbf{A}] + \log \# \hat{\mathscr{E}}_{T;k,n,m} \Big) \Big\} \geq -z.$$

We write

$$\Theta = S^{-1} (y_1 + S)^2 + \frac{2}{\rho_1} (y_{\tau_A/\log \varepsilon^{-1}} - u),$$

where

$$y_r := r\beta \log \varepsilon^{-1} - \beta\tau_A + \log z + (\beta-2)\log(\beta-2) - \beta \log d(S-\beta\tau_A) - \log u + u$$

for $r \in [0,1]$. Note that y is convex in β, with

$$\frac{\partial y_r}{\partial \beta} = r\log\varepsilon^{-1} - \tau_A + \log(\beta-2) - \log d(S-\beta\tau_A) + \frac{S}{S-\beta\tau_A}.$$

If $S \geq 3\tau_A + \frac{e^{1-\tau_A}}{d}\varepsilon^{-r}$ then $\partial y_r/\partial\beta > 0$ for $\beta = 3$. If $\kappa_1 \geq 3$ this implies that the minimum value of y_r on the allowable range $\beta \geq \kappa_1$ is at κ_1, so

$$y_r \geq \kappa_1\left(r - \frac{2\rho_1}{2\rho_1+1}\right)\log\varepsilon^{-1} - \kappa_1\tau_A - \kappa_1 \log dS - \log u + u + (\kappa_1 \vee 3 + \delta)\log\log\varepsilon^{-1}.$$

Minimizing each term individually on the interval $\beta \in [2,3]$ for the case $\kappa_1 = 2$ we then have for all $\beta \in [\kappa_1, \infty)$

$$y_r \geq \kappa_1\left(r - \frac{2\rho_1}{2\rho_1+1}\right)\log\varepsilon^{-1} - (\kappa_1 \vee 3)\Big(\tau_A + \log dS\Big) - \log u + u + (\kappa_1 \vee 3 + \delta)\log\log\varepsilon^{-1}. \quad (34)$$

We consider separately the large and small values of S. If $S \geq \log\varepsilon^{-1}$ we write

$$\Theta \geq S + \frac{2\rho_1+2}{\rho_1}y_r - \frac{2}{\rho_1}u \quad \text{with} \quad r = \frac{\rho_1}{\rho_1+1}.$$

Applying (34) we see that

$$\Theta \geq S - \frac{2\rho_1+2}{\rho_1}(\kappa_1 \vee 3)\log S \quad - \frac{\rho_1}{2\rho_1^2+3\rho_1+1}\log\varepsilon^{-1} + u - C,$$
$$\text{where} \quad C := \frac{2\rho_1+2}{\rho_1}(\kappa_1 \vee 3)\,(\tau_A + \log d) - \frac{2\rho_1+2}{\rho_1}\log\frac{2\rho_1+2}{\rho_1}.$$

For ε sufficiently small we have $\log \varepsilon^{-1} \geq C+$, and for all $S \geq 2\log \varepsilon^{-1}$

$$S - \frac{2\rho_1 + 2}{\rho_1}(\kappa_1 \vee 3)\log S \geq \frac{4}{3}\log \varepsilon^{-1},$$

hence $\Theta \geq 1 + u$.

For $S < 2\log \varepsilon^{-1}$ we use $(S+y)^2 \geq 4Sy$ to obtain

$$\Theta \geq 4y_1 + \frac{2}{\rho_1}(y_{\tau_A/\log \varepsilon^{-1}} - u) \geq \frac{4\rho_1 + 2}{\rho_1} y_r - \frac{2}{\rho_1} u, \quad \text{where } r = \frac{2\rho_1}{2\rho_1 + 1}. \tag{35}$$

Applying (34) we have

$$\begin{aligned}\Theta \geq \frac{4\rho_1 + 2}{\rho_1}\Big[(\kappa_1 \vee 3 + \delta)\log(2\log \varepsilon^{-1}) - (\kappa_1 \vee 3)\log S\Big] + 2u - C', \\ \text{where} \quad C' := 2C + \frac{4\rho_1 + 2}{\rho_1}(\kappa_1 \vee 3 + \delta)\log 2 \\ \geq 4\delta \log\log \varepsilon^{-1} + 2u - C'.\end{aligned}$$

Thus $\Theta \geq 1 + u$ for ε sufficiently small, completing the proof. □

Lemma 12 implies that

$$\sum_{T=T_0}^{\infty} \mathbb{P}\left\{ T^{-1} \max_{1 \leq k,n,m \leq T} \left(\log \#\hat{\mathscr{E}}_{T;k,n,m} + \max_{\hat{\mathbf{e}} \in \hat{\mathscr{E}}_{T;k,n,m}} \hat{\mathbf{e}}[\mathbf{X}, \mathbf{A}] \right) \geq z \right\} < \infty.$$

By the Borel–Cantelli Lemma, this implies that with probability 1 this event occurs only finitely often. It follows that the limsup is no greater than z almost surely, and hence, by (26), that

$$a(\varepsilon) - a(0) \leq \varepsilon^{\bar{\eta}} \cdot (\log \varepsilon^{-1})^c. \tag{36}$$

It remains only to clear away the assumption that that κ_j and ρ_j are both minimized at site 1. We do this by stratifying the excursions further by their diameter (recall the definition from section

5). Define

$$\breve{\rho}(\kappa) := \min\Big\{\rho_j \,:\, \kappa_j \le \kappa\Big\}.$$

If $\mathbf{e}$ is an excursion with diameter κ, then any site j included in $\mathbf{e}$ has $\kappa_j \le \kappa$, hence also $\rho_j \ge \breve{\rho}(\kappa)$. Furthermore,

$$\min_{1\le j\le d-1} 2\rho_j\kappa_j/(1+2\rho_j) = \min_{2\le\kappa\le d-1} 2\breve{\rho}(\kappa)\kappa/(1+2\breve{\rho}(\kappa)) \tag{37}$$

The maximum in (26) may be written as a maximum over $(k_2,\ldots,k_{d-1})$, representing the number of excursions whose diameter is $2,3,\ldots,d-1$, with the constraint $\sum k_\kappa = k$. We write $\hat{\mathscr{E}}^{(\kappa)}_{T;k,n,m}$ for the excursion sequences consisting of k excursions, all of which have diameter κ; and $\hat{\mathscr{E}}_{T;(k_\kappa),n,m}$ for the set of excursion sequences that have exactly k_κ excursions with diameter κ. Then $\hat{\mathscr{E}}_{T;(k_\kappa),n,m}$ is naturally included in the direct sum of $\hat{\mathscr{E}}^{(\kappa)}_{T;k_\kappa,n,m}$. (A sequence of mixed diameters $\hat{\mathbf{e}} \in \hat{\mathscr{E}}_{T;(k_\kappa),n,m}$ may be decomposed into sequences $\hat{\mathbf{e}}_{(\kappa)} \in \hat{\mathscr{E}}^{(\kappa)}_{T;k_\kappa,n,m}$ of excursions with each particular diameter. Referring back to the example in Figure 13, this would entail making one excursion sequence by dropping out the green excursions, and a separate one by dropping out the red excursions.) Thus

$$\begin{aligned}
\max_{\hat{\mathbf{e}}\in\hat{\mathscr{E}}_{T;k,n,m}} \hat{\mathbf{e}}[\mathbf{X},\mathbf{A}] &= \max_{\sum k_\kappa = k} \max_{\hat{\mathbf{e}}\in\hat{\mathscr{E}}_{T;(k_\kappa),n,m}} \hat{\mathbf{e}}[\mathbf{X},\mathbf{A}] \\
&\le \max_{\sum k_\kappa = k} \max_{\hat{\mathbf{e}}_{(\kappa)}\in\hat{\mathscr{E}}^{(\kappa)}_{T;k_\kappa,n,m}} \sum_\kappa \hat{\mathbf{e}}_{(\kappa)}[\mathbf{X},\mathbf{A}] \\
&\le \sum_\kappa \max_{1\le k_\kappa\le T} \max_{\hat{\mathbf{e}}_{(\kappa)}\in\hat{\mathscr{E}}^{(\kappa)}_{T;k_\kappa,n,m}} \hat{\mathbf{e}}_{(\kappa)}[\mathbf{X},\mathbf{A}],
\end{aligned}$$

using the general fact that the maximum of a sum is smaller than the sum of maxima. Thus we have

$$\max_{1\le k,n,m\le T}\left(\log\#\hat{\mathscr{E}}_{T;k,n,m} + \max_{\hat{\mathbf{e}}\in\hat{\mathscr{E}}_{T;k,n,m}} \hat{\mathbf{e}}[\mathbf{X},\mathbf{A}]\right) \le \sum_\kappa \max_{1\le k_\kappa,n_\kappa,m_\kappa\le T} \max_{\hat{\mathbf{e}}_{(\kappa)}\in\hat{\mathscr{E}}^{(\kappa)}_{T;k_\kappa,n_\kappa,m_\kappa}} \hat{\mathbf{e}}_{(\kappa)}[\mathbf{X},\mathbf{A}].$$

Here n_κ and m_κ simply index the choice of total length and number of change-points among all

of the excursions of length κ.

Because all excursions in $\hat{\mathscr{E}}^{(\kappa)}_{T;k_\kappa,n_\kappa,m_\kappa}$ pass through only sites j with $\rho_j \geq \check{\rho}(\kappa)$, the same argument used for the upper bound in (36) may be applied to show that almost surely for $\varepsilon > 0$ sufficiently small

$$\limsup_{T\to\infty} T^{-1} \max_{1\leq k_\kappa,n_\kappa,m_\kappa\leq T} \max_{\hat{\mathbf{e}}_{(\kappa)}\in\hat{\mathscr{E}}^{(\kappa)}_{T;k_\kappa,n_\kappa,m_\kappa}} \hat{\mathbf{e}}_{(\kappa)}[\mathbf{X},\mathbf{A}] \leq c'_\kappa(\log\varepsilon^{-1})^{c_\kappa}\varepsilon^{2\kappa\check{\rho}(\kappa)/(1+2\check{\rho}(\kappa))}.$$

It follows that for $c' := (2d-2)\cdot\max c'_\kappa$ and $c := \max c_\kappa$,

$$\begin{aligned}\limsup_{T\to\infty} T^{-1} \max_{1\leq k,n,m\leq T}\left(\log\#\hat{\mathscr{E}}_{T;k,n,m} + \max_{\hat{\mathbf{e}}\in\hat{\mathscr{E}}_{T;k,n,m}} \hat{\mathbf{e}}[\mathbf{X},\mathbf{A}]\right)& \\ \leq \sum_\kappa c'_\kappa(\log\varepsilon^{-1})^{c_\kappa}\varepsilon^{2\kappa\check{\rho}(\kappa)/(1+2\check{\rho}(\kappa))}& \\ \leq (2d-2)(\log\varepsilon^{-1})^{c}\varepsilon^{\min_{1\leq j\leq d-1}2\kappa_j\rho_j/(1+2\rho_j)}&\end{aligned}$$

by (37), which completes the proof for the case $A_t(i,i)\equiv 0$.

When $A_t(0,0)\geq A_t(j,j)$ the contribution $-\varepsilon\Delta^0$ from (26) becomes relevant, but the additional term

$$\limsup_{T\to\infty} T^{-1} \max_{1\leq k,n,m\leq T}\left(\log\#\hat{\mathscr{E}}_{T;k,n,m} + \max_{\hat{\mathbf{e}}\in\hat{\mathscr{E}}_{T;k,n,m}} \hat{\mathbf{e}}[\mathbf{X},\mathbf{A}]\right),$$

can only decrease. This proves (9).

Changing from $A(i,i)\equiv 0$ to general A is equivalent to replacing $\widetilde{X}^{(i)}_t$ in the above calculation by $\widetilde{X}'^{(i)}_t = \widetilde{X}^{(i)}_t + \varepsilon(A_t(i,i)-A_t(0,0))$. (There is, in addition, a bounded multiple of ε^2, that will have no effect on the limit.) The only effect is to replace $\tilde{\mu}_i$ by $\tilde{\mu}'_i = \tilde{\mu}_i + \varepsilon(\Delta^0-\Delta^i)$, and τ_i by $(\sqrt{\tau_i}+\varepsilon\sqrt{\tau_\Delta})^2$ (by Lemma 5), meaning that ρ_j is replaced by $\rho'_j \geq \rho_j - c'\varepsilon$ where

$$c' = \max_j\left(\mathbb{E}\left[\Big(A_t(j,j)-A_t(0,0)\Big)_+\right] + 3\max\left\{1,\frac{\tau_\Delta}{\tau_j}\right\}\rho_j\right).$$

Thus we may apply the earlier result with $\bar{\eta}$ replaced by a power arbitrarily close before send-

ing ε to 0. (Note that the earlier calculations assumed the power to be fixed as ε goes to 0, so we cannot directly infer that $a(\varepsilon) - a(0) \leq \varepsilon^{\bar{\eta}-c\varepsilon}(\log \varepsilon^{-1})^c$.)

Derivation of the lower bound

We show that the lower bound applies for each j; it will then hold in particular for the j at which $\kappa_j \rho_j$ attains its minimum. We may assume without loss of generality that this best site is $j = 1$, and we will write simply κ, $\tilde{\mu}$, and $\underline{\rho}$ for κ_1, $\tilde{\mu}_1$, and $\underline{\rho}_1$. We begin with the assumption that the penalty $A_t(j,j)$ is identically 0.

Let $0 = j_0, j_1, j_2, \ldots, j_I = 1, j_{I+1}, \ldots, j_{\kappa-1}, j_\kappa = 0$, be a cycle from 0 in $\mathcal{M}$, passing through 1. We may fix a real number A_* and $p > 0$ such that

$$\mathbb{P}\left\{ \sum_{i=0}^{\kappa-1} \left(\widetilde{X}_{t+i}^{(j_i)} + \log A_{t+i}(j_{i+1}, j_i) \right) > -\kappa A_* \right\} \geq p \text{ almost surely,} \tag{38}$$

where $\mathcal{D}$ is the sigma-algebra generated by all the matrices $D_t(0)$.

Assume that $\varepsilon \leq \mathrm{e}^{-1}$ and $T > \log \varepsilon^{-1}$. Defining $k = \lfloor T/m \rfloor$ and $m = \lfloor \kappa \log \varepsilon^{-1}/\tilde{\mu} \rfloor + \kappa$, we will apply (25) by considering only excursions of length exactly $m-1$, which proceed exactly through the sequence of sites $0 = j_0, j_1, \ldots, j_{I-1}, 1, \ldots, 1, j_{I+1}, \ldots, j_{\kappa-1}, j_\kappa = 0$, where site 1 is repeated exactly $m - \kappa + 1$ times. The idea is that the excursion fills a time block of length m, proceeding as quickly as possible from 0 to 1, remaining as long as possible at 1, and then returning as quickly as possible to 0.

We define the standard excursion $\mathbf{e}_\circ := (j_1, \ldots, j_{I-1}, 1, \ldots, 1, j_{I+1}, \ldots, j_{\kappa-1})$, with $m - \kappa + 1$ repetitions of site 1; and an excursion sequence $\hat{\mathbf{e}}_\circ$ consisting of those pairs $(\ell m + 1, \mathbf{e}_\ell)$ for which

$$Y_\ell := \mathbf{e}_\circ \Big[\ell m + 1; \mathbf{X}, \mathbf{A} \Big] > 0. \tag{39}$$

That is, $\hat{\mathbf{e}}_\circ$ is assembled from identical excursions of form $\mathbf{e}_\circ$ which can start only at times $\ell m + 1$. Each one of the k possible excursions is included precisely when its contribution to the sum

would be positive.

We have

$$\hat{\mathbf{e}}[\mathbf{X},\mathbf{A}] = \sum_{\ell=0}^{k-1} (Y_\ell)_+ .$$

Since the excursion contributions Y_ℓ are all independent, combining (25) with the Strong Law of Large Numbers yields

$$a(\varepsilon) - a(0) \geq \frac{\mathbb{E}[(Y_\ell)_+]}{m}. \tag{40}$$

We now observe that for any ℓ

$$Y_\ell = \sum_{i=0}^{I-1} \alpha_{\ell m+i}(j_i, j_{i+1}) + \sum_{i=I}^{\kappa-1} \alpha_{(\ell+1)m-\kappa+i}(j_i, j_{i+1}) + \sum_{t=\ell m+I}^{(\ell+1)m-\kappa+I-1} \widetilde{X}_t^{(1)} + \varepsilon\left(A_t(1,1) - A_t(0,0)\right).$$

Note that the α_t terms are independent of the $\widetilde{X}_t^{(1)}$ terms, because they draw their component random variables from different times.

By (38) and the definition of m, which makes $\kappa \log \varepsilon^{-1} \leq \tilde{\mu}(m-\kappa+1)$, for any $y > 0$

$$\begin{aligned}\mathbb{P}\Big\{Y_\ell > (m-\kappa+1)y\Big\} &\geq p\mathbb{P}\Big\{(m-\kappa+1)Z > (m-\kappa+1)(y+\tilde{\mu}) + \kappa\Big(\log\varepsilon^{-1} + A_*\Big)\Big\} \\ &\geq p\mathbb{P}\Big\{Z > \Big(y+2\tilde{\mu}\Big) + \frac{\kappa}{m-\kappa+1}\cdot\frac{A_*}{\log\varepsilon^{-1}}\Big)\Big\},\end{aligned}$$

where

$$Z := (m-\kappa+1)^{-1} \sum_{t=\ell m+I}^{(\ell+1)m-\kappa+I-1} (\widetilde{X}_t^{(1)} + \tilde{\mu}).$$

is the mean of independent mean-zero random variables whose moment generating functions have Legendre transform $I(z) \leq \frac{z^2}{2\underline{\tau}^2} = \frac{\rho z^2}{2\tilde{\mu}}$. By Cramér's Theorem (Dembo and Zeitouni 2009)

$$\begin{aligned}\lim_{m\to\infty}(m-\kappa+1)^{-1}\log\mathbb{P}\Big\{Z > \Big(y+\tilde{\mu}(2+\delta)\Big)\Big\} &\geq -I(y+2\tilde{\mu}) \\ &\geq -\frac{\rho}{2\tilde{\mu}}(y+2\tilde{\mu})^2.\end{aligned} \tag{41}$$

Holding y fixed, it follows that for any $\delta > 0$ we have for m sufficiently large (which is to say, for

ε sufficiently small)

$$\mathbb{P}\left\{\frac{Y_\ell}{m} > y\right\} \geq p\exp\left\{-\frac{m(y+2\tilde{\mu})^2\underline{\rho}}{2\tilde{\mu}} - m\frac{\tilde{\mu}\delta}{4\kappa}\right\}.$$

Taking ε sufficiently small that $\log\varepsilon^{-1}\delta \geq 1/2m\tilde{\mu}\underline{\rho}$ and $\log\varepsilon^{-1} \leq \varepsilon^{-\delta/2}$, it follows, by Lemma 14, that

$$\begin{aligned}
\mathbb{E}\Big[(Y_\ell)_+\Big] &\geq \frac{p}{m\underline{\rho}}\exp\left\{-\frac{m(2\tilde{\mu})^2\underline{\rho}}{2\tilde{\mu}} - \frac{\delta}{4}\log\varepsilon^{-1} - \frac{1}{8m\tilde{\mu}\underline{\rho}}\right\} \\
&\geq \frac{p}{m\underline{\rho}}\exp\left\{-\left(2\kappa\underline{\rho} + \frac{\delta}{2}\right)\log\varepsilon^{-1}\right\} \\
&\geq \frac{p\kappa}{\mu\underline{\rho}}\varepsilon^{2\kappa\underline{\rho}+\delta}.
\end{aligned}$$

Combining this with (20) and (40) completes the proof for the case $A_t(j,j) \equiv 0$. The statements for general A follow then from the linear contribution of $\varepsilon\sum A_t(0,0)$ to (40). The rest follows as in the conclusion of section 5 by considering the lower bound, applied to a modified $\widetilde{X}_t^{(j)}$, taken as $X_t^{(j)} - X_t^{(0)} + \varepsilon\left(A_t(j,j) - A_t(0,0)\right)$, showing that the contribution of $A(j,j) - A(0,0)$ over all non-trivial excursions can shift the power $\underline{\eta}$ by an arbitrarily small amount, which may be incorporated into δ.

Remark 13. *If the random variables $\widetilde{X}_t^{(1)}$ are actually Gaussian, rather than sub-Gaussian, then Z is simply a standard normal random variable, so we can use what we know about Gaussian tails for fixed m rather than relying on the limit as $m \to \infty$ (41). This improves the lower bound from $\varepsilon^{2\kappa\rho+\delta}$ to $C\varepsilon^{2\kappa\rho}(\log\varepsilon^{-1})^{-1}$ for some constant C.*

Lemma 14. *Let Z be a random variable with tail bound*

$$\mathbb{P}\Big\{Z > z\Big\} \geq \mathrm{e}^{-(z_0+z)^2/2\tau_Z} \textit{ for all } z > 0.$$

Then

$$\mathbb{E}[Z_+] \geq \mathrm{e}^{-z_0^2/2\tau_Z} \cdot \mathrm{e}^{-\tau_Z/2z_0^2} \cdot \frac{\tau_Z}{\mathrm{e}z_0}.$$

Proof.

$$\begin{aligned}
\mathbb{E}[Z_+] &= \int_0^\infty \mathbb{P}\Big\{Z > z\Big\}\,\mathrm{d}z \\
&\geq \mathrm{e}^{-z_0^2/2\tau_Z} \int_0^\infty \mathrm{e}^{-z(2z_0+z)/2\tau_Z}\,\mathrm{d}z \\
&\geq z_* \mathrm{e}^{-z_0^2/\tau_Z} \mathrm{e}^{-z_*(2z_0+z_*)/2\tau_Z}
\end{aligned}$$

for any fixed $z_* > 0$, since the integrand is decreasing. Taking $z_* = \tau_Z/z_0$ gives us the result. $\square$