

Derivatives of the Stochastic Growth Rate

David Steinsaltz
Department of Statistics
University of Oxford
1 South Parks Road
Oxford OX1 3TG
United Kingdom
steinsal@stats.ox.ac.uk

Shripad Tuljapurkar
Biology Department
Stanford University
Stanford, CA 94305
USA

Carol Horvitz
Biology Department
University of Miami
P.O. Box 249118
Coral Gables, FL 33124
USA

March 13, 2011

Abstract

We consider stochastic matrix models for population driven by random environments which form a Markov chain. The top Lyapunov exponent a , which describes the long-term growth rate, depends smoothly on the demographic parameters (represented as matrix entries) and on the parameters that define the stochastic matrix of the driving Markov chain. The derivatives of a — the “stochastic elasticities” — with respect to changes in the demographic parameters were derived by Tuljapurkar (1990). These results are here extended to a formula for the derivatives with respect to changes in the Markov chain driving the environments. We supplement these formulas with rigorous bounds on computational estimation errors, and with rigorous derivations of both the new and the old formulas.

1 Introduction

Stochastic matrix models for structured populations are widely used in evolutionary biology, demographic forecasting, ecology, and population viability analysis (e.g., Tuljapurkar (1990); Lee and Tuljapurkar (1994); Morris and Doak (2002); Caswell (2001); Lande et al. (2003)). In these models, a discrete-time stochastic process drives changes in environmental conditions that determine the population’s stage-transition rates (survival, fertility, growth, regression and so on). Population dynamics are described by a product of randomly chosen population projection matrices. In most biological situations the population’s stage structure converges to a time-varying but stable structure Cohen (1977), and in the long run the population grows at a stochastic growth rate a that is not random and is the leading Lyapunov exponent of the random product of population projection matrices Furstenberg and Kesten (1960a); Cohen (1977); Lange (1979); Lange and Holmes (1981); Tuljapurkar and Orzack (1980). This growth rate a is of considerable biological interest, as a fitness measure for a stage-structured phenotype Tuljapurkar (1982), as a determinant of population viability and persistence Tuljapurkar and Orzack (1980); Morris and Doak (2002); Lande et al. (2003), and in a variety of invasion problems in evolution and epidemiology Metz et al. (1992).

The map between environments and projection matrices describes how phenotype change depends on the environment — the phenotypic norm of response — and we are often interested in how populations respond to changes in, say, the mean or variance of the projection matrix elements. Such questions are answered by computing the derivatives of a with respect to changes in the projection matrices, using a formula derived by Tuljapurkar (1990). Tuljapurkar et al. (2003) called these derivatives stochastic elasticities, to contrast with the elasticity of the dominant eigenvalue of a fixed projection matrix to the elements of that matrix (Caswell, 2001). Stochastic elasticity has been used to examine evolutionary questions (Haridas and Tuljapurkar, 2005) and the effects of climate change (Morris et al., 2008). At the same time, a is also a function of the stochastic process that drives environments. Many processes, such as climate change (Boyce et al., 2006), will result in changes in the frequencies of, or the probabilities of transition between, environmental states. How is a affected by a change in the pattern and distribution of environments, rather than by a change in the population projection matrices? To answer this question, we consider a model in which the environment makes transitions among several discrete states, according to a Markov chain. Then what we want is the derivative of a with respect to changes in the transition probabilities of

this Markov chain. This derivative exists (at least away from the boundaries of the space of stochastic matrices), and in fact we know from Peres (1992) that a is an analytic function of the parameters of both the projection matrices and the parameters defining the stochastic matrix, in an open neighborhood of the set of stochastic matrices. In deterministic models (Caswell, 2001), the growth rate is represented as $\lambda = e^r$; then sensitivities are derivatives of the form $(\partial\lambda/\partial x)$ with respect to a parameter x whereas elasticities are proportional derivatives of the form $(\partial r/\partial \log x)$. In stochastic models we compute derivatives of a , and these can be used to compute elasticities (as in Tuljapurkar et al. (2003)) or sensitivities.

Our first contribution here is a new formula for computing the derivative of a with respect to changes in the transition probabilities of the environmental Markov chain, given in abstract form as equation (40). To obtain this result we show how an initial environmental state affects future growth, using coupling and importance sampling; this analysis may be of independent interest. Even with a formula in hand we must compute derivatives of a by numerical simulation which is subject to both sampling (Monte Carlo) error and bias. Our second contribution here is to show how one can bound these estimation errors. Our third contribution is a rigorous proof of the heuristically derived formula given by Tuljapurkar (1990) for the derivatives of a to the elements of the population projection matrices.

In Section 2 of this paper we set out the model and assumptions, the approach to finding derivatives, along with necessary facts about the convergence of population structures and distributions. In Section 3 we discuss systematic and sampling errors and show how we can bound them. We illustrate this approach in Section 4 by presenting bounds (in Theorem 1) for simulation estimates of the stochastic growth rate a and (in Theorem 2) for the derivatives of a with respect to projection matrix elements. In Section 5 we define a measure of the effect of an initial environmental state on subsequent population growth and show how to estimate this measure using coupling arguments. Section 6 presents (in Theorem 4) the formula, algorithm, and error bounds for the derivative of a with respect to the elements of the Markov chain that drives environments. We end by discussing how these theorems can be applied and some related issues concerning parameter estimation in such models. Proofs are in the Appendix.

2 The Model

We consider a population whose individuals exist in K different stages (these may be, for example, ages, developmental stages or size classes). Newborns are in stage $i = 1$. The progression between stages occurs at discrete time intervals at rates that depend on the environment in each time interval. The environment e_t in period t is in one of M possible states; we denote the set of possible environments by $\mathcal{M} = \{1, \dots, M\}$. Individuals in stage i at time t move to stage j at a rate $X_{e_{t+1}}(j, i)$. These rates are elements of a nonnegative population projection matrix, and at time t when the environment is e_t this matrix is denoted by X_{e_t} ; there are M such matrices, one for each environmental state. We assume that allocation of individuals to classes and the identification of environment states are certain. We also assume that the total number of individuals in the population is large enough that we can ignore sampling variation. Successive environments are chosen according to a Markov process with transition matrix P whose elements are $P(e, \check{e})$ and whose stationary distribution is $\nu = \{\nu(e)\}$. We follow the standard convention for Markov chains, that $P(e, \check{e})$

represents the probability of a transition from state e to state \check{e} ; note that this is the reverse of the convention used in matrix population models.

To guarantee demographic weak ergodicity (Cohen 1977) we assume that

- (i) There exists some $R > 0$ such that any product $X_{e_1} \cdots X_{e_R}$ has all entries positive. This implies, in particular, that each population projection matrix is *row-allowable* and *column-allowable*. That is, every row and every column has at least one positive entry.
- (ii) The chain is ergodic (so transitive and aperiodic), and environments are in the stationary distribution of P , which will also be denoted by $\nu(e)$.

The population in year t is represented by a vector $N_t \in \mathbb{R}_+^K := \{(n_t(1), \dots, n_t(K))^T : n_t(i) > 0\}$. The superscript T will always mean transpose; here it indicates that population vectors are column vectors. (There may be population classes early on that have 0 members; condition (ii) above forces all classes eventually to have positive membership, and so we assume without loss of generality that we start with all population classes occupied.) The population structure changes according to $N_{t+1} = X_{e_{t+1}} N_t$, and

$$N_t = X_{e_t} X_{e_{t-1}} \cdots X_{e_1} N_0. \quad (1)$$

The normalized population structure $N_t / (\sum_i N_t(i))$ does not converge to a fixed limit (as it would if the environment were constant) but it does converge in distribution.

At each stage i , we define

$$a := \lim_{t \rightarrow \infty} t^{-1} \log N_t(i) = \lim_{t \rightarrow \infty} t^{-1} \log \frac{\sum_i N_t(i)}{\sum_i N_0(i)} \quad (2)$$

A fundamental result in this area of research is the Furstenberg-Kesten Theorem (Furstenberg and Kesten (1960b)), which tells us that the long-run growth rate exists and is not random. This a is called the “stochastic growth rate”. The proof of this fact is fairly straightforward, within the framework of modern Markov chain theory. We present a version in section 3.1 and a proof in the appendix, both for the reader’s convenience and because it illustrates some of the important basic ideas that we draw on throughout this work.

The main result of this paper is formula (40), which represents the derivative of a with respect to an incremental change in the probability of transitioning from e to \check{e} . If we denote this derivative by $A_{e,\check{e}}$, then the change corresponding to shifting P in the direction of a matrix W is a linear functional

$$\frac{da(P + \epsilon W)}{d\epsilon} = \sum_{e,\check{e} \in \mathcal{M}} A_{e,\check{e}} W_{e,\check{e}}.$$

We note that, while the existence of such a derivative follows from the general theory of Peres (1992), the computable formula is new.

Counting the K^2 parameters in each matrix, there are at most $(MK^2 + M^2)$ parameters. While all parameters must be nonnegative, and all elements of the population projection matrices but the birth rates must be ≤ 1 , the only universal constraint is $\sum_{e \in \mathcal{M}} P(\check{e}, e) = 1$. (There may, however, be further constraints imposed, as some transitions may be impossible. If we are considering age-structured populations, the matrices are Leslie matrices, each with only $2K - 1$ potentially nonzero parameters.) The sensitivities we examine are derivatives of a with respect

to these parameters. We will define a one-dimensional family of population matrices or environment transition matrices, and we refer to change “in the direction of” the derivative of this family. This will be made precise in the analyses that follow.

3 Technical background

3.1 Contraction

We denote the K -dimensional column vector with 1’s in all places by $\mathbf{1}$. By default we use the L^1 norm $\|x\| := \sum |x_i|$ when x is a vector in \mathbb{R}^K , and write $\|X\| := \sum_{i,j=1}^K |X_{i,j}|$ when X is a $K \times K$ matrix. Note that $\|X\| = \|X\mathbf{1}\|$ when X is a nonnegative matrix. Our assumptions imply that there are positive constants \hat{k} and \hat{r} , such that for any environments e_1, \dots, e_m ,

$$|\log \|X_{e_m} \cdots X_{e_1}\|| \leq \hat{k} + m\hat{r} \quad (3)$$

We use the Hilbert projective metric ρ , described in Golubitsky et al. (1975) and in section 3.1 of Seneta (1981), defined by

$$\rho(x, y) := \log \max_{1 \leq i \leq K} x(i)/y(i) + \log \max_{1 \leq i \leq K} y(i)/x(i). \quad (4)$$

This is a pseudometric on \mathbb{R}_+^K that is a metric on $\mathcal{S} := \{(x(1), \dots, x(K))^T : x(i) > 0 \text{ and } \sum x(i) = 1\}$. The distance between two vectors is defined by the ray from the origin; that is, $\rho(x, y) = \rho(x/\|x\|, y/\|y\|)$. It is straightforward to show — cf. Bushell (1973) — that

$$\frac{1}{2} [\min\{x(i)\} + \min\{y(i)\}] e^{-\rho(x,y)} \rho(x, y) \leq \|x - y\| \leq e^{\rho(x,y)} - 1$$

for any $x, y \in \mathcal{S}$. (The upper bound is shown by Bushell with respect to the Euclidean norm, but the same argument holds for any L^p norm.) Thus, convergence in the projective metric implies that the projections onto \mathcal{S} converge in the standard norms. It is also straightforward to show that

$$\max_{1 \leq i \leq K} \left| \log \frac{x(i)}{y(i)} \right| \leq \rho(x, y) + \left| \log \frac{\|x\|}{\|y\|} \right|. \quad (5)$$

The *Birkhoff contraction coefficient* (cf. Seneta (1981)) is defined for nonnegative matrices X as

$$\tau_B(X) := \sup_{u \neq v \in \mathcal{S}} \frac{\rho(Xu, Xv)}{\rho(u, v)}. \quad (6)$$

It is clear that τ_B is sub-multiplicative (that is, $\tau_B(XY) \leq \tau_B(X)\tau_B(Y)$). We also have the formula (Theorem 3.12 of Seneta (1981))

$$\tau_B(X) = \frac{1 - \phi(X)^{1/2}}{1 + \phi(X)^{1/2}}, \quad \text{where } \phi(X) = \begin{cases} \min_{i,j,k,\ell} \frac{X_{ik}X_{j\ell}}{X_{jk}X_{i\ell}} & \text{if } X > 0, \\ 0 & \text{otherwise.} \end{cases} \quad (7)$$

(This depends on the assumption that X is row-allowable. We will be applying this result only to cases in which X is positive, so our bound on τ_B will be strictly less than 1.)

An immediate consequence is Lemma 3.1.

Lemma 3.1. For any fixed R , set $r := \max_{e_1, \dots, e_R \in \mathcal{M}} \tau_B(X_{e_R} \cdots X_{e_1})^{1/R}$. Then for any $e_1, \dots, e_n \in \mathcal{M}$,

$$\tau_B(X_{e_n} \cdots X_{e_1}) \leq k_1 r^n, \quad (8)$$

where $k_1 := r^{1/R-1}$. In particular, if R is chosen so that every product $X_{e_R} \cdots X_{e_1}$ is positive, then $r < 1$.

We also know that for any $u, \check{u} \in \mathcal{S}$, and v^T any positive row vector,

$$\log \min \left\{ \frac{u(i)}{\check{u}(i)} \right\} \leq \log \frac{v^T u}{v^T \check{u}} \leq \log \max \left\{ \frac{u(i)}{\check{u}(i)} \right\}.$$

Since the left-hand side is ≤ 0 and the right-hand side ≥ 0 , so

$$|\log v^T u - \log v^T \check{u}| \leq \rho(u, \check{u}). \quad (9)$$

Following Lemma 1 of Lange (1979), we define a compact subset $\mathcal{U} \subset \mathcal{S}$ which is stable under the transformations $u \rightarrow X_e u / \|X_e u\|$ for any $e \in \mathcal{M}$ and includes the vector $\mathbf{1}/K$, as well as all vectors of the form $X_{e_R} \cdots X_{e_1} y / \|X_{e_R} \cdots X_{e_1} y\|$ where $y \in \mathcal{S}$; and a compact subset $\mathcal{V} \subset \mathcal{S}^T$ which is stable under the transformations $v^T \rightarrow v^T X_e / \|v^T X_e\|$ and includes the vector $\mathbf{1}^T/K$ as well as all vectors of the form $y^T X_{e_R} \cdots X_{e_1} / \|y^T X_{e_R} \cdots X_{e_1}\|$, where $y \in \mathcal{S}$. Note that we could take \mathcal{U} to be any \mathcal{U}_n , defined to be the union of all products $X_{e_1} \cdots X_{e_n} \mathcal{S}$. These are a decreasing sequence of compact sets, with $X_e \mathcal{U}_n \subset \mathcal{U}_{n+1}$. whose limit \mathcal{U}_∞ is in general a fractal set, which is the support of the stationary distribution π . For practicality we are likely to work with a \mathcal{U} which is a finite-stage approximation to \mathcal{U}_∞ .

Since \mathcal{U} is a compact subset of \mathcal{S} , its diameter $\Delta := \sup_{u, \check{u} \in \mathcal{U}} \rho(u, \check{u})$ is finite. Combining this with Lemma 3.1, there is a constant $k_2 = k_1 \Delta$ such that

$$\rho(X_{e_m} \cdots X_{e_1} u, X_{e_m} \cdots X_{e_1} \check{u}) \leq k_1 r^m \rho(u, \check{u}) \leq k_2 r^m. \quad (10)$$

(Cf. Theorem 2 of Lange and Holmes (1981).) Of course, the constants may be chosen so that the same relation holds for the transposed matrices, with $u^T, \check{u}^T \in \mathcal{V}$. We use then the definition

$$\Delta := \max \left\{ \sup_{u, \check{u} \in \mathcal{U}} \rho(u, \check{u}), \sup_{v, \check{v} \in \mathcal{U}} \rho(v, \check{v}) \right\}. \quad (11)$$

We also define

$$\Delta_{\mathcal{U}}(u) := \sup_{\check{u} \in \mathcal{U}} \rho(u, \check{u}); \quad \Delta_{\mathcal{V}}(v) := \sup_{\check{v} \in \mathcal{V}} \rho(v, \check{v}). \quad (12)$$

Since for any positive vectors u, u', u'' we have $\rho(u+u', u'') \leq \max\{\rho(u, u''), \rho(u', u'')\}$, the maximum of $\rho(u, u')$ among vectors in a cone is taken between extreme points of the cone, we can compute Δ as the maximum of all distances of the form $\rho(X_{e_R} \cdots X_{e_1} \mathbf{1}, X_{e'_R} \cdots X_{e'_1} \mathbf{1})$; and distances of the form $\rho(\mathbf{1} X_{e_R} \cdots X_{e_1}, \mathbf{1} X_{e'_R} \cdots X_{e'_1})$.

It follows (as in Lemma 2 of Lange and Holmes (1981)) that for any $u, \check{u} \in \mathcal{U}$ and environments e, e_1, \dots, e_m ,

$$\left| \log \frac{\|X_e X_{e_m} \cdots X_{e_1} u\|}{\|X_{e_m} \cdots X_{e_1} u\|} - \log \frac{\|X_e X_{e_m} \cdots X_{e_1} \check{u}\|}{\|X_{e_m} \cdots X_{e_1} \check{u}\|} \right| \leq k_1 r^m \rho(u, \check{u}) \leq k_2 r^m. \quad (13)$$

The same relation holds when the matrices X are replaced by their transposes, with $u^T, \check{u}^T \in \mathcal{V}$.

Since $\|X\| = \|X\mathbf{1}\|$, and $\mathbf{1}/K$ is in \mathcal{U} , it immediately follows that if $\check{e}_1, \dots, \check{e}_i$ are any other environments,

$$\left| \log \frac{\|X_e X_{e_m} \cdots X_{e_1}\|}{\|X_{e_m} \cdots X_{e_1}\|} - \log \frac{\|X_e X_{e_m} \cdots X_{e_{i+1}} X_{\check{e}_i} \cdots X_{\check{e}_1}\|}{\|X_{e_m} \cdots X_{e_{i+1}} X_{\check{e}_i} \cdots X_{\check{e}_1}\|} \right| \leq k_2 r^{m-i}. \quad (14)$$

We note that the results in Lange and Holmes (1981) depend only on the set of matrices X being compact, not on it being finite. In section 5 and beyond we will be letting the matrices X_e and/or the transition matrix P depend smoothly on a parameter x , which will take values either in $[-x_0, x_0]$ or $[0, x_0]$. We may then choose the sets \mathcal{U} and \mathcal{V} and constants k_1, k_2, r such that the properties above — in particular, the stability of \mathcal{U} and \mathcal{V} and the bounds (10) and (13) — hold simultaneously for all values of the parameter.

One slightly more complicated bound is proved in the appendix.

Lemma 3.2. *Given $v, \check{v}, u, \check{u} \in \mathbb{R}_+^K$, and constants $k_u, k_v \leq \frac{1}{4}$ with $\max |\log v_i / \check{v}_i| \leq k_v$ and $\max |\log u_j / \check{u}_j| \leq k_u$, we have*

$$\left| \log \frac{v^T u}{\check{v}^T \check{u}} - \log \frac{v^T \check{u}}{\check{v}^T u} \right| \leq 8k_u k_v. \quad (15)$$

This condition is satisfied if

$$k_v := \rho(v, \check{v}) + |\log \|v\| / \|\check{v}\|| \leq \frac{1}{4} \text{ and } k_u := \rho(u, \check{u}) + |\log \|v\| / \|\check{v}\|| \leq \frac{1}{4}.$$

3.2 Time reversal and the stationary distribution

The transition matrix for the *time-reversal* of P will be denoted \tilde{P} , and is given by

$$\tilde{P}(e, \check{e}) := \frac{\nu(\check{e})}{\nu(e)} P(\check{e}, e).$$

A standard result (for example, see Theorem 6.5.1 of Grimmett and Stirzaker (2001)) tells us that if $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_m$ form a stationary Markov chain with transition matrix P , for a fixed m , the reversed sequence $\mathbf{e}_m, \mathbf{e}_{m-1}, \dots, \mathbf{e}_1$ is a Markov chain with transition matrix \tilde{P} . (We use the boldface \mathbf{e} to represent a random environment — a random variable with values in \mathcal{M} — and e to represent a particular environment or a realisation of a random variable \mathbf{e} .)

As described in Lange and Holmes (1981), if $\mathbf{e}_1, \mathbf{e}_2, \dots$ forms a stationary Markov chain with transition probabilities P , there is a unique distribution π on $\mathcal{U} \times \mathcal{M}$ which is stable under the transformation $(u, \mathbf{e}_t) \mapsto (X_{\mathbf{e}_t} u / \|X_{\mathbf{e}_t} u\|, \mathbf{e}_{t+1})$. That is, if the normalized population structure $N_t / \|N_t\|$ paired with \mathbf{e}_{t+1} is chosen from the distribution π , then the pair $(N_{t+1} / \|N_{t+1}\|, \mathbf{e}_{t+2})$ will also be in the distribution π .

This Markov chain is naturally represented as an *iterated function system*; that is, a step in the chain is made by choosing randomly from a collection of possible functions, and applying it to the current position of the chain. In this case, the function is $u \mapsto X_{\mathbf{e}} u / \|X_{\mathbf{e}} u\|$, where \mathbf{e} is the next random environment. Time reversal is a convenient approach for studying the asymptotic behavior of such systems, and more general random dynamical systems. For applications, see Barnsley and Elton (1988); Elton (1990); Steinsaltz (1999).

To understand how this works, suppose we extend our Markov chain of environments forward and backward in time to infinity: $\dots, \mathbf{e}_{-2}, \mathbf{e}_{-1}, \mathbf{e}_0, \mathbf{e}_1, \mathbf{e}_2, \dots$. Starting at any s , positive or negative, moving forward we have a Markov chain with transition matrix P : $\mathbf{e}_s, \mathbf{e}_{s+1}, \mathbf{e}_{s+2}, \dots$. Moving backward, $\mathbf{e}_s, \mathbf{e}_{s-1}, \mathbf{e}_{s-2}, \dots$ is a Markov chain with transition matrix \tilde{P} . Choose any population vector $u_0 \in \mathcal{U}$, and define for $s > t$,

$$U_{s,t} := \frac{X_{\mathbf{e}_s} X_{\mathbf{e}_{s-1}} \cdots X_{\mathbf{e}_t} u_0}{\|X_{\mathbf{e}_s} X_{\mathbf{e}_{s-1}} \cdots X_{\mathbf{e}_t} u_0\|}. \quad (16)$$

From (10), we know that for any $t_1, t_2 \geq m$, we have the bound $\rho(U_{s,-t_1}, U_{s,-t_2}) \leq k_2 r^{m+s}$, which means that $(U_{s,-t})_{t=0}^\infty$ is a Cauchy sequence. We may denote its limit by $U_{s,-\infty}$.

We note four important features of this backward iterated sequence:

- (i) If the environments all shift one to the left, so that we start at \mathbf{e}_{s+1} instead of \mathbf{e}_s , the distribution remains exactly the same. Thus, the distribution of $U_{s,-\infty}$ is the same for any s .
- (ii) If we start our population Markov chain in the state $(U_{0,-\infty}, \mathbf{e}_1)$, then the next state will be

$$\left(\frac{X_{\mathbf{e}_1} U_{0,-\infty}}{\|X_{\mathbf{e}_1} U_{0,-\infty}\|}, \mathbf{e}_2 \right) = (U_{1,-\infty}, \mathbf{e}_2).$$

Thus, the distribution of $(U_{0,-\infty}, \mathbf{e}_1)$ is a stationary distribution π for the chain.

- (iii) For each fixed t , $U_{0,-t}$ has the same distribution as

$$U_{t,0} = \frac{X_{\mathbf{e}_t} X_{\mathbf{e}_{t-1}} \cdots X_{\mathbf{e}_0} u_0}{\|X_{\mathbf{e}_t} X_{\mathbf{e}_{t-1}} \cdots X_{\mathbf{e}_0} u_0\|}, \quad (17)$$

which is a realization of the normalized population vector $N_t/\|N_t\|$, when it is started at the vector u_0 . Thus, the population-environment Markov chain $(N_t/\|N_t\|)$ converges in distribution to the same stationary distribution π , no matter what the initial condition.

- (iv) We will use frequently the fact that $(U_{0,-t}, \mathbf{e}_1)$ may be considered an approximate sample from the distribution π , with an error that is bounded by

$$\begin{aligned} \rho(U_{0,-t}, U_{0,-\infty}) &= \rho(X_{\mathbf{e}_0} \cdots X_{\mathbf{e}_{-t}} u_0, X_{\mathbf{e}_0} \cdots X_{\mathbf{e}_{-t}} \cdot U_{-t-1,-\infty}) \\ &\leq \Delta_{\mathcal{U}}(u_0) \tau_B(X_{\mathbf{e}_0} \cdots X_{\mathbf{e}_{-t}}), \end{aligned}$$

which is straightforward to bound (*cf.* section 3.1), independent of any information about $U_{-t,-\infty}$.

The same holds true, of course, if we reverse the matrix multiplication: Starting from any nonnegative K -dimensional row vector v_0^T , we define the sequence of row vectors

$$V_{0,t}^T := \frac{v_0^T X_{\mathbf{e}_t} \cdots X_{\mathbf{e}_0}}{\|v_0^T X_{\mathbf{e}_t} \cdots X_{\mathbf{e}_0}\|} \in \mathcal{V}.$$

Then $V_{0,t}$ converges pointwise to a vector $V_{0,\infty} := \lim_{t \rightarrow \infty} V_{0,t}$. We denote the distribution of $(V_{0,\infty}, \mathbf{e}_{-1})$ by $\tilde{\pi}$. As before, $\tilde{\pi}$ is the stationary distribution for the Markov chain on $\mathcal{V} \times \mathcal{M}$, defined by taking (v^T, \mathbf{e}_{-t}) at time t to $(v^T X_{\mathbf{e}_{-t}} / \|v^T X_{\mathbf{e}_{-t}}\|, \mathbf{e}_{-(t+1)})$ at time $t+1$.

We also define the regular conditional distributions π_e on \mathcal{U} as follows: pick (U, \mathbf{e}) from the distribution π , conditioned on \mathbf{e} being e , and take π_e to be the distribution of U . We define $\tilde{\pi}_e$ similarly. In the case of i.i.d. environments, of course, π and $\tilde{\pi}$ would be simply products of an independent population vector and the stationary environment with distribution ν .

4 Errors and How to Bound Them

A standard approach to estimating a — and a similar approach to the derivatives that we describe later on — is to choose a fixed starting vector $u_0 \in \mathcal{U}$, simulate sequences $\mathbf{e}_0(i), \mathbf{e}_1(i), \dots, \mathbf{e}_{m_i}(i)$ independently from the stationary Markov chain with transition probabilities P ($i = 1, \dots, J$), and then compute

$$\begin{aligned} a_m &:= \mathbb{E} \left[\log \frac{\|X_{\mathbf{e}_{m_i}(i)} X_{\mathbf{e}_{m_i-1}(i)} \cdots X_{\mathbf{e}_0(i)} u_0\|}{\|X_{\mathbf{e}_{m_i-1}(i)} \cdots X_{\mathbf{e}_0(i)} u_0\|} \right] \\ &\approx \frac{1}{J} \sum_{i=1}^J \left(\log \|X_{\mathbf{e}_{m_i}(i)} X_{\mathbf{e}_{m_i-1}(i)} \cdots X_{\mathbf{e}_0(i)} u_0\| \right. \\ &\quad \left. - \log \|X_{\mathbf{e}_{m_i-1}(i)} \cdots X_{\mathbf{e}_0(i)} u_0\| \right). \end{aligned} \tag{18}$$

It is important not only to know what would be an appropriate approximation to a or its derivatives in the sense of being asymptotically correct, but also to have rigorous bounds for the error arising from any finite simulation procedure, such as (18). There are two sources of error: systematic error, arising from the fact that $(X_{\mathbf{e}_0}, u_0)$ is not exactly a sample from the distribution π ; and sampling error, arising from the fact that we have estimated the expectation by averaging over a random sample.

For obtaining an unbiased estimate there is no reason the sample Markov chain realizations should need to be independent. For instance, a standard approach would be to take a single realization of the chain $\mathbf{e}_0, \mathbf{e}_1, \mathbf{e}_2, \dots$, and take $\mathbf{e}_t(i) = \mathbf{e}_t$, with $m_i = B + i$ for some “burn-in time” B . The problem with this approach is that it becomes more difficult to bound the errors; in principle, the convergence could be extremely slow for these kinds of sums. We leave the problem of bounding errors for these cumulative simulations to a future work.

4.1 Systematic error

By “systematic error” we mean the error in our estimate of a arising from the difference between the distribution we are aiming for and the distribution we are actually sampling from. The quantity we are trying to estimate may be represented as $a = \pi[F]$, the expectation of a certain function F with respect to the distribution π . If we can simulate Z_1, \dots, Z_J from π , then $\hat{a}_J := J^{-1} \sum_{j=1}^J F(Z_j)$ is an unbiased estimator of $\pi[F]$, and will be consistent under modest assumptions on F and the independence of the samples. Suppose, though, that what we have are not samples from π , but samples Z_j^* from a “similar” distribution π^* . Then we can bound the error by

$$|a - \hat{a}| \leq \left| J^{-1} \sum_{j=1}^J F(Z_j^*) - \pi^*[F] \right| + \left| \pi[F] - \pi^*[F] \right|. \tag{19}$$

Here the first term on the right-hand side is the sampling error, and the second term is the bias, the expected value of systematic error. The problem is that the bounds we can obtain for the bias are likely to be crude, absent good computational tools for the distribution π . (And if we could compute analytically from π , we wouldn't need to be simulating).

An alternative is to couple the samples Z_j^* from the approximate distribution π' to exact samples Z_j from the distribution π , we can break up the error in a slightly different way:

$$\begin{aligned} |a - \hat{a}| &\leq \left| J^{-1} \sum_{j=1}^J F(Z_j) - \pi[F] \right| + \left| J^{-1} \sum_{j=1}^J F(Z_j) - J^{-1} \sum_{j=1}^J F(Z_j^*) \right| \\ &\leq \left| J^{-1} \sum_{j=1}^J F(Z_j) - \pi[F] \right| + J^{-1} \sum_{j=1}^J \left| F(Z_j) - F(Z_j^*) \right| \end{aligned} \quad (20)$$

Bounds for the sampling error in (19) will generally also be bounds for the first term in (20). The second term in (20), on the other hand, which takes the place of the bias, is a random variable, computed from the samples Z_j^* . Its expectation is still a bound on the bias. The crucial fact is that the last line may be computable without knowing in detail what the “true” sample Z_j is.

A small disadvantage of this approach is that the systematic error varies with the sample. To achieve a particular fixed error bound we need an adaptive approach, whereby we successively extend our sequence of matrices until the error crosses the desired threshold. We note here that this approach to estimating the systematic error in simulations is essentially just a version of the Propp-Wilson algorithm (*cf.* Chapter 10 of Häggström (2002)). Unlike standard applications, though, the space is continuous, so the systematic error never reaches 0.

4.2 Sampling error

The sampling error is difficult to control with current techniques, because the distribution of the samples is so poorly understood — the very reason why we resort to the Monte Carlo approximation in the first place. The best we can do for a rigorous bound is to use Hoeffding's inequality (see Hoeffding (1963)), relying on crude bounds on the terms in the expectation. Hoeffding's inequality tells us that if X_1, \dots, X_J are i.i.d. random variables such that $\alpha \leq X_i \leq \beta$ almost surely, then for any $z > 0$,

$$\mathbb{P} \left\{ \left| \frac{1}{J} \sum X_i - \mathbb{E}[X] \right| > z \right\} \leq 2e^{-2Jz^2/(\beta-\alpha)^2}. \quad (21)$$

This is essentially the same bound that we would estimate from the normal approximation if the standard deviation of X were $(\beta - \alpha)/2$. Generally we will want to fix p_0 , the confidence level, and compute the corresponding z , which will be

$$z_0 = (\beta - \alpha) \sqrt{-\frac{1}{2J} \log(p_0/2)}. \quad (22)$$

Of course, the standard deviation will be smaller than this, but we do not know how much smaller. An alternative approach then would be to use the bound $2\tau(z\sqrt{J}/\hat{\sigma})$, where $\hat{\sigma}$ is the standard deviation of the simulated samples, and τ is the cumulative distribution function of the Student t distribution with $J - 1$ degrees

of freedom. This will be a smaller bound, in that sense “better”, but not precisely true for finite samples, to the extent that the sample distribution is not normal. The corresponding bound on the error, at probability level p_0 , is

$$z_0 = \frac{\hat{\sigma}}{\sqrt{J}} t_{1-p_0/2}(J-1), \quad (23)$$

where $t_p(J-1)$ is the p quantile of the Student T distribution with $J-1$ degrees of freedom; that is, if T has this distribution then $P\{T > t_p(J-1)\} = p$.

These asymptotic bounds are perfectly conventional in Markov chain Monte Carlo analysis (see, for example, Asmussen and Glynn (2007)), and there is no reason particularly to eschew them in this context. They are likely to be quite accurate, and superior to the rigorous bounds, and might also be applied to the setting where the expectations are estimated not from independent samples, but from a single run of the environment chain. We have nonetheless emphasised the Hoeffding-based rigorous bounds in the statements of our theorems for three reasons: First, they are likely to be less familiar, and the reader may require more guidance in applying them to the individual cases; second, because some residue of skepticism must remain for the asymptotic bounds, while these results may be applied to calculations that are otherwise analytically precise; and third, as a spur to further thought on the best ways to bound errors rigorously in these sorts of problems.

5 Growth Rate and Sensitivity to Projection Matrices

We present here extensions of two known results. In these cases (and in later results) we start by defining an estimator that converges to the the quantity we desire, and follow that by bounds on the systematic and sampling errors, as well as an error bound for estimates from a simulation estimator. We state our results on error bounds in the form “The quantity Q may be approximated by the expectation of A , with systematic error bounded by B and sampling error bounded by $C(J, p)$.” This means that if A_1, \dots, A_J are independent realizations of A , then the probability that the true value of Q is not in the interval $J^{-1} \sum A_i \pm [B + C(J, p)]$ is no bigger than p . When describing an adaptive bound on the systematic error, B will depend upon the particular simulation result. Again, the sampling error may be bounded either by a universally valid Hoeffding bound, based on known upper bounds on the samples, or by the Student t distribution, using the standard deviation estimated from the sample, which provides a generally much superior bound, but which can only be treated as an approximation.

5.1 Computing a

The stochastic growth rate a is commonly estimated by numerical simulation but, as discussed with examples by Caswell (2001), there is no general way to bound the errors in the estimated values. The following result provides suitable bounds.

Theorem 5.1. *Let u_0 be any fixed element of \mathcal{U} , and $Y_m := X_{\mathbf{e}_m} X_{\mathbf{e}_{m-1}} \cdots X_{\mathbf{e}_1}$, where $\mathbf{e}_0, \mathbf{e}_1, \dots$ form a Markov chain with transition rates P . The stochastic growth rate may be approximated by the simulated expectation of*

$$\log \frac{\|X_{\mathbf{e}_{m+1}} Y_m u_0\|}{\|Y_m u_0\|}, \quad (24)$$

with systematic error bounded by $k_2 r^m$ and sampling error at level p on J samples bounded by

$$\left(\log \frac{\sup_{u \in \mathcal{U}} \max_{e \in \mathcal{M}} \|X_e u\|}{\inf_{u \in \mathcal{U}} \min_{e \in \mathcal{M}} \|X_e u\|} \right) \left(\frac{-\log p}{2J} \right)^{1/2}. \quad (25)$$

When the simulated expectation is

$$\frac{1}{J} \sum_{i=1}^J \log \frac{\|X_{e_{m+1}}(i) Y_m(i) u_0\|}{\|Y_m(i) u_0\|}$$

we may also bound the systematic error by

$$\frac{1}{J} \sum_{i=1}^J \sup_{u, \check{u} \in \mathcal{U}} \rho(Y_m(i)u, Y_m(i)\check{u}) \leq \frac{\Delta_{\mathcal{U}}(u_0)}{J} \sum_{i=1}^J \tau_B(Y_m(i)), \quad (26)$$

where $\Delta_{\mathcal{U}}(u_0)$ is defined as in (12) and τ_B is defined as in (6).

5.2 Derivatives with respect to Projection Matrices

We need care in defining derivatives of a with respect to elements of the population projection matrices. As discussed in Tuljapurkar and Horvitz (2003) we must define how the matrix entries change, e.g., do we change fertility rates in a particular environment, or in all possible environments? Although the main formula here is known, Tuljapurkar's (1990) derivation did not justify a crucial exchange of limits (between taking the perturbation to zero and time to infinity). We provide a rigorous proof (see Appendix) and of course the error bounds here are new.

We will suppose that the matrices X_e depend smoothly on a parameter ϵ , so that we may define $X'_e := \partial X_e / \partial \epsilon$, and we define the base matrices to be at $\epsilon = 0$. In some cases, the parametrization will be defined only for $\epsilon \geq 0$, and in those cases we will understand the partial derivatives to be one-sided derivatives, and the limits $\lim_{\epsilon \rightarrow 0}$ will be the one-sided limits $\lim_{\epsilon \downarrow 0}$.

Theorem 5.2. *Let U_e and V_e be independent random variables with distributions π_e and $\tilde{\pi}_e$ (the conditional stationary distributions defined in section 3.2). Then*

$$a'(0) = \sum_{e \in \mathcal{M}} \nu_e \mathbb{E} \begin{bmatrix} V_e^T X'_e U_e \\ V_e^T X_e U_e \end{bmatrix} \quad (27)$$

Each term may be approximated by averaging samples of the form

$$\sum_{e \in \mathcal{M}} \nu_e \frac{V^{(m)T} X'_e U^{(m)}}{V^{(m)T} X_e U^{(m)}}, \quad (28)$$

where u_0, v_0^T are any fixed elements of \mathcal{U}, \mathcal{V} respectively, $U^{(m)} = X_{\tilde{\mathbf{e}}_1} \cdots X_{\tilde{\mathbf{e}}_m} u_0$ and $V^{(m)T} = v_0^T X_{\mathbf{e}_m} \cdots X_{\mathbf{e}_1}$, $e = \tilde{\mathbf{e}}_0, \tilde{\mathbf{e}}_1, \dots, \tilde{\mathbf{e}}_m$ form a sample from the Markov chain \tilde{P} , and $e = \mathbf{e}_0, \mathbf{e}_1, \dots, \mathbf{e}_m$ form a sample from the Markov chain P . The systematic error may be bounded uniformly by

$$2(\exp(4k_2 r^m) - 1)|a'(0)|, \quad (29)$$

while the sampling error at level p on J samples is bounded by

$$\left(\frac{-\log(p/2)}{2J}\right)^{1/2} \sup_{\substack{u \in \mathcal{U}, v^T \in \mathcal{V} \\ \check{u} \in \mathcal{U}, \check{v}^T \in \mathcal{V}}} \sum_{e \in \mathcal{M}} \nu_e \left(\frac{v^T X'_e u}{v^T X_e u} - \frac{\check{v}^T X'_e \check{u}}{\check{v}^T X_e \check{u}} \right). \quad (30)$$

Suppose the simulated expectation is

$$\frac{1}{J} \sum_{j=1}^J \sum_{e \in \mathcal{M}} \nu_e \frac{(V^{(m)}(j))^T X'_e U^{(m)}(j)}{(V^{(m)}(j))^T X_e U^{(m)}(j)},$$

where

$$U^{(m)}(j) = X_{\mathbf{e}_1(j)} \cdots X_{\mathbf{e}_m(j)} u_0 =: \tilde{Y}_m(j) u_0 \text{ and} \\ (V^{(m)}(j))^T = v_0^T X_{\mathbf{e}_m(j)} \cdots X_{\mathbf{e}_1(j)} =: v_0^T Y_m(j).$$

Let

$$\mathcal{U}(j) := \tilde{Y}_m(j) \mathcal{U} = \{\tilde{Y}_m(j) u : u \in \mathcal{U}\}, \\ \mathcal{V}(j) := \mathcal{V} Y_m(j) = \{v^T Y_m(j) : v^T \in \mathcal{V}\}.$$

Then we may also bound the systematic error by

$$\begin{aligned} & \sum_{e \in \mathcal{M}} \frac{\nu_e}{J} \sum_{j=1}^J \sup_{\substack{u \in \mathcal{U}(j) \\ v^T \in \mathcal{V}(j)}} \left| \frac{v^T X'_e u}{v^T X_e u} - \frac{(V^{(m)}(j))^T X'_e U^{(m)}(j)}{(V^{(m)}(j))^T X_e U^{(m)}(j)} \right| \\ & \leq \sum_{e \in \mathcal{M}} \frac{\nu_e}{J} \sum_{j=1}^J \left(\exp \left\{ 2 \sup_{u \in \mathcal{U}(j)} \rho(u, U^{(m)}(j)) + 2 \sup_{v^T \in \mathcal{V}(j)} \rho(v, V^{(m)}(j)) \right\} - 1 \right) \\ & \quad \times \sup_{\substack{u \in \mathcal{U}(j) \\ v^T \in \mathcal{V}(j)}} \frac{v^T |X'_e| u}{v^T X_e u} \\ & \leq \sum_{e \in \mathcal{M}} \frac{\nu_e}{J} \sum_{j=1}^J \sup_{\substack{u \in \mathcal{U}(j) \\ v^T \in \mathcal{V}(j)}} \frac{v^T |X'_e| u}{v^T X_e u} \left(\exp \left\{ 2\Delta \left(\tau_B(\tilde{Y}_m(j)) + \tau_B(Y_m(j)^T) \right) \right\} - 1 \right). \end{aligned} \quad (31)$$

Note that the bound (29) is given as a proportion of the unknown $a'(0)$. It can be turned into an explicit bound by using an upper bound on $a'(0)$. We have the trivial bound

$$|a'(0)| \leq \sum_{e \in \mathcal{M}} \nu_e \sup_{\substack{u \in \mathcal{U} \\ v^T \in \mathcal{V}}} \frac{|v^T X'_e u|}{v^T X_e u}. \quad (32)$$

We recall that Δ and τ_B may be bounded according to the formulas given in section 3.1.

It may seem surprising that the vectors U_e and V_e are independent. If we think of the environments as a doubly infinite sequence with $\mathbf{e}_0 = e$, then the vector U_e in equation (27) depends only on the past (that is, \mathbf{e}_i with $i < 0$) and V_e depends only on the future (\mathbf{e}_i with $i > 0$). By the Markov property, these two are independent, when conditioned on $\mathbf{e}_0 = e$.

6 Environments and Coupling

Suppose we change the transition matrix P to a slightly different matrix $P^{(\epsilon)}$, and want to compare population growth along environmental sequences generated by the original and the perturbed matrix. We expect that the perturbed environmental sequences will only occasionally deviate from the environment that we “would have had” in the original distribution of environments. Computing the derivative of a is then a matter of measuring the cumulative deviations due to these changes. These may be split into two parts: First, the process moves under $P^{(\epsilon)}$ to a state \check{e} , different from the state e that it would have moved to under P . Then there is a sequence of environments following on from \check{e} that is different from the sequence that would have followed from e , until the Markov chain gradually “forgets” its starting point. The change to a new sequence of matrices induces two separate changes on the growing population: The new sequence accumulates a difference in magnitude on its way to stationarity; and it produces a different stationary distribution of unit vectors depending on the starting environment \check{e} rather than e .

In this section we examine the first effect. We fix the transition matrix P and compare the growth of total population size when starting from environment e to the growth starting from the stationary distribution ν . A standard method for doing this is *coupling*. For an outline of coupling techniques in MCMC, see Kendall (2005) and Roberts and Rosenthal (2004). We use coupling in two ways, corresponding to the two components of the Markov chain: the environment and the population vector.

Fix environments e and \check{e} (possibly the same). We define sequences $\mathbf{e}_0, \mathbf{e}_1, \dots$; $\check{\mathbf{e}}_0, \check{\mathbf{e}}_1, \dots$; and $\bar{\mathbf{e}}_0, \bar{\mathbf{e}}_1, \dots$: all three are Markov chains with transition probabilities P , but with $\mathbf{e}_0 = e$, $\check{\mathbf{e}}_0 = \check{e}$ and $\bar{\mathbf{e}}_0$ having distribution ν (so that $(\bar{\mathbf{e}}_i)$ is stationary). We define the total population effect of starting in state e rather than \check{e} as

$$\begin{aligned} \check{e}\zeta_e &:= \lim_{t \rightarrow \infty} \left(\mathbb{E} [\log \|X_{\mathbf{e}_t} \cdots X_{\mathbf{e}_0}\|] - \mathbb{E} [\log \|X_{\check{\mathbf{e}}_t} \cdots X_{\check{\mathbf{e}}_0}\|] \right) \\ \zeta_e &:= \lim_{t \rightarrow \infty} \left(\mathbb{E} [\log \|X_{\mathbf{e}_t} \cdots X_{\mathbf{e}_0}\|] - \mathbb{E} [\log \|X_{\bar{\mathbf{e}}_t} \cdots X_{\bar{\mathbf{e}}_0}\|] \right) \\ &= \sum_{\check{e} \in \mathcal{M}} \nu_{\check{e}} \cdot \check{e}\zeta_e. \end{aligned} \tag{33}$$

This is one of the terms that will come into formula (40), the derivative of a with respect to shifting transition probabilities from e to \check{e} .

Note that when the environments are i.i.d. — so $P(e, \check{e}) = \nu_{\check{e}}$ — we have

$$\check{e}\zeta_e = \mathbb{E} \left[\log \frac{\|V^T X_e\|}{\|V^T X_{\check{e}}\|} \right],$$

where V^T has the distribution π .

Computing ζ_e depends on coupling the version of the Markov chain starting at e , to another version starting in the distribution ν . We define the *coupling time* τ to be the first time such that $\mathbf{e}_\tau = \bar{\mathbf{e}}_\tau$; after this time the chains follow identical trajectories. If we know the distribution of τ and of the sequences followed by the two chains from time 0 to τ , we can average the differences in (33) to find ζ . The advantage of coupling is, first, that it reduces the variability of the estimates, and second, that we know from the simulation when the coupling time has been achieved, which gives bounds on the error. A suitable choice is Griffeath’s maximal

coupling (Griffeath, 1975) which we will apply in Pitman's (Pitman, 1976) path-decomposition representation. (The coupling is "maximal" in the sense of making the coupling time, and hence the variance of the estimate, as small as possible.) However we must be careful about sampling values of τ , because they may be large if the Markov chain mixes slowly. To deal with this we overweight large coupling times that generate a large contribution to ζ .

Beginning with a fixed environment e , the procedure is as follows:

- (i) Define the sequence of vectors $\alpha_t := P^t(e, \cdot) - \nu(\cdot)$. We also define α_t^+ and α_t^- to be the vectors of pointwise positive and negative parts respectively. Let $C(t)$ be a bound on $|\log \|X_{e_1} \cdots X_{e_t}\| - \log \|X_{\check{e}_1} \cdots X_{\check{e}_t}\||$, where the e_i and \check{e}_i are any environments. From (3) we know that $2\hat{k} + 2t\hat{r}$ is a possible choice for $C(t)$.
- (ii) For pairs (t, \mathbf{e}) , where t is a positive integer and $e_* \in \mathcal{M}$, define a probability distribution

$$q(t, \check{e}) := \begin{cases} \nu_e & \text{if } e = \check{e}, t = 0; \\ 0 & \text{if } e \neq \check{e}, t = 0; \\ [\alpha_{t-1}^+ P](\check{e}) - \alpha_t^+(\check{e}) & \text{otherwise.} \end{cases}$$

This is the distribution of the pair (τ, \mathbf{e}_τ) for the maximally coupled chain. Define

$$A := \sum_{t=1}^{\infty} \sum_{\check{e} \in \mathcal{M}} q(t, \check{e}) C(t),$$

and a probability distribution on $\mathbb{N} \times \mathcal{M}$

$$\hat{q}(t, \check{e}) := \frac{q(t, \check{e}) C(t)}{A}.$$

- (iii) Average J independent realizations of the following random variable: Let (τ, \mathbf{e}_0) be chosen from the distribution \hat{q} on $\mathbb{N} \times \mathcal{M}$, and $\check{\mathbf{e}}_0$ independently from the distribution ν . From these starting states \mathbf{e}_0 and $\check{\mathbf{e}}_0$, let $(\mathbf{e}_0, \dots, \mathbf{e}_\tau, \dots, \mathbf{e}_m)$ and $(\check{\mathbf{e}}_0, \dots, \check{\mathbf{e}}_\tau, \dots, \check{\mathbf{e}}_m)$ be realizations of the coupled pair of Markov chains with transition probabilities P , conditioned on the coupling time being τ and $\mathbf{e}_\tau = \check{\mathbf{e}}_\tau = \check{\mathbf{e}}$. These realizations are generated from independent inhomogeneous Markov chains running backward, with transition probabilities

$$\mathbb{P}\{\mathbf{e}_{i-1} = x \mid \mathbf{e}_i = y\} = \frac{\alpha_{i-1}^+(x) P(x, y)}{\sum_{z \in \mathcal{M}} \alpha_{i-1}^+(z) P(z, y)},$$

$$\mathbb{P}\{\check{\mathbf{e}}_{i-1} = x \mid \check{\mathbf{e}}_i = y\} = \frac{\alpha_{i-1}^-(x) P(x, y)}{\sum_{z \in \mathcal{M}} \alpha_{i-1}^-(z) P(z, y)}.$$

We extend the chain past τ , requiring $\mathbf{e}_t = \check{\mathbf{e}}_t$ for $t > \tau$, as a realization of the Markov transition probabilities P , to obtain a total sample of predetermined length m . The random variable is then

$$Z := \frac{A}{C(\tau)} \log \frac{\|X_{\mathbf{e}_m} \cdots X_{\mathbf{e}_\tau} \cdots X_{\mathbf{e}_0}\|}{\|X_{\mathbf{e}_m} \cdots X_{\mathbf{e}_{\tau+1}} X_{\check{\mathbf{e}}_\tau} \cdots X_{\check{\mathbf{e}}_0}\|}.$$

(Note that the realizations corresponding to $\tau = 0$ are identically 0. The possibility of $\tau = 0$ has been included only to simplify the notation. In practice, we are free to condition on $\tau > 0$.)

The change from q to \hat{q} is an example of importance sampling (cf. Chapter V.1 in Asmussen and Glynn (2007)). We oversample the values of the random variable with high τ to reduce the variability of the estimate. The importance sampling makes $Z(j)$ a bounded random variable, with bound A . Imagine that we had a source of perfect samples $V^T(j)$ from the distribution $\tilde{\pi}_{\mathbf{e}_m}$, and define

$$\tilde{Z}(j) := \frac{A}{C(\tau(j))} \log \frac{\|V^T(j)X_{\mathbf{e}_m(j)} \cdots X_{\mathbf{e}_0(j)}\|}{\|V^T(j)X_{\check{\mathbf{e}}_m(j)} \cdots X_{\check{\mathbf{e}}_0(j)}\|}.$$

Let $Y(j) := X_{\mathbf{e}_m(j)} \cdots X_{\mathbf{e}_0(j)}$ and $\check{Y}(j) := X_{\check{\mathbf{e}}_m(j)} \cdots X_{\check{\mathbf{e}}_0(j)}$. Then

$$|\tilde{Z}(j) - Z(j)| \leq \frac{A\Delta\nu(\mathbf{1})}{C(\tau(j))} \left(\tau_B(Y(j)) + \tau_B(\check{Y}(j)) \right). \quad (34)$$

Since $\mathbb{E}[\tilde{Z}] = \zeta_e$, we may use (22) to compute the bound

$$\mathbb{P} \left\{ \left| \zeta_e - n^{-1} \sum_{j=1}^n \tilde{Z}(j) \right| > 2A \sqrt{-\frac{1}{2J} \log(p_0/2)} \right\} \leq p_0. \quad (35)$$

Lemma 6.1. *The limits defining the coefficients ${}_\varepsilon\zeta_e$ and ζ_e exist and are finite. We may approximate ζ_e by*

$$\frac{1}{J} \sum_{j=1}^J \frac{A}{C(\tau(j))} \log \frac{\|X_{e_\tau(j)} \cdots X_{e_0(j)}\|}{\|X_{\check{e}_\tau(j)} \cdots X_{\check{e}_0(j)}\|}. \quad (36)$$

If $0 < p_0 \leq 1$, the probability is no more than p_0 that the error in this estimation is larger than

$$\frac{1}{J} \sum_{j=1}^J \frac{A\Delta\nu(\mathbf{1})}{C(\tau(j))} \left(\tau_B(Y(j)) + \tau_B(\check{Y}(j)) \right) + 2A \sqrt{-\frac{1}{2J} \log(p_0/2)}, \quad (37)$$

It remains to bound A . From standard Markov chain theory, for any vector $v \in \mathcal{S}$,

$$\|v^T P^t - \nu^T\| \leq D\xi^t. \quad (38)$$

Setting setting $Q := P - \mathbf{1}\nu^T$, α_t^T is the e -th row of Q^t , which may also be written as

$$\alpha_t = \mathbf{1}_e^T (Q^t)^T,$$

where $\mathbf{1}_e$ is the column vector with 1 in place e and 0 elsewhere. Thus

$$\|\alpha_t\| \leq D\xi^t.$$

If we use the bound $C(t) = \hat{k} + t\hat{r}$, then

$$\begin{aligned} A &= \sum_{t=1}^{\infty} (\hat{k} + t\hat{r}) (\|\alpha_{t-1}\| - \|\alpha_t\|) \\ &= \hat{k}\|\alpha_1\| + \hat{r} \sum_{t=1}^{\infty} \|\alpha_t\| \\ &\leq D\xi \left(\hat{k} + \frac{D\hat{r}}{1-\xi} \right) \end{aligned} \quad (39)$$

7 Derivatives with respect to Environmental Transitions

We are now ready to compute derivatives of a with respect to changes in the distribution of environments, as determined by P . Complicating the notation slightly is the constraint $\{P : \sum_{\check{e}} P(e, \check{e}) = 1 \text{ for each } e\}$; thus, there can be no sense in speaking of the derivative with respect to changes in $P(e, \check{e})$ for some particular e, \check{e} . Instead, we must compute directional derivatives along the direction of some matrix W , in the plane $\sum_{\check{e}} W_{e, \check{e}} = 0$.

For the purposes of this result we write $a(\epsilon) = a(P^{(\epsilon)})$, where $P^{(\epsilon)}$ is a differentiable curve of $M \times M$ matrices, where the parameter ϵ takes values either in a two-sided interval $[-\epsilon_0, \epsilon_0]$, or a one-sided interval $[0, \epsilon_0]$. Let $W := \partial P^{(\epsilon)} / \partial \epsilon$, an $M \times M$ matrix whose rows all sum to 0. The perturbations are such that $P^{(\epsilon)}$ retains the ergodicity and irreducibility of P . (The result should be the same whether ϵ is positive or negative. If P is on the boundary of the set of possible values, one or the other sign may be impossible. Some choices of W may be impossible in both directions.) In the special case in which $W_{e, \check{e}} = 1$ and $W_{e, \check{e}} = -1$, with all other entries 0, we are computing the derivative corresponding to a small increase in the rate of transitioning from environment e to \check{e} , and a decrease in the rate of transitioning to \check{e} .

In this section the matrices X_1, \dots, X_M are assumed fixed.

Theorem 7.1. *The derivative of the stochastic growth rate is*

$$a'(0) = \sum_{e, \check{e} \in \mathcal{M}} \nu_{\check{e}} W_{\check{e}, e} \left(\zeta_e + \mathbb{E} \left[\log \frac{V_e^T X_e X_{\check{e}} U_{\check{e}}}{\|V_e^T X_e\|} \right] \right), \quad (40)$$

where $U_{\check{e}}, V_e^T$ are independent random variables with distributions $\pi_{\check{e}}$ and $\tilde{\pi}_e$ respectively.

The quantities ζ_e may be approximated, with error bounds, according to the algorithm described in section 6.

The other part of the expression may be approximated by averaging samples of the form

$$\sum_{e, \check{e} \in \mathcal{M}} \nu_{\check{e}} W_{\check{e}, e} \log \frac{V_e^{(m)T} X_e X_{\check{e}} U_{\check{e}}^{(m)}}{\|V_e^{(m)T} X_e\|}, \quad (41)$$

where

$$U_{\check{e}}^{(m)} = \frac{X_{\check{e}_1} \cdots X_{\check{e}_m} u_0}{\|X_{\check{e}_1} \cdots X_{\check{e}_m} u_0\|} \text{ and}$$

$$V_e^{(m)T} = \frac{v_0^T X_{e_m} \cdots X_{e_1}}{\|v_0^T X_{e_m} \cdots X_{e_1}\|},$$

and $\mathbf{e} = \mathbf{e}_0, \mathbf{e}_1, \dots, \mathbf{e}_m$ is a Markov chain with transition matrix P , and $\tilde{\mathbf{e}} = \tilde{\mathbf{e}}_0, \tilde{\mathbf{e}}_1, \dots, \tilde{\mathbf{e}}_m$ is an independent Markov chain with transition probabilities \tilde{P} , and $u_0 \in \mathcal{U}$ and $v_0^T \in \mathcal{V}$.

The systematic error may be bounded uniformly by $2k_2 r^m \|(\nu^T |W|)\|$, while the sampling error at level p on J samples bounded by

$$2\|\nu^T |W|\| \sup_{v^T \in \mathcal{V}} \max_{u \in \mathcal{U}} \max_{e \in \mathcal{M}} |\log v^T X_e u| \left(\frac{-\log p}{2J} \right)^{1/2}. \quad (42)$$

Suppose the simulated expectation is

$$\frac{1}{J} \sum_{j=1}^J \sum_{\tilde{e}, e \in \mathcal{M}} \nu_{\tilde{e}} W_{e, \tilde{e}} \log \frac{V^{(m, \tilde{e})T}(j) X_{\tilde{e}} X_e U^{(m, e)}(j)}{\|V^{(m, \tilde{e})T}(j) X_{\tilde{e}}\|},$$

where

$$U_{\tilde{e}}^{(m)}(j) = \frac{X_{\tilde{e}_1(j)} \cdots X_{\tilde{e}_m(j)} u_0}{\|X_{\tilde{e}_1(j)} \cdots X_{\tilde{e}_m(j)} u_0\|} =: \frac{\tilde{Y}_{\tilde{e}}^{(m)}(j) u_0}{\|\tilde{Y}_{\tilde{e}}^{(m)}(j) u_0\|} \text{ and}$$

$$V_e^{(m)T}(j) = \frac{v_0^T X_{\mathbf{e}_m(j)} \cdots X_{\mathbf{e}_1(j)}}{\|v_0^T X_{\mathbf{e}_m(j)} \cdots X_{\mathbf{e}_1(j)}\|} =: \frac{v_0^T Y_e^{(m)}(j)}{\|v_0^T Y_e^{(m)}(j)\|}.$$

We may also bound the systematic error by

$$\begin{aligned} & \frac{1}{J} \sum_{j=1}^J \sum_{\tilde{e}, e \in \mathcal{M}} \nu_{\tilde{e}} |W_{\tilde{e}, e}| \left(\sup_{u, \tilde{u} \in \mathcal{U}} \rho(\tilde{Y}_{\tilde{e}}^{(m)}(j) u, \tilde{Y}_{\tilde{e}}^{(m)}(j) \tilde{u}) \right. \\ & \qquad \qquad \qquad \left. + \sup_{v^T, \tilde{v}^T \in \mathcal{V}} \rho(v^T Y_e^{(m)}(j) X_e, \tilde{v}^T Y_e^{(m)}(j) X_e) \right) \quad (43) \\ & \leq \frac{\Delta}{J} \sum_{j=1}^J \sum_{\tilde{e}, e \in \mathcal{M}} \nu_{\tilde{e}} |W_{e, \tilde{e}}| \left(\tau_B(\tilde{Y}_{\tilde{e}}^{(m)}(j)) + \tau_B(X_e^T Y_e^{(m)}(j)^T) \right). \end{aligned}$$

We note that expressions like (40) are examples of what Brémaud (1992) calls “ersatz derivatives”. In a rather different class of applications Brémaud suggests applying maximal coupling.

8 Discussion

Our results provide analytical formulas and simulation estimators for the derivatives of stochastic growth rate with respect to the transition probability matrix or the population projection matrices. We have concentrated here on the theoretical results; although this may not be obvious, we have made considerable effort at brevity. Partly for this reason, we will present elsewhere numerical applications of these results. We expect that our results should carry over to integral population models (IPMs), given the strong parallels between the stochastic ergodicity properties of IPMs and matrix models (Ellner and Rees, 2007).

Our results apply not only to stochastic structured populations but to any stochastic system in which a Lyapunov exponent of a product of random matrices determines stability or other dynamic properties. Examples include the net reproductive rate in epidemic models and some models of network dynamics. An obvious application of our results is to the analysis of optimal life histories, i.e., environment-to-projection matrix maps that maximize the stochastic growth rate. As discussed by McNamara (1997), this optimization problem translates into what is called an average reward problem in stochastic control theory, and so our results may be more generally useful in such control problems.

9 Acknowledgements

We thank NIA (BSR) for support under 1P01 AG22500. David Steinsaltz was supported by a New Dynamics of Ageing grant, a joint interdisciplinary program of the UK research councils.

Appendix

Proofs of the theorems.

A.1 Proof of Lemma 3.2

We have

$$\begin{aligned}
\left| \log \frac{v^T u}{\check{v}^T u} - \log \frac{v^T \check{u}}{\check{v}^T \check{u}} \right| &= \left| \log \frac{\sum_{i,j} v_i u_i \check{v}_j \check{u}_j}{\sum_{i,j} v_i \check{u}_i \check{v}_j u_j} \right| \\
&= \left| \log \frac{\sum_i u_i v_i \check{u}_i \check{u}_i + \sum_{i < j} (v_i u_i \check{v}_j \check{u}_j + v_j u_j \check{v}_i \check{u}_i)}{\sum_i u_i v_i \check{u}_i \check{u}_i + \sum_{i < j} (v_i \check{u}_i \check{v}_j u_j + v_j \check{u}_j \check{v}_i u_i)} \right| \\
&\leq \max_{i,j} \left| \log \frac{v_i u_i \check{v}_j \check{u}_j + v_j u_j \check{v}_i \check{u}_i}{v_i \check{u}_i \check{v}_j u_j + v_j \check{u}_j \check{v}_i u_i} \right|
\end{aligned}$$

For some choice of i, j , let

$$\alpha := \log \frac{u_i}{\check{u}_i}, \quad \beta := \log \frac{v_i}{\check{v}_i}, \quad \gamma := \log \frac{u_j}{\check{u}_j}, \quad \delta := \log \frac{v_j}{\check{v}_j}.$$

Note that by (5),

$$\begin{aligned}
|\alpha| &\leq \rho(u, \check{u}) + |\log \|u\| / \|\check{u}\|| \leq \frac{1}{4}, \\
|\beta| &\leq \rho(v, \check{v}) + |\log \|v\| / \|\check{v}\|| \leq \frac{1}{4}, \\
|\gamma| &\leq \rho(u, \check{u}) + |\log \|u\| / \|\check{u}\|| \leq \frac{1}{4}, \\
|\delta| &\leq \rho(v, \check{v}) + |\log \|v\| / \|\check{v}\|| \leq \frac{1}{4}.
\end{aligned} \tag{44}$$

Then we need to bound

$$\begin{aligned}
\left| \log \frac{e^{\alpha+\beta} + e^{\gamma+\delta}}{e^{\beta+\delta} + e^{\alpha+\gamma}} \right| &\leq \frac{|e^{\alpha+\beta} + e^{\gamma+\delta} - e^{\beta+\delta} - e^{\alpha+\gamma}|}{\min\{e^{\alpha+\beta} + e^{\gamma+\delta}, e^{\beta+\delta} + e^{\alpha+\gamma}\}} \\
&\leq |e^{\alpha+\beta} + e^{\gamma+\delta} - e^{\beta+\delta} - e^{\alpha+\gamma}|,
\end{aligned} \tag{45}$$

where we have used the relation $|\log x| \leq \max\{x, 1/x\} - 1$.

We now use the Taylor series expansion

$$e^x = \sum_{k=0}^n \frac{x^k}{k!} + \frac{C x^{k+1}}{(k+1)!},$$

where $|C| \leq \max\{e^x, 1\}$. This turns the right-hand side of (45) into

$$\begin{aligned}
& 2 \sum_{k=2}^{\infty} \frac{1}{k!} |(\alpha + \beta)^k + (\gamma + \delta)^k - (\beta + \delta)^k - (\alpha + \gamma)^k| \\
&= 2 \sum_{k=2}^{\infty} \left| \sum_{\ell=1}^{k-1} \frac{1}{\ell!(k-\ell)!} (\alpha^\ell \beta^{k-\ell} + \gamma^\ell \delta^{k-\ell} - \beta^\ell \delta^{k-\ell} - \alpha^\ell \gamma^{k-\ell}) \right| \\
&\leq 2 \sum_{k=0}^{\infty} \frac{1}{(k-2)!} \left(\alpha \beta \sum_{\ell=0}^{k-2} \binom{k-2}{\ell} \alpha^\ell \beta^{k-2-\ell} + |\gamma \delta| \sum_{\ell=0}^{k-2} \binom{k-2}{\ell} \gamma^\ell \delta^{k-2-\ell} \right. \\
&\quad \left. + |\beta \delta| \sum_{\ell=0}^{k-2} \binom{k-2}{\ell} \beta^\ell \delta^{k-2-\ell} + |\alpha \gamma| \sum_{\ell=0}^{k-2} \binom{k-2}{\ell} \alpha^\ell \gamma^{k-2-\ell} \right).
\end{aligned}$$

Applying the bounds (44) yields finally the upper bound in (15).

A.2 Estimating the stochastic growth rate

We prove here Theorem 5.1. The quantity we are trying to compute is

$$a = \mathbb{E}[\log \|X_{\mathbf{e}} U\|], \text{ where } (U, \mathbf{e}) \text{ is selected from the distribution } \pi. \quad (46)$$

Let $\mathbf{e}_0, \mathbf{e}_1, \mathbf{e}_2, \dots$ be a realization of the stationary Markov chain with transition matrix P . Let $Y_m := X_{\mathbf{e}_m} X_{\mathbf{e}_{m-1}} \cdots X_{\mathbf{e}_1}$. Choose $u_0 \in \mathcal{U}$, and let U be a random variable with distribution $\pi_{\mathbf{e}_0}$. Then $a = \mathbb{E}[\log \|X_{\mathbf{e}_{m+1}} Y_m U\| / \|Y_m U\|]$, which may be approximated by $\mathbb{E}[\log \|X_{\mathbf{e}_{m+1}} Y_m u_0\| / \|Y_m u_0\|]$.

If we identify systematic error with bias, this is

$$\text{Error}_{\text{sys}} = \left| \mathbb{E} \left[\log \frac{\|X_{\mathbf{e}_{m+1}} Y_m u_0\|}{\|Y_m u_0\|} \right] - \mathbb{E} \left[\log \frac{\|X_{\mathbf{e}_{m+1}} Y_m U\|}{\|Y_m U\|} \right] \right|,$$

since $(Y_m U / \|Y_m U\|, \mathbf{e}_m)$ also has the distribution π (if Y_m and U are taken to be independent, conditioned on \mathbf{e}_0). Thus

$$\begin{aligned}
\text{Error}_{\text{sys}} &\leq \mathbb{E} \left[\left| \log \frac{\|X_{\mathbf{e}_{m+1}} Y_m u_0\|}{\|Y_m u_0\|} - \log \frac{\|X_{\mathbf{e}_{m+1}} Y_m U\|}{\|Y_m U\|} \right| \right] \\
&\leq \mathbb{E} \left[\sup_{u, \check{u} \in \mathcal{U}} \left| \log \frac{\|X_{\mathbf{e}_{m+1}} Y_m u\|}{\|Y_m u\|} - \log \frac{\|X_{\mathbf{e}_{m+1}} Y_m \check{u}\|}{\|Y_m \check{u}\|} \right| \right] \\
&\leq k_2 r^m,
\end{aligned}$$

by (13). The corresponding bound on the sampling error may be computed from (22).

For a particular choice of $\mathbf{e}_1, \dots, \mathbf{e}_{m+1}$ and U we can also represent the random systematic error as

$$\left| \log \frac{\|X_{\mathbf{e}_{m+1}} Y_m u_0\|}{\|Y_m u_0\|} - \log \frac{\|X_{\mathbf{e}_{m+1}} Y_m U\|}{\|Y_m U\|} \right|,$$

which may be bounded by the summand in (26).

A.3 Estimating sensitivities: Matrix entries

We prove here Theorem 5.2. As discussed at the end of section 3.1, we may assume that the compact sets \mathcal{U} and \mathcal{V} are stable and satisfy the bounds of section 3.1 simultaneously for all $X_e^{(\epsilon)}$. The stationary distributions corresponding to products of the perturbed matrices are denoted $\pi^{(\epsilon)}$ and $\tilde{\pi}^{(\epsilon)}$, and the corresponding regular conditional distributions are $\pi_e^{(\epsilon)}$ and $\tilde{\pi}_e^{(\epsilon)}$

The derivative $a'(0)$ may be written as

$$\begin{aligned} & \lim_{\epsilon \rightarrow 0} \epsilon^{-1} \left(\lim_{m \rightarrow \infty} \mathbb{E} \left[\log \frac{\|X_{\mathbf{e}_m}^{(\epsilon)} X_{\mathbf{e}_{m-1}}^{(\epsilon)} \cdots X_{\mathbf{e}_0}^{(\epsilon)} u_0\|}{\|X_{\mathbf{e}_{m-1}}^{(\epsilon)} \cdots X_{\mathbf{e}_0}^{(\epsilon)} u_0\|} \right] - \lim_{m \rightarrow \infty} \mathbb{E} \left[\log \frac{\|X_{\mathbf{e}_m} X_{\mathbf{e}_{m-1}} \cdots X_{\mathbf{e}_0} u_0\|}{\|X_{\mathbf{e}_{m-1}} \cdots X_{\mathbf{e}_0} u_0\|} \right] \right) \\ &= \left(\lim_{m \rightarrow \infty} \mathbb{E} \left[\log \frac{\|X_{\tilde{\mathbf{e}}_0}^{(\epsilon)} X_{\tilde{\mathbf{e}}_1}^{(\epsilon)} \cdots X_{\tilde{\mathbf{e}}_m}^{(\epsilon)} u_0\|}{\|X_{\tilde{\mathbf{e}}_1}^{(\epsilon)} \cdots X_{\tilde{\mathbf{e}}_m}^{(\epsilon)} u_0\|} \right] - \lim_{m \rightarrow \infty} \mathbb{E} \left[\log \frac{\|X_{\tilde{\mathbf{e}}_0} X_{\tilde{\mathbf{e}}_1} \cdots X_{\tilde{\mathbf{e}}_m} u_0\|}{\|X_{\tilde{\mathbf{e}}_1} \cdots X_{\tilde{\mathbf{e}}_m} u_0\|} \right] \right) \\ &= \lim_{\epsilon \rightarrow 0} \lim_{m \rightarrow \infty} \sum_{s=0}^m a_{s,m}(\epsilon) =: \lim_{\epsilon \rightarrow 0} \lim_{m \rightarrow \infty} A(m, \epsilon), \end{aligned}$$

where

$$\begin{aligned} a_{s,m}(\epsilon) := & \mathbb{E} \left[\epsilon^{-1} \left(\log \frac{\|X_{\tilde{\mathbf{e}}_0}^{(\epsilon)} X_{\tilde{\mathbf{e}}_1}^{(\epsilon)} \cdots X_{\tilde{\mathbf{e}}_s}^{(\epsilon)} X_{\tilde{\mathbf{e}}_{s+1}} \cdots X_{\tilde{\mathbf{e}}_m} u_0\|}{\|X_{\tilde{\mathbf{e}}_1}^{(\epsilon)} \cdots X_{\tilde{\mathbf{e}}_s}^{(\epsilon)} X_{\tilde{\mathbf{e}}_{s+1}} \cdots X_{\tilde{\mathbf{e}}_m} u_0\|} \right. \right. \\ & \left. \left. - \log \frac{\|X_{\tilde{\mathbf{e}}_0}^{(\epsilon)} X_{\tilde{\mathbf{e}}_1}^{(\epsilon)} \cdots X_{\tilde{\mathbf{e}}_{s-1}}^{(\epsilon)} X_{\tilde{\mathbf{e}}_s} \cdots X_{\tilde{\mathbf{e}}_m} u_0\|}{\|X_{\tilde{\mathbf{e}}_1}^{(\epsilon)} \cdots X_{\tilde{\mathbf{e}}_{s-1}}^{(\epsilon)} X_{\tilde{\mathbf{e}}_s} \cdots X_{\tilde{\mathbf{e}}_m} u_0\|} \right) \right]. \end{aligned}$$

Here $\mathbf{e}_0, \mathbf{e}_1, \dots$ is the stationary Markov chain with transition probabilities P , and $\tilde{\mathbf{e}}_0, \tilde{\mathbf{e}}_1, \dots$ is the reverse stationary Markov chain, with transition probabilities \tilde{P} .

By (13), for $\epsilon > 0$ sufficiently small

$$|a_{s,m}(\epsilon)| \leq \epsilon^{-1} k_2 r^{s-1} \sup_{u \in \mathcal{U}} \mathbb{E} [\rho(X_{\tilde{\mathbf{e}}_s}^{(\epsilon)} u, X_{\tilde{\mathbf{e}}_s} u)] \leq 2k_2 C r^s,$$

where

$$C := r^{-1} \max_{e \in \mathcal{M}} \max_{1 \leq \ell \leq K} \left\{ \max_{u \in \mathcal{U}} \frac{|X'_e u|_\ell}{(X_e u)_\ell}, \max_{v \in \mathcal{V}} \frac{|v^T X'_e|_\ell}{(v^T X_e)_\ell} \right\}.$$

If we define

$$\begin{aligned} V_{i,j} &:= \frac{\mathbf{1}^T X_{\tilde{\mathbf{e}}_i}^{(\epsilon)} X_{\tilde{\mathbf{e}}_{i+1}}^{(\epsilon)} \cdots X_{\tilde{\mathbf{e}}_j}^{(\epsilon)}}{\|\mathbf{1}^T X_{\tilde{\mathbf{e}}_i}^{(\epsilon)} X_{\tilde{\mathbf{e}}_{i+1}}^{(\epsilon)} \cdots X_{\tilde{\mathbf{e}}_j}^{(\epsilon)}\|} \\ U_{i,j} &:= \frac{X_{\tilde{\mathbf{e}}_i} X_{\tilde{\mathbf{e}}_{i+1}} \cdots X_{\tilde{\mathbf{e}}_j} u_0}{\|X_{\tilde{\mathbf{e}}_i} X_{\tilde{\mathbf{e}}_{i+1}} \cdots X_{\tilde{\mathbf{e}}_j} u_0\|}, \end{aligned}$$

then for $m \leq n$,

$$\begin{aligned} |a_{s,m}(\epsilon) - a_{s,n}(\epsilon)| &\leq \epsilon^{-1} \mathbb{E} \left[\left| \log \frac{V(0, s-1)^T X_{\tilde{\mathbf{e}}_s}^{(\epsilon)} U(s+1, m)}{V(0, s-1)^T X_{\tilde{\mathbf{e}}_s} U(s+1, m)} - \log \frac{V(0, s-1)^T X_{\tilde{\mathbf{e}}_s}^{(\epsilon)} U(s+1, n)}{V(0, s-1)^T X_{\tilde{\mathbf{e}}_s} U(s+1, n)} \right| \right. \\ &\quad \left. + \left| \log \frac{V_{1,s-1}^T X_{\tilde{\mathbf{e}}_s}^{(\epsilon)} U_{s+1, m}}{V_{1,s-1}^T X_{\tilde{\mathbf{e}}_s} U_{s+1, m}} - \log \frac{V_{1,s-1}^T X_{\tilde{\mathbf{e}}_s}^{(\epsilon)} U_{s+1, n}}{V_{1,s-1}^T X_{\tilde{\mathbf{e}}_s} U_{s+1, n}} \right| \right] \end{aligned}$$

We note by (10) that $\rho(U_{s+1,m}, U_{s+1,n}) \leq k_2 r^{m-s}$; and for $\epsilon > 0$ sufficiently small and any $v \in \mathcal{V}$ we have

$$\max_{\substack{1 \leq j \leq K, \\ e \in \mathcal{M}}} \left| \log \frac{(v^T X_e^{(\epsilon)})_j}{(v^T X_e)_j} \right| \leq 2C\epsilon.$$

It follows by Lemma 3.2 that

$$|a_{s,m}(\epsilon) - a_{s,n}(\epsilon)| \leq \min \{16k_2 C r^{m-s}, 4k_2 C r^s\}. \quad (47)$$

Putting these together, we get, for all $n \geq m$,

$$\begin{aligned} |A(m, \epsilon) - A(n, \epsilon)| &\leq \sum_{s=m+1}^n |a_{s,n}(\epsilon)| + \sum_{s=0}^m |a_{s,m}(\epsilon) - a_{s,n}(\epsilon)| \\ &\leq \sum_{s=m+1}^n 2k_2 C r^s + \sum_{1 \leq s \leq m/2} 16k_2 C r^{m-s} + \sum_{m/2 < s \leq m} 4k_2 C r^s \\ &\leq \frac{20k_2 C}{1-r} r^{m/2}. \end{aligned}$$

By Lemma A.1 we may exchange the order of the limits, to see that

$$\begin{aligned} a'(0) &= \lim_{m \rightarrow \infty} \lim_{\epsilon \rightarrow 0} \epsilon^{-1} \sum_{s=1}^m \left(\mathbb{E} \left[\log \frac{\|X_{\mathbf{e}_m}^{(\epsilon)} X_{\mathbf{e}_{m-1}}^{(\epsilon)} \cdots X_{\mathbf{e}_s}^{(\epsilon)} X_{\mathbf{e}_{s-1}} \cdots X_{\mathbf{e}_0} u_0\|}{\|X_{\mathbf{e}_{m-1}}^{(\epsilon)} \cdots X_{\mathbf{e}_s}^{(\epsilon)} X_{\mathbf{e}_{s-1}} \cdots X_{\mathbf{e}_0} u_0\|} \right] \right. \\ &\quad \left. - \mathbb{E} \left[\log \frac{\|X_{\mathbf{e}_m}^{(\epsilon)} X_{\mathbf{e}_{m-1}}^{(\epsilon)} \cdots X_{\mathbf{e}_{s+1}}^{(\epsilon)} X_{\mathbf{e}_s} \cdots X_{\mathbf{e}_0} u_0\|}{\|X_{\mathbf{e}_{m-1}}^{(\epsilon)} \cdots X_{\mathbf{e}_{s+1}}^{(\epsilon)} X_{\mathbf{e}_s} \cdots X_{\mathbf{e}_0} u_0\|} \right] \right). \end{aligned} \quad (48)$$

This limit is the same for any choice of u_0 , hence would also be the same if we replaced u_0 by a random U , with any distribution on \mathcal{U} . We choose U to have the distribution $\pi_{\mathbf{e}_0}$, independent of the rest of the Markov chain. By the invariance property of the distributions π ,

$$\begin{aligned} a'(0) &= \lim_{m \rightarrow \infty} \lim_{\epsilon \rightarrow 0} \epsilon^{-1} \left(\sum_{s=1}^m \mathbb{E} \left[\log \frac{\|X_{\mathbf{e}_m}^{(\epsilon)} X_{\mathbf{e}_{m-1}}^{(\epsilon)} \cdots X_{\mathbf{e}_s}^{(\epsilon)} X_{\mathbf{e}_{s-1}} \cdots X_{\mathbf{e}_0} U\|}{\|X_{\mathbf{e}_{m-1}}^{(\epsilon)} \cdots X_{\mathbf{e}_s}^{(\epsilon)} X_{\mathbf{e}_{s-1}} \cdots X_{\mathbf{e}_0} U\|} \right] \right. \\ &\quad \left. - \mathbb{E} \left[\log \frac{\|X_{\mathbf{e}_m}^{(\epsilon)} X_{\mathbf{e}_{m-1}}^{(\epsilon)} \cdots X_{\mathbf{e}_{s+1}}^{(\epsilon)} X_{\mathbf{e}_s} \cdots X_{\mathbf{e}_0} U\|}{\|X_{\mathbf{e}_{m-1}}^{(\epsilon)} \cdots X_{\mathbf{e}_{s+1}}^{(\epsilon)} X_{\mathbf{e}_s} \cdots X_{\mathbf{e}_0} U\|} \right] \right) \\ &= \lim_{m \rightarrow \infty} \sum_{s=0}^m \lim_{\epsilon \rightarrow 0} \epsilon^{-1} \mathbb{E} \left[\log \frac{\|X_{\mathbf{e}_m}^{(\epsilon)} X_{\mathbf{e}_{m-1}}^{(\epsilon)} \cdots X_{\mathbf{e}_s}^{(\epsilon)} U_{s-1}\|}{\|X_{\mathbf{e}_{m-1}}^{(\epsilon)} \cdots X_{\mathbf{e}_s}^{(\epsilon)} U_{s-1}\|} \right. \\ &\quad \left. - \log \frac{\|X_{\mathbf{e}_m}^{(\epsilon)} X_{\mathbf{e}_{m-1}}^{(\epsilon)} \cdots X_{\mathbf{e}_{s+1}}^{(\epsilon)} X_{\mathbf{e}_s} U_{s-1}\|}{\|X_{\mathbf{e}_{m-1}}^{(\epsilon)} \cdots X_{\mathbf{e}_{s+1}}^{(\epsilon)} X_{\mathbf{e}_s} U_{s-1}\|} \right], \end{aligned} \quad (49)$$

where (U_{s-1}, \mathbf{e}_s) has distribution π .

For $m \geq s \geq 1$ define functions

$$f_s(\epsilon, \delta) := \frac{\mathbf{1}^T X_{\mathbf{e}_m}^{(\delta)} X_{\mathbf{e}_{m-1}}^{(\delta)} \cdots X_{\mathbf{e}_{s+1}}^{(\delta)} X_{\mathbf{e}_s}^{(\epsilon)} U_{s-1}}{\mathbf{1}^T X_{\mathbf{e}_{m-1}}^{(\delta)} \cdots X_{\mathbf{e}_{s+1}}^{(\delta)} X_{\mathbf{e}_s}^{(\epsilon)} U_{s-1}},$$

where the denominator is understood to be 1 for $s = m$. The summand on the right of (49) may be written as

$$\lim_{\epsilon \rightarrow 0} \epsilon^{-1} \mathbb{E} [\log f_s(\epsilon, \epsilon) - \log f_s(0, \epsilon)]. \quad (50)$$

By (13),

$$\epsilon^{-1} |\log f_s(\epsilon, \delta) - \log f_s(0, \delta)| \leq \epsilon^{-1} k_2 r^s \rho(X_{e_s}^{(\epsilon)} U, X_{e_s} U) < 2Ck_2 r^s$$

for ϵ in a neighborhood of 0, so the Bounded Convergence Theorem turns (50) into

$$\mathbb{E} \left[\lim_{\epsilon \rightarrow 0} \epsilon^{-1} (\log f_s(\epsilon, \epsilon) - \log f_s(0, \epsilon)) \right] = \mathbb{E} \left[\frac{\partial \log f_s}{\partial \epsilon}(0, 0) \right]. \quad (51)$$

We have, by linearity of the matrix product and $\|\cdot\|$,

$$\begin{aligned} \frac{\partial \log f_s}{\partial \epsilon}(0, 0) &= \frac{\mathbf{1}^T X_{e_m} \cdots X_{e_{s+1}} \frac{\partial}{\partial \epsilon} X_{e_s}^{(\epsilon)} U_{s-1}}{\|X_{e_m} \cdots X_{e_s} U_{s-1}\|} \\ &\quad - \frac{\mathbf{1}^T X_{e_{m-1}} \cdots X_{e_{s+1}} \frac{\partial}{\partial \epsilon} X_{e_s}^{(\epsilon)} U_{s-1}}{\|X_{e_{m-1}} \cdots X_{e_s} U_{s-1}\|} \\ &= \begin{cases} \left(\frac{\mathbf{1}^T X_{e_m} \cdots X_{e_{s+1}} X'_{e_s} U_{s-1}}{\|X_{e_m} \cdots X_{e_s} U_{s-1}\|} - \frac{\mathbf{1}^T X_{e_{m-1}} \cdots X_{e_{s+1}} X'_{e_s} U_{s-1}}{\|X_{e_{m-1}} \cdots X_{e_s} U_{s-1}\|} \right) & \text{for } 1 \leq s \leq m-1, \\ \frac{\mathbf{1}^T X'_{e_m} U_{m-1}}{\|X_{e_m} U_{m-1}\|} & \text{for } s = m. \end{cases} \end{aligned}$$

Combining this with (49) yields the telescoping sum

$$\begin{aligned} a'(0) &= \lim_{m \rightarrow \infty} \left(\sum_{t=1}^m \mathbb{E} \left[\frac{\mathbf{1}^T X_{e_t} \cdots X_{e_1} X'_{e_0} U}{\mathbf{1}^T X_{e_t} \cdots X_{e_1} X_{e_0} U} \right] \right. \\ &\quad \left. - \sum_{t=1}^{m-1} \mathbb{E} \left[\frac{\mathbf{1}^T X_{e_t} \cdots X_{e_1} X'_{e_0} U}{\mathbf{1}^T X_{e_t} \cdots X_{e_1} X_{e_0} U} \right] \right) \\ &= \lim_{m \rightarrow \infty} \mathbb{E} \left[\frac{\mathbf{1}^T X_{e_m} \cdots X_{e_1} X'_{e_0} U}{\mathbf{1}^T X_{e_m} \cdots X_{e_1} X_{e_0} U} \right], \end{aligned}$$

where in the last line (U, \mathbf{e}_0) has the distribution π . Define $V_m^T := \mathbf{1}^T X_{e_m} \cdots X_{e_1} / \|\mathbf{1}^T X_{e_m} \cdots X_{e_1}\|$. Then V_m^T converges in distribution to V , with distribution $\tilde{\pi}_{\mathbf{e}_0}$ (conditioned on \mathbf{e}_0), and so

$$a'(0) = \lim_{m \rightarrow \infty} \mathbb{E} \left[\frac{V_m^T X'_{e_0} U}{V_m^T X_{e_0} U} \right] = \sum_{e \in \mathcal{M}} \nu_e \mathbb{E} \left[\frac{V_e^T X'_e U_e}{V_e^T X_e U_e} \right],$$

which is identical to (27).

Now we estimate the error. We use the representation

$$U := \frac{X_{\tilde{e}_1} \cdots X_{\tilde{e}_m} U_0}{\|X_{\tilde{e}_1} \cdots X_{\tilde{e}_m} U_0\|} \text{ and } V^T := \frac{V_0^T X_{e_m} \cdots X_{e_1}}{\|V_0^T X_{e_m} \cdots X_{e_1}\|},$$

where U_0 and V_0^T are assumed to have distributions $\pi_{\tilde{e}_m}$ and $\tilde{\pi}_{\mathbf{e}_m}$ respectively. Then

$$\begin{aligned} \left| \frac{V^{(m)T} X'_e U^{(m)}}{V^{(m)T} X_e U^{(m)}} - \frac{V^T X'_e U}{V^T X_e U} \right| &\leq 2 \left(e^{2\rho(V, V^{(m)T}) + 2\rho(U, U^{(m)})} - 1 \right) \frac{|V X'_e U|}{|V X_e U|} \\ &\leq 2 \left(e^{4k_2 r^m} - 1 \right) \frac{|V X'_e U|}{|V X_e U|}, \end{aligned} \quad (52)$$

by (13). This implies the uniform bound on systematic error, and the bound on sampling error (30) follows from applying (22) to a trivial bound on the terms in the average. The simulated bound (31) also follows directly from (52).

Lemma A.1. *Let $A(m, \epsilon)$ be a two-dimensional array of real numbers, indexed by $m \in \mathbb{N}$ and $\epsilon > 0$, with $A(m, 0) := \lim_{\epsilon \downarrow 0} A(m, \epsilon)$ existing for each m , and $A(\infty, \epsilon) := \lim_{m \rightarrow \infty} A(m, \epsilon)$ existing for all ϵ sufficiently small (independent of m). Suppose A satisfies*

$$\lim_{M \rightarrow \infty} \limsup_{\epsilon \downarrow 0} \sup_{m, n > M} |A(m, \epsilon) - A(n, \epsilon)| = 0. \quad (53)$$

Then the two limits

$$\lim_{m \rightarrow \infty} \lim_{\epsilon \downarrow 0} A(m, \epsilon) \text{ and } \lim_{\epsilon \downarrow 0} \lim_{m \rightarrow \infty} A(m, \epsilon)$$

are equal; in particular, if one exists the other exists as well.

Proof. Suppose $A_* := \lim_{\epsilon \downarrow 0} A(\infty, \epsilon)$ exists. Then we need to show that $A_* = \lim_{m \rightarrow \infty} A(m, 0)$.

Choose any $\delta > 0$, and choose M such that

$$\limsup_{\epsilon \downarrow 0} \sup_{m, n > M} |A(m, \epsilon) - A(n, \epsilon)| < \delta.$$

This means that we may find $\epsilon_0 > 0$ such that $|A(m, \epsilon) - A(n, \epsilon)| < 2\delta$ when $0 < \epsilon < \epsilon_0$ and $m, n > M$, and such that also $|A(\infty, \epsilon) - A_*| < 2\delta$ for all $0 < \epsilon < \epsilon_0$. It follows, in particular, that $|A(m, \epsilon) - A_*| < 4\delta$ when $m > M$ and $0 < \epsilon < \epsilon_0$. Thus $|A(m, 0) - A_*| < 4\delta$ for $m > M$, and consequently $|\limsup_{m \rightarrow \infty} A(m, 0) - A_*| < 4\delta$. Since δ is arbitrary, it follows that $\lim_{m \rightarrow \infty} A(m, 0) = A_*$.

The converse result (starting from the assumption that $\lim_{m \rightarrow \infty} A(m, 0)$ exists) follows identically. \square

A.4 Estimating sensitivities: Markov environments

We derive here the formula (40) by a combination of the coupling method and importance sampling. We use importance sampling for the actual computation, but coupling provides a more direct path to validating the crucial exchange of limits. As in the proof of Theorem 5.2, the error bounds are an obvious consequence of the formula (40) and the general formulas for errors described in section 4.

Given two distributions q and q^* on $\{1, \dots, M\}$, we define a *standard coupling* between q and q^* . Suppose we are given a uniform random variable ω on $[0, 1]$. Let $\mathcal{M}_- := \{e : q_e < q_e^*\}$ and $\mathcal{M}_+ := \{e : q_e^* < q_e\}$. Let $\delta := \sum_{e \in \mathcal{M}_-} (q_e^* - q_e) = \sum_{e \in \mathcal{M}_+} (q_e - q_e^*)$. We define three random variables \check{e} on \mathcal{M} , \check{e}_+ on \mathcal{M}_+ , and \check{e}_- on \mathcal{M}_- , according to the following distributions:

$$\begin{aligned} \mathbb{P}\{\check{e} = e\} &= \min\{q_e, q_e^*\} / (1 - \delta), \\ \mathbb{P}\{\check{e}_+ = e\} &= (q_e - q_e^*)_+ / \delta, \\ \mathbb{P}\{\check{e}_- = e\} &= (q_e^* - q_e)_+ / \delta. \end{aligned}$$

The joint distribution is irrelevant, but for definiteness we let them be independent. Then we define the coupled pair $(\mathbf{e}, \mathbf{e}^*)$ to have the values

$$\begin{aligned} &(\check{e}, \check{e}) \text{ if } \omega > \delta; \\ &(\check{e}_+, \check{e}_-) \text{ if } \omega \leq \delta. \end{aligned} \quad (54)$$

Then \mathbf{e} has distribution q , \mathbf{e}^* has distribution q^* , and $\mathbf{e} = \mathbf{e}^*$ with probability $1 - \delta$. This δ is called the *total-variation distance* between q and q^* .

We write \mathbb{E}_P for the expectation with respect to the distribution that makes e_0, \dots, e_m a stationary Markov chain with transition matrix P . Define $\nu^{(\epsilon)}$ to be the stationary distribution corresponding to $P^{(\epsilon)}$, and define $\tilde{P}^{(\epsilon)}$ to be the time-reversed chain of $P^{(\epsilon)}$. We define

$$g(m, \epsilon; u) := \mathbb{E}_{P^{(\epsilon)}} \left[\log \frac{\|X_{\mathbf{e}_m} X_{\mathbf{e}_{m-1}} \cdots X_{\mathbf{e}_0} u\|}{\|X_{\mathbf{e}_{m-1}} \cdots X_{\mathbf{e}_0} u\|} \right] - \mathbb{E}_P \left[\log \frac{\|X_{\mathbf{e}_m} X_{\mathbf{e}_{m-1}} \cdots X_{\mathbf{e}_0} u\|}{\|X_{\mathbf{e}_{m-1}} \cdots X_{\mathbf{e}_0} u\|} \right].$$

By the time-reversal property,

$$g(m, \epsilon; u) = \mathbb{E}_{\tilde{P}^{(\epsilon)}} \left[\log \frac{\|X_{\mathbf{e}_0} X_{\mathbf{e}_1} \cdots X_{\mathbf{e}_m} u\|}{\|X_{\mathbf{e}_1} \cdots X_{\mathbf{e}_m} u\|} \right] - \mathbb{E}_{\tilde{P}} \left[\log \frac{\|X_{\mathbf{e}_0} X_{\mathbf{e}_1} \cdots X_{\mathbf{e}_m} u\|}{\|X_{\mathbf{e}_1} \cdots X_{\mathbf{e}_m} u\|} \right].$$

For $\epsilon > 0$ we couple a sequence $\mathbf{e}_0, \dots, \mathbf{e}_m$ selected from the distribution \tilde{P} to a sequence $\mathbf{e}_0^{(\epsilon)}, \dots, \mathbf{e}_m^{(\epsilon)}$ selected from the distribution $\tilde{P}^{(\epsilon)}$ as follows: We start by choosing $(\mathbf{e}_0, \mathbf{e}_0^{(\epsilon)})$ according to the standard coupling of $(\nu, \nu^{(\epsilon)})$. Assume now that we have produced sequences of length i , ending in \mathbf{e}_{i-1} and $\mathbf{e}_{i-1}^{(\epsilon)}$. We then produce $(\mathbf{e}_i, \mathbf{e}_i^{(\epsilon)})$ according to the standard coupling of row \mathbf{e}_{i-1} of \tilde{P} to row $\mathbf{e}_{i-1}^{(\epsilon)}$ of $\tilde{P}^{(\epsilon)}$. (To simplify the typography in some places we use $e(i)$ and $e^{(\epsilon)}(i)$ interchangeably with e_i and $e_i^{(\epsilon)}$.)

Let $\delta = \delta(\epsilon)$ be the maximum of the total variation distance between ν and $\nu^{(\epsilon)}$, and all of the pairs of rows. It is easy to see that there is a constant c such that $\delta \leq c\epsilon$ for ϵ sufficiently small. Define $\omega_1, \omega_2, \dots$ to be an i.i.d. sequence of uniform random variables on $[0, 1]$, and two sequences of random times as follows: $T_0 := S_0 := -1$, and

$$\begin{aligned} T_{i+1} &= \min\{t > S_i : \omega_t \leq \delta\}, \\ S_{i+1} &= \min\{t > T_{i+1} : \mathbf{e}_t^{(\epsilon)} = \mathbf{e}_t\}. \end{aligned}$$

Thus, $\mathbf{e}_t^{(\epsilon)} = \mathbf{e}_t$ for all $S_i \leq t < T_{i+1}$. Define for any $u_0 \in \mathcal{U}$ the random vector

$$U_t := \lim_{m \rightarrow \infty} \frac{X_{\mathbf{e}(t)} \cdots X_{\mathbf{e}(t+m)} u_0}{\|X_{\mathbf{e}(t)} \cdots X_{\mathbf{e}(t+m)} u_0\|},$$

and define a version of g conditioned on T_1 and T_2

$$\begin{aligned} g(m, \epsilon; u; T_1, T_2) &:= \mathbb{E}_{\tilde{P}^{(\epsilon)}} \left[\log \frac{\|X_{\mathbf{e}_0} X_{\mathbf{e}_1} \cdots X_{\mathbf{e}_m} u\|}{\|X_{\mathbf{e}_1} \cdots X_{\mathbf{e}_m} u\|} \mid T_1, T_2 \right] \\ &\quad - \mathbb{E}_{\tilde{P}} \left[\log \frac{\|X_{\mathbf{e}_0} X_{\mathbf{e}_1} \cdots X_{\mathbf{e}_m} u\|}{\|X_{\mathbf{e}_1} \cdots X_{\mathbf{e}_m} u\|} \mid T_1, T_2 \right]. \end{aligned}$$

Then for any $u \in \mathcal{U}$,

$$\nabla_W a(P) = a'(0) = \lim_{\epsilon \downarrow 0} \lim_{m \rightarrow \infty} \epsilon^{-1} \mathbb{E} [g(m, \epsilon; u; T_1, T_2)]. \quad (55)$$

We also define

$$\begin{aligned} \gamma(\epsilon; T_1, T_2) &:= \mathbb{E} \left[\log \frac{\|X_{\mathbf{e}^{(\epsilon)}(0)} \cdots X_{\mathbf{e}^{(\epsilon)}(S_1-1)} U_{S_1}\|}{\|X_{\mathbf{e}^{(\epsilon)}(1)} \cdots X_{\mathbf{e}^{(\epsilon)}(S_1-1)} U_{S_1}\|} \right. \\ &\quad \left. - \log \frac{\|X_{\mathbf{e}(0)} X_{\mathbf{e}(1)} \cdots X_{\mathbf{e}(S_1-1)} U_{S_1}\|}{\|X_{\mathbf{e}(1)} \cdots X_{\mathbf{e}(S_1-1)} U_{S_1}\|} \mid T_1, T_2 \right] \end{aligned}$$

We break up these expectations into their portion overlapping three different events:

- (i) $\{T_1 > m\}$;
- (ii) $\{T_2 > m \geq T_1\}$;
- (iii) $\{m \geq T_2\}$.

On the event $\{T_1 > m\}$ we have $g(m, \epsilon; u; T_1, T_2) = 0$, and $T_1 - m$ is geometrically distributed with parameter δ . By (13), γ is bounded by $k_2 r^{T_1 - 1}$.

On the event $\{T_2 > m \geq T_1\}$: We have $\mathbf{e}^{(\epsilon)}(i) = \mathbf{e}(i)$ for $i < T_1$ and for $S_1 \leq i \leq m$. If $S_1 \leq m$,

$$\begin{aligned} U_{S_1} &= X_{\mathbf{e}(S_1)} \cdots X_{\mathbf{e}(m)} U_{m+1} / \|X_{\mathbf{e}(S_1)} \cdots X_{\mathbf{e}(m)} U_{m+1}\| \\ &= X_{\mathbf{e}^{(\epsilon)}(S_1)} \cdots X_{\mathbf{e}^{(\epsilon)}(m)} U_{m+1} / \|X_{\mathbf{e}^{(\epsilon)}(S_1)} \cdots X_{\mathbf{e}^{(\epsilon)}(m)} U_{m+1}\| \end{aligned}$$

Thus we may write

$$\begin{aligned} &|\gamma(\epsilon; T_1, T_2) - g(m, \epsilon; u; T_1, T_2)| \\ &\leq \left| \mathbb{E} \left[\log \frac{\|X_{\mathbf{e}(0)} \cdots X_{\mathbf{e}(T_1-1)} X_{\mathbf{e}^{(\epsilon)}(T_1)} \cdots X_{\mathbf{e}^{(\epsilon)}(m)} \check{U}^{(\epsilon)}\|}{\|X_{\mathbf{e}(1)} \cdots X_{\mathbf{e}(T_1-1)} X_{\mathbf{e}^{(\epsilon)}(T_1)} \cdots X_{\mathbf{e}^{(\epsilon)}(m)} U^{(\epsilon)}\|} \middle| T_1, T_2 \right] \right. \\ &\quad \left. - \mathbb{E} \left[\log \frac{\|X_{\mathbf{e}(0)} \cdots X_{\mathbf{e}(T_1-1)} X_{\mathbf{e}^{(\epsilon)}(T_1)} \cdots X_{\mathbf{e}^{(\epsilon)}(m)} u\|}{\|X_{\mathbf{e}(1)} \cdots X_{\mathbf{e}(T_1-1)} X_{\mathbf{e}^{(\epsilon)}(T_1)} \cdots X_{\mathbf{e}^{(\epsilon)}(m)} u\|} \middle| T_1, T_2 \right] \right| \\ &\quad + \left| \mathbb{E} \left[\log \frac{\|X_{\mathbf{e}(0)} X_{\mathbf{e}(1)} \cdots X_{\mathbf{e}(m)} \check{U}\|}{\|X_{\mathbf{e}(1)} \cdots X_{\mathbf{e}(m)} \check{U}\|} \right. \right. \\ &\quad \left. \left. - \log \frac{\|X_{\mathbf{e}(0)} \cdots X_{\mathbf{e}(T_1-1)} X_{\mathbf{e}(T_1)} \cdots X_{\mathbf{e}(m)} u\|}{\|X_{\mathbf{e}(1)} \cdots X_{\mathbf{e}(T_1-1)} X_{\mathbf{e}(T_1)} \cdots X_{\mathbf{e}(m)} u\|} \middle| T_1, T_2 \right] \right| \\ &\leq 2k_2 r^m, \end{aligned} \tag{56}$$

where $\check{U}^{(\epsilon)} = \check{U} = U_{m+1}$ if $S_1 \leq m$; otherwise

$$\begin{aligned} \check{U}^{(\epsilon)} &= \frac{X_{\mathbf{e}^{(\epsilon)}(m+1)} \cdots X_{\mathbf{e}^{(\epsilon)}(S_1-1)} U_{S_1}}{\|X_{\mathbf{e}^{(\epsilon)}(m+1)} \cdots X_{\mathbf{e}^{(\epsilon)}(S_1-1)} U_{S_1}\|}; \\ \check{U} &= \frac{X_{(m+1)} \cdots X_{(S_1-1)} U_{S_1}}{\|X_{(m+1)} \cdots X_{(S_1-1)} U_{S_1}\|}. \end{aligned}$$

On the event $\{T_2 \leq m\}$: The above approach shows that

$$|\gamma(\epsilon; T_1, T_2) - g(m, \epsilon; u; T_1, T_2)| \leq 2k_2 r^{T_2 - 1}. \tag{57}$$

Combining these bounds, we obtain

$$\begin{aligned} &|\gamma(\epsilon; T_1, T_2) - g(m, \epsilon; u; T_1, T_2)| \\ &\leq k_2 r^{T_1 - 1} \mathbf{1}_{\{T_1 > m\}} + 2k_2 r^m \mathbf{1}_{\{T_1 \leq m\}} + 2k_2 r^{T_2 - 1}. \end{aligned} \tag{58}$$

Taking the expectation with respect to the distribution of T_1 and T_2 , using the fact that T_1 and $T_2 - S_1$ are independent with distribution geometric with parameter δ , we obtain

$$\begin{aligned} & \mathbb{E} \left[\left| \gamma(\epsilon; T_1, T_2) - g(m, \epsilon; u; T_1, T_2) \right| \right] \\ & \leq \frac{k_2}{1-r} r^{m-1} \delta + 2k_2 \delta m r^m + \frac{2k_2}{r^2(1-r)^2} \delta^2. \end{aligned} \quad (59)$$

Since δ is bounded by a constant times $|\epsilon|$, we may find a constant C such that (by the triangle inequality) for all ϵ , positive integers m , and $u \in \mathcal{U}$,

$$\begin{aligned} & \left| \mathbb{E} [\gamma(\epsilon; T_1, T_2)] - \mathbb{E} [g(m, \epsilon; u; T_1, T_2)] \right| \\ & \leq \mathbb{E} \left[\left| \gamma(\epsilon; T_1, T_2) - g(m, \epsilon; u; T_1, T_2) \right| \right] \\ & \leq C(mr^m |\epsilon| + \epsilon^2). \end{aligned} \quad (60)$$

This bound allows us to exchange the limits in (55):

$$\begin{aligned} a'(0) &= \lim_{\epsilon \rightarrow 0} \lim_{m \rightarrow \infty} \epsilon^{-1} \mathbb{E} [g(m, \epsilon; u; T_1, T_2)] \\ &= \lim_{\epsilon \rightarrow 0} \epsilon^{-1} \mathbb{E} [\gamma(\epsilon; T_1, T_2)] \\ &= \lim_{m \rightarrow \infty} \lim_{\epsilon \rightarrow 0} \epsilon^{-1} \mathbb{E} [g(m, \epsilon; u; T_1, T_2)] \\ &= \lim_{m \rightarrow \infty} \frac{d}{d\epsilon} \Big|_{\epsilon=0} \mathbb{E}_{P(\epsilon)} \left[\log \frac{\|X_{\mathbf{e}_m} X_{\mathbf{e}_{m-1}} \cdots X_{\mathbf{e}_0} u\|}{\|X_{\mathbf{e}_{m-1}} \cdots X_{\mathbf{e}_0} u\|} \right] \end{aligned} \quad (61)$$

Now we apply the method of importance sampling. We may assume without loss of generality that $W(e, e') = 0$ whenever $P(e, e') = 0$ (using the analyticity of a , and the fact that the formula (40) is nonsingular on the nonnegative orthant). For any function $Z : \mathcal{M}^{m+1} \rightarrow \mathbb{R}$,

$$\mathbb{E}_{P(\epsilon)} [Z(\mathbf{e}_0, \dots, \mathbf{e}_m)] = \mathbb{E}_P [Z(\mathbf{e}_0, \dots, \mathbf{e}_m) F(\epsilon; \mathbf{e}_0, \dots, \mathbf{e}_m)],$$

where F is the Radon-Nikodym derivative

$$\begin{aligned} F(\epsilon; \mathbf{e}_0, \dots, \mathbf{e}_m) &= \frac{dP(\epsilon)}{dP}(\mathbf{e}_0, \dots, \mathbf{e}_m) \\ &= \frac{\nu_{\mathbf{e}_0}^{(\epsilon)}}{\nu_{\mathbf{e}_0}} \prod_{i=0}^{m-1} \frac{P(\epsilon)(\mathbf{e}_i, \mathbf{e}_{i+1})}{P(\mathbf{e}_i, \mathbf{e}_{i+1})}. \end{aligned}$$

This allows us to rewrite

$$a'(0) = \lim_{m \rightarrow \infty} \frac{d}{d\epsilon} \Big|_{\epsilon=0} \mathbb{E}_P \left[\frac{\nu_{\mathbf{e}_0}^{(\epsilon)}}{\nu_{\mathbf{e}_0}} \prod_{i=0}^{m-1} \frac{P(\epsilon)(\mathbf{e}_i, \mathbf{e}_{i+1})}{P(\mathbf{e}_i, \mathbf{e}_{i+1})} \log \frac{\|X_{\mathbf{e}_m} X_{\mathbf{e}_{m-1}} \cdots X_{\mathbf{e}_0} u\|}{\|X_{\mathbf{e}_{m-1}} \cdots X_{\mathbf{e}_0} u\|} \right] \quad (62)$$

For any fixed m , there is an upper bound on $\epsilon^{-1}(F(\epsilon; \mathbf{e}_0, \dots, \mathbf{e}_m) - 1)$, so we may

move the differentiation inside the expectation, to obtain

$$\begin{aligned}
a'(0) &= \lim_{m \rightarrow \infty} \mathbb{E}_P \left[\frac{d}{d\epsilon} \Big|_{\epsilon=0} \frac{\nu_{\mathbf{e}_0}^{(\epsilon)}}{\nu_{\mathbf{e}_0}} \prod_{i=0}^{m-1} \frac{P^{(\epsilon)}(\mathbf{e}_i, \mathbf{e}_{i+1})}{P(\mathbf{e}_i, \mathbf{e}_{i+1})} \log \frac{\|X_{\mathbf{e}_m} X_{\mathbf{e}_{m-1}} \cdots X_{\mathbf{e}_0} u\|}{\|X_{\mathbf{e}_{m-1}} \cdots X_{\mathbf{e}_0} u\|} \right] \\
&= \lim_{m \rightarrow \infty} \mathbb{E}_P \left[(\nu_{\mathbf{e}_0})^{-1} \frac{d\nu_{\mathbf{e}_0}^{(\epsilon)}}{d\epsilon} \Big|_{\epsilon=0} \log \frac{\|X_{\mathbf{e}_m} X_{\mathbf{e}_{m-1}} \cdots X_{\mathbf{e}_0} u\|}{\|X_{\mathbf{e}_{m-1}} \cdots X_{\mathbf{e}_0} u\|} \right] \\
&\quad + \lim_{m \rightarrow \infty} \sum_{i=0}^{m-1} \mathbb{E}_P \left[\frac{W(\mathbf{e}_i, \mathbf{e}_{i+1})}{P(\mathbf{e}_i, \mathbf{e}_{i+1})} \log \frac{\|X_{\mathbf{e}_m} X_{\mathbf{e}_{m-1}} \cdots X_{\mathbf{e}_0} u\|}{\|X_{\mathbf{e}_{m-1}} \cdots X_{\mathbf{e}_0} u\|} \right]
\end{aligned} \tag{63}$$

The first limit is 0. To see this, rewrite it as a sum over possible values of \mathbf{e}_0 :

$$\begin{aligned}
&\lim_{m \rightarrow \infty} \sum_{e \in \mathcal{M}} \nu_e \mathbb{E}_{P_e} \left[(\nu_e)^{-1} \frac{d\nu_e^{(\epsilon)}}{d\epsilon} \Big|_{\epsilon=0} \log \frac{\|X_{\mathbf{e}_m} X_{\mathbf{e}_{m-1}} \cdots X_e u\|}{\|X_{\mathbf{e}_{m-1}} \cdots X_e u\|} \right] \\
&= \lim_{m \rightarrow \infty} \sum_{e \in \mathcal{M}} \frac{d\nu_e^{(\epsilon)}}{d\epsilon} \mathbb{E}_{P_e} \left[\log \frac{\|X_{\mathbf{e}_m} X_{\mathbf{e}_{m-1}} \cdots X_{\mathbf{e}_1} X_e u\|}{\|X_{\mathbf{e}_{m-1}} \cdots X_{\mathbf{e}_1} X_e u\|} \right]
\end{aligned}$$

Since $\nu^{(\epsilon)}$ is a probability distribution, it must be that $\sum_{\tilde{e} \in \mathcal{M}} \frac{d\nu_{\tilde{e}}^{(\epsilon)}}{d\epsilon} = 0$. Thus, the expression in the limit becomes 0 if we replace the expectation by a constant, independent of e . By Lemma A.2 it follows that the limit is 0.

To compute the other limit, we sum over all possible pairs $(\mathbf{e}_i, \mathbf{e}_{i+1}) = (\tilde{e}, e)$. The summand becomes

$$\sum_{\tilde{e}, e \in \mathcal{M}} \nu(\tilde{e}) W(\tilde{e}, e) \mathbb{E} \left[\log \frac{\|X_{\mathbf{e}_m} X_{\mathbf{e}_{m-1}} \cdots X_{\mathbf{e}_0} u\|}{\|X_{\mathbf{e}_{m-1}} \cdots X_{\mathbf{e}_0} u\|} \Big| \mathbf{e}_i = \tilde{e}, \mathbf{e}_{i+1} = e \right] \tag{64}$$

In order to analyze this, we need to consider the distribution of $\mathbf{e}_0, \dots, \mathbf{e}_m$, conditioned on $\mathbf{e}_i = \tilde{e}$ and $\mathbf{e}_{i+1} = e$. By the Markov property, this splits into two independent Markov chains: $e = \mathbf{e}_{i+1}, \dots, \mathbf{e}_m$ is a Markov chain of length $m - i$, with transition probabilities P and starting point e , while $\tilde{e} = \mathbf{e}_i, \mathbf{e}_{i-1}, \dots, \mathbf{e}_0$ is a Markov chain of length $i + 1$ with transition probabilities \tilde{P} and starting point \tilde{e} . Define two independent infinite sequences $\tilde{\mathbf{e}}_0, \tilde{\mathbf{e}}_1, \dots$ and $\mathbf{e}_0, \mathbf{e}_1, \dots$, which are Markov chains with transitions \tilde{P} and P respectively, beginning in $\tilde{\mathbf{e}}_0 = \tilde{e}$ and $\mathbf{e}_{i+1} = e$. Define for $i \geq 1$, $\check{U}_i(\tilde{e}) := X_{\tilde{e}} X_{\tilde{\mathbf{e}}_1} \cdots X_{\tilde{\mathbf{e}}_i} u$ with $\check{U}_0(\tilde{e}) := \mathbf{1}$, and $\check{V}_i^T(e) := \mathbf{1}^T X_{\mathbf{e}_{i-1}} \cdots X_{\mathbf{e}_1} X_e$ with $\check{V}_0^T(e) := \mathbf{1}^T$. Also define

$$U_i(e) := \frac{\check{U}_i(e)}{\|\check{U}_i(e)\|}, \quad V_i^T(e) := \frac{\check{V}_i^T(e)}{\|\check{V}_i^T(e)\|}.$$

Since $\|u\| = \mathbf{1}^T u$ for any nonnegative column vector u , the expression (63) becomes

$$\begin{aligned}
a'(0) &= \sum_{\tilde{e}, e \in \mathcal{M}} \nu(\tilde{e}) W(\tilde{e}, e) \lim_{m \rightarrow \infty} \left\{ \sum_{i=0}^{m-1} \left(\mathbb{E} [\log \check{V}_{i+1}^T(e) \check{U}_{m-i}(\tilde{e})] \right. \right. \\
&\quad \left. \left. - \mathbb{E} [\log \check{V}_i^T(e) \check{U}_{m-i}(\tilde{e})] \right) \right\} \\
&= \sum_{\tilde{e}, e \in \mathcal{M}} \nu(\tilde{e}) W(\tilde{e}, e) \lim_{m \rightarrow \infty} \left\{ \sum_{i=0}^{m-1} \left(\mathbb{E} [\log \|\check{V}_{i+1}^T(e)\|] - \mathbb{E} [\log \|\check{V}_i^T(e)\|] \right. \right. \\
&\quad \left. \left. + \mathbb{E} [\log V_{i+1}^T(e) U_i(\tilde{e})] - \mathbb{E} [\log V_i^T(e) U_i(\tilde{e})] \right) \right\} \tag{65} \\
&= \sum_{\tilde{e} \in \mathcal{M}} \nu(\tilde{e}) \lim_{m \rightarrow \infty} \sum_{e \in \mathcal{M}} W(\tilde{e}, e) \mathbb{E} [\log \|\check{V}_m^T(e)\|] \\
&\quad + \sum_{e, \tilde{e} \in \mathcal{M}} \nu(\tilde{e}) W(\tilde{e}, e) \lim_{m \rightarrow \infty} \sum_{i=0}^{m-1} \mathbb{E} \left[\log \frac{V_{i+1}^T(e) U_{m-i}(\tilde{e})}{V_i^T(e) U_{m-i}(\tilde{e})} \right]
\end{aligned}$$

In the last line we have used the fact that $\sum_{e \in \mathcal{M}} W(\tilde{e}, e) = 0$, which means that $\sum_{e \in \mathcal{M}} W(\tilde{e}, e) \mathbb{E} [\log \|\check{V}_0^T(e)\|] = 0$ as well, since $\check{V}_0^T(e) = \mathbf{1}^T$ is independent of e . The same reasoning implies that if we define $\check{V}_i^T(\nu)$ to be the version of \check{V}_i^T started in the stationary distribution — for instance, starting from realizations of $\check{V}_i^T(e)$, define $\check{V}_i^T(\nu)$ to be equal to $\check{V}_i^T(e)$ with probability ν_e — then $\sum_{e \in \mathcal{M}} W(\tilde{e}, e) \mathbb{E} [\log \|\check{V}_m^T(\nu)\|] = 0$. The first term on the right-hand side of (65) may then be written as

$$\sum_{\tilde{e}, e \in \mathcal{M}} \nu(\tilde{e}) W(\tilde{e}, e) \lim_{m \rightarrow \infty} \mathbb{E} \left[\log \frac{\|\check{V}_m^T(e)\|}{\|\check{V}_m^T(\nu)\|} \right] = \sum_{\tilde{e}, e \in \mathcal{M}} \nu(\tilde{e}) W(\tilde{e}, e) \zeta_e. \tag{66}$$

To compute the second term, we note that $U_\infty(\tilde{e}) := \lim_{i \rightarrow \infty} U_i(\tilde{e})$ exists, with distribution $\pi_{\tilde{e}}$, and $\rho(U_i(\tilde{e}), U_\infty(\tilde{e})) \leq k_2 r^i$; similarly, $V_\infty^T(e) := \lim_{i \rightarrow \infty} V_i^T(e)$ exists, with distribution $\tilde{\pi}_e$, and $\rho(V_i(e), V_{i+1}(e)) \leq k_2 r^i$. Thus

$$\left| \log V_{i+1}^T(e) U_{m-i}(\tilde{e}) - \log V_i^T(e) U_{m-i}(\tilde{e}) \right| \leq \rho(V_i(e), V_{i+1}(e)) \leq k_2 r^i,$$

We break up the sum on the right-hand side of (65) into three pieces:

$$\begin{aligned}
&\sum_{0 \leq i \leq m-1} \mathbb{E} [\log V_{i+1}^T(e) U_{m-i}(\tilde{e}) - \log V_i^T(e) U_{m-i}(\tilde{e})] \\
&= \sum_{0 \leq i \leq m/2} \mathbb{E} [\log V_{i+1}^T(e) U_\infty(\tilde{e}) - \log V_i^T(e) U_\infty(\tilde{e})] \\
&\quad + \sum_{0 \leq i \leq m/2} \mathbb{E} \left[\log \frac{V_{i+1}^T(e) U_{m-i}(\tilde{e})}{V_{i+1}^T(e) U_\infty(\tilde{e})} - \log \frac{V_i^T(e) U_{m-i}(\tilde{e})}{V_i^T(e) U_\infty(\tilde{e})} \right] \\
&\quad + \sum_{m/2 < i \leq m-1} \mathbb{E} [\log V_{i+1}^T(e) U_{m-i}(\tilde{e}) - \log V_i^T(e) U_{m-i}(\tilde{e})].
\end{aligned}$$

The first sum telescopes to

$$\mathbb{E} [\log V_{1+m/2}^T(e) U(\tilde{e}) - \log V_0^T(e) U(\tilde{e})] = \mathbb{E} [\log V_{1+m/2}^T(e) U(\tilde{e})] - \log K,$$

applying the fact that $V_0^T = \mathbf{1}^T/K$, so that $V_0^T(e)U(\tilde{e}) = \|U(\tilde{e})\|/K = 1/K$. Applying (9), the second and third sums are bounded by

$$\sum_{0 \leq i \leq m/2} 2k_2 r^{m-i} + \sum_{m/2 < i \leq m-1} k_2 r^i \leq \frac{3k_2 r^{m/2}}{1-r}.$$

Thus

$$\lim_{m \rightarrow \infty} \sum_{i=0}^{m-1} \mathbb{E} \left[\log \frac{V_{i+1}^T(e)U_{m-i}(\tilde{e})}{V_i^T(e)U_{m-i}(\tilde{e})} \right] = \mathbb{E} [V^T(e)U(\tilde{e})], \quad (67)$$

completing the proof of Theorem 7.1.

Lemma A.2. *For any $u, \check{u} \in \mathcal{U}$ and $e, e^* \in \mathcal{M}$, if we let $\mathbf{e}_0, \mathbf{e}_1, \dots$ and $\mathbf{e}_0^*, \mathbf{e}_1^*, \dots$ be realisations of the Markov chain P starting at $\mathbf{e}_0 = e$ and $\mathbf{e}_0^* = e^*$ respectively. Then*

$$\left| \mathbb{E} \left[\log \frac{\|X_{\mathbf{e}_{m+1}} \cdots X_{\mathbf{e}_0} u\|}{\|X_{\mathbf{e}_m} \cdots X_{\mathbf{e}_0} u\|} \right] - \mathbb{E} \left[\log \frac{\|X_{\mathbf{e}_{m+1}^*} \cdots X_{\mathbf{e}_0^*} \check{u}\|}{\|X_{\mathbf{e}_m^*} \cdots X_{\mathbf{e}_0^*} \check{u}\|} \right] \right| \leq \frac{k_2 D}{1-r} (m+1)(\xi \vee r)^m, \quad (68)$$

where ξ and D are the constants that satisfy (38).

Proof. Using the maximal coupling, we create coupled versions of $(\mathbf{e}_i, \mathbf{e}_i^*)$, such that the coupling time τ satisfies

$$\mathbb{P}\{\tau \geq t\} \leq \|\nu - P^t(e, \cdot)\| + \|\nu - P^t(e^*, \cdot)\| \leq 2D\xi^t.$$

Define

$$u_\tau := \frac{X_{\mathbf{e}_{\tau-1}} \cdots X_{\mathbf{e}_0} u}{\|X_{\mathbf{e}_{\tau-1}} \cdots X_{\mathbf{e}_0} u\|}, \quad \check{u}_\tau := \frac{X_{\mathbf{e}_{\tau-1}^*} \cdots X_{\mathbf{e}_0^*} \check{u}}{\|X_{\mathbf{e}_{\tau-1}^*} \cdots X_{\mathbf{e}_0^*} \check{u}\|}.$$

Then by the bound (13),

$$\begin{aligned} & \left| \mathbb{E} \left[\log \frac{\|X_{\mathbf{e}_{m+1}} \cdots X_{\mathbf{e}_0} u\|}{\|X_{\mathbf{e}_m} \cdots X_{\mathbf{e}_0} u\|} \right] - \mathbb{E} \left[\log \frac{\|X_{\mathbf{e}_{m+1}^*} \cdots X_{\mathbf{e}_0^*} \check{u}\|}{\|X_{\mathbf{e}_m^*} \cdots X_{\mathbf{e}_0^*} \check{u}\|} \right] \right| \\ & \leq \mathbb{E} \left[\left| \log \frac{\|X_{\mathbf{e}_{m+1}} \cdots X_{\mathbf{e}_0} u\|}{\|X_{\mathbf{e}_m} \cdots X_{\mathbf{e}_0} u\|} - \log \frac{\|X_{\mathbf{e}_{m+1}^*} \cdots X_{\mathbf{e}_0^*} \check{u}\|}{\|X_{\mathbf{e}_m^*} \cdots X_{\mathbf{e}_0^*} \check{u}\|} \right| \right] \\ & = \mathbb{E} \left[\left| \log \frac{\|X_{\mathbf{e}_{m+1}} \cdots X_{\mathbf{e}_\tau} u_\tau\|}{\|X_{\mathbf{e}_m} \cdots X_{\mathbf{e}_0} u\|} - \log \frac{\|X_{\mathbf{e}_{m+1}^*} \cdots X_{\mathbf{e}_\tau} \check{u}_\tau\|}{\|X_{\mathbf{e}_m^*} \cdots X_{\mathbf{e}_0^*} \check{u}\|} \right| \right] \\ & \leq \mathbb{E} [k_2 r^{m-\tau}] \\ & \leq \frac{k_2 D}{1-r} \sum_{t=0}^m r^{m-t} \xi^t \\ & \leq \frac{k_2 D}{1-r} (m+1)(\xi \vee r)^m. \end{aligned}$$

□

A.5 A version of the Furstenberg-Kesten Theorem

Theorem A.3 (Furstenberg-Kesten). *The limit defined in (2) exists almost surely, and is deterministic, given by (46).*

Proof. Let $\mathbf{e}_1, \mathbf{e}_2, \dots$ be a stationary Markov chain with transition matrix P . If we choose $N_0 \in \mathcal{U}$, this induces a Markov chain $Y_t := (N_{t-1}/\|N_{t-1}\|, \mathbf{e}_t)$ with state space $\mathcal{U} \times \mathcal{M}$. We extend the Hilbert metric to $\mathcal{U} \times \mathcal{M}$ by $\rho((u, e), (u_*, e_*)) = \rho(u, u_*) + \mathbf{1}_{\{e \neq e_*\}}$.

If we define the function $F : \mathcal{U} \times \mathcal{M} \rightarrow \mathbb{R}$ by

$$F(u, m) := \log \|X_m u\|,$$

then

$$\begin{aligned} \log \|N_t\| &= \sum_{i=1}^t \log \frac{\|N_i\|}{\|N_{i-1}\|} \\ &= \sum_{i=1}^t \log \left\| X_{\mathbf{e}_i} \frac{N_{i-1}}{\|N_{i-1}\|} \right\| \\ &= \sum_{i=1}^t F(Y_i). \end{aligned}$$

If (Y_i) were a chain on a finite space, we would invoke the law of large numbers for Markov chains, also known as the pathwise Individual Ergodic Theorem (*cf.* Theorem 1.10.2 of Norris (1998)) to see that

$$\lim_{t \rightarrow \infty} t^{-1} \log \|N_t\| = \lim_{t \rightarrow \infty} t^{-1} \sum_{i=1}^t F(Y_i) \text{ exists and equals } \pi[F] := \sum \pi_y F(y),$$

where π is the unique stationary distribution. The same would be true for a positive recurrent Markov chain on a countable state space.

On a more general state space, the pathwise ergodic theorem still holds as long as the Markov chain is uniquely ergodic; that is, it has a unique stationary distribution π (see Theorem 6.1 of Hernández-Lerma and Lasserre (1998), or Chapter 6 of Walters (1982)), and π is ergodic.

We first show the existence of π . Choose $u_0 \in \mathcal{U}$, and let $\dots, \mathbf{e}_{-1}, \mathbf{e}_0, \mathbf{e}_1, \mathbf{e}_2, \dots$ be a stationary Markov chain on \mathcal{M} with transitions P , infinite in both directions (as in section 3.2). Then

$$\rho(U_{-1,-n}, U_{-1,-n-1}) = \rho(X_{\mathbf{e}_{-1}} \cdots X_{\mathbf{e}_{-n}} u_0, X_{\mathbf{e}_{-1}} \cdots X_{\mathbf{e}_{-n}} X_{\mathbf{e}_{-n-1}} u_0) \leq k_2 r^n,$$

implying that $(U_{-1,-n})_{n=1}^\infty$ is always a Cauchy sequence, hence converges to a random variable $U_{-1,-\infty} \in \mathcal{U}$. Define π to be the distribution of the pair $(U_{-1,-\infty}, \mathbf{e}_0)$.

Now, conditioned on starting at $Y_0 = (U_{-1,-\infty}, \mathbf{e}_0)$, the next step Y_1 has the distribution of $(X_{\mathbf{e}_0} U_{-1,-\infty} / \|X_{\mathbf{e}_0} U_{-1,-\infty}\|, \mathbf{e}_1)$. Notice that

$$\frac{X_{\mathbf{e}_0} U_{-1,-\infty}}{\|X_{\mathbf{e}_0} U_{-1,-\infty}\|} = U_{0,-\infty}$$

has the same distribution as $U_{-1,-\infty}$, so that π is a stationary distribution for (Y_t) .

It remains to show that π is uniquely ergodic. Suppose that we start the chain in an alternative distribution μ . Define $Y_0 = (U, \mathbf{e}_1)$ to have distribution π and $Y_0^* = (U^*, \mathbf{e}_1^*)$ to have distribution μ . Let $\mathbf{e}_1, \mathbf{e}_2, \dots$ and $\mathbf{e}_1^*, \mathbf{e}_2^*, \dots$ be coupled versions of the environment chain, started in \mathbf{e}_1 and \mathbf{e}_1^* respectively. For definiteness, we say that they are independent up to the first time τ when $\mathbf{e}_\tau = \mathbf{e}_\tau^*$ — and then $\mathbf{e}_{\tau+i} = \mathbf{e}_{\tau+i}^*$ for $i \geq 0$. Then

$$Y_m := \left(\frac{X_{\mathbf{e}_m} \cdots X_{\mathbf{e}_1} U}{\|X_{\mathbf{e}_m} \cdots X_{\mathbf{e}_1} U\|}, \mathbf{e}_{m+1} \right) \text{ and } Y_m^* := \left(\frac{X_{\mathbf{e}_m^*} \cdots X_{\mathbf{e}_1^*} U^*}{\|X_{\mathbf{e}_m^*} \cdots X_{\mathbf{e}_1^*} U^*\|}, \mathbf{e}_{m+1}^* \right)$$

are both realisations of the Markov chain Y , with initial distribution of π and μ respectively. We write μ_m for the distribution of Y_m^* . We see that

$$\rho(Y_m, Y_m^*) \leq k_2 r^{m-\tau} + \mathbf{1}_{\{\tau > m\}} \rightarrow_P 0 \text{ as } m \rightarrow \infty.$$

Thus, $\mu_m \rightarrow \pi$ weakly. By Theorem 6.12 of Walters (1982), it follows that π is ergodic (in fact, strong mixing). Furthermore, if μ were an invariant distribution then $\mu_m = \mu$, so the convergence implies $\mu = \pi$. Therefore, π is uniquely ergodic. \square

References

- Asmussen, S. and Glynn, P. (2007). *Stochastic simulation: Algorithms and analysis*. Springer Verlag.
- Barnsley, M. F. and Elton, J. H. (1988). A new class of Markov processes for image encoding. *Annals of Applied Probability*, 20:14–32.
- Boyce, M., Haridas, C., and Lee, C. (2006). Demography in an increasingly variable world. *Trends in Ecology & Evolution*, 21(3):141–148.
- Brémaud, P. (1992). Maximal coupling and rare perturbation sensitivity analysis. *Queueing Systems*, 11:307–33.
- Bushell, P. J. (1973). Hilbert’s metric and positive contraction mappings in a Banach space. *Archive for Rational Mechanics and Analysis*, 52(4):330–338.
- Caswell, H. (2001). *Matrix population models: construction, analysis and interpretation*. Sinauer Associates inc. Publishers, Sunderland, Mass., 2nd edition.
- Cohen, J. (1977). Ergodicity of age structure in populations with Markovian vital rates. II. General states. *Advances in Applied Probability*, pages 18–37.
- Ellner, S. and Rees, M. (2007). Stochastic stable population growth in integral projection models: theory and application. *Journal of Mathematical Biology*, 54(2):227–256.
- Elton, J. H. (1990). A multiplicative ergodic theorem for Lipschitz maps. *Stochastic Processes Appl.*, 34(1):39–47.
- Furstenberg, H. and Kesten, H. (1960a). Products of random matrices. *The Annals of Mathematical Statistics*, 31(2):457–469.

- Furstenberg, H. and Kesten, H. (1960b). Products of random matrices. *The Annals of Mathematical Statistics*, 31(2):457–69.
- Golubitsky, M., Keeler, E. B., and Rothschild, M. (1975). Convergence of the age structure: applications of the projective metric. *Theoretical population biology*, 7(1):84.
- Griffeath, D. (1975). A maximal coupling for Markov chains. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 31:95–106.
- Grimmett, G. and Stirzaker, D. (2001). *Probability and random processes*. Oxford University Press, USA.
- Häggström, O. (2002). *Finite Markov Chains and Algorithmic Applications*. Cambridge University Press, Cambridge.
- Haridas, C. V. and Tuljapurkar, S. (2005). Elasticities in variable environments: properties and implications. *Am Nat*, 166(4):481–95.
- Hernández-Lerma, O. and Lasserre, J. B. (1998). Ergodic theorems and ergodic decomposition for Markov chains. *Acta Applicandae Mathematicae*, 54:99–119.
- Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, pages 13–30.
- Kendall, W. S. (2005). Notes on perfect simulation. In F., L. and J-S., W., editors, *Markov Chain Monte Carlo: Innovations and Applications*, pages 93–146. World Scientific.
- Lande, R., Engen, S., and Saether, B.-E. (2003). *Stochastic Population Dynamics in Ecology and Conservation*. Oxford University Press, Oxford.
- Lange, K. (1979). On cohen’s stochastic generalization of the strong ergodic theorem of demography. *Journal of Applied Probability*, pages 496–504.
- Lange, K. and Holmes, W. (1981). Stochastic stable population growth. *Journal of Applied Probability*, 18(2):325–334.
- Lee, R. D. and Tuljapurkar, S. (1994). Stochastic population forecasts for the united states: Beyond high, medium, and low. *Journal of the American Statistical Association*, 89(428):1175–1189.
- McNamara, J. (1997). Optimal life histories for structured populations in fluctuating environments. *Theoretical Population Biology*, 51(2):94–108.
- Metz, J., Nisbet, R., and Geritz, S. (1992). How should we define ‘fitness’ for general ecological scenarios? *Trends in Ecology & Evolution*, 7(6):198–202.
- Morris, W. and Doak, D. (2002). *Quantitative conservation biology*. Sinauer Associates, inc. Publishers, Sunderland, Massachusetts, USA.
- Morris, W., Pfister, C., Tuljapurkar, S., Haridas, C., Boggs, C., Boyce, M., Bruna, E., Church, D., Coulson, T., Doak, D., et al. (2008). Longevity can buffer plant and animal populations against changing climatic variability. *Ecology*, 89(1):19–25.

- Norris, J. (1998). *Markov Chains*. Cambridge University Press.
- Peres, Y. (1992). Domains of analytic continuation for the top lyapunov exponent. *Annales de l'Institut Henri Poincaré. Probabilités et Statistiques*, 28(1):131–48.
- Pitman, J. (1976). On coupling of markov chains. *Probability Theory and Related Fields*, 35(4):315–322.
- Roberts, G. O. and Rosenthal, J. S. (2004). General state space Markov chains and MCMC algorithms. *Probab. Surv.*, 1:20–71 (electronic).
- Seneta, E. (1981). *Non-negative matrices and Markov chains*. Springer Verlag.
- Steinsaltz, D. (1999). Locally contractive iterated function systems. *Annals of Probability*, pages 1952–1979.
- Tuljapurkar, S. (1982). Population dynamics in variable environments. III. Evolutionary dynamics of r-selection. *Theoretical Population Biology*, 21(1):141–165.
- Tuljapurkar, S. (1990). *Population dynamics in variable environments*. Lecture notes in biomathematics, 85. Springer-Verlag, New York.
- Tuljapurkar, S., Horvitz, C. C., and Pascarella, J. B. (2003). The many growth rates and elasticities of populations in random environments. *The American Naturalist*, 162(4):489–502.
- Tuljapurkar, S. D. and Orzack, S. H. (1980). Population dynamics in variable environments. I. Long-run growth rates and extinction. *Theoretical Population Biology*, 18:314–342.
- Walters, P. (1982). *An introduction to ergodic theory*. Springer Verlag, New York, Heidelberg, Berlin.