

LIE ALGEBRA SOLUTION OF POPULATION MODELS BASED ON TIME-INHOMOGENEOUS MARKOV CHAINS

THOMAS HOUSE,* *University of Warwick*

Abstract

Many natural populations are well modelled through time-inhomogeneous stochastic processes. Such processes have been analysed in the physical sciences using a method based on Lie algebras, but this methodology is not widely used for models with ecological, medical, and social applications. In this paper we present the Lie algebraic method, and apply it to three biologically well-motivated examples. The result of this is a solution form that is often highly computationally advantageous.

Keywords: Lie algebra; Markov chain; time inhomogeneous; epidemic; birth–death process

2010 Mathematics Subject Classification: Primary 60J22

Secondary 17B80; 92D25

1. Introduction

Stochastic models based on Markov chains are important in many ecological, medical, and social contexts. In these contexts, where populations are modelled, there are often external influences that act on the system in a manner that varies over time, leading to a time-inhomogeneous Markov chain model and corresponding technical difficulties for analysis [10].

Wei and Norman [14] proposed a method for dealing analytically with time-inhomogeneous Markov chains, based on Lie algebraic methods. The idea of combining Lie algebras and symmetry considerations with Markov chains has continued to attract theoretical interest in a variety of contexts [6], [9], [12], [15].

At the same time, there is a more applied desire to have numerically efficient methods to analyse Markov chain population models, one option for which is the use of matrix exponentials [8], [10]. The aim of this paper is to explain how Lie algebraic methods can be used to derive matrix exponential solutions to time-inhomogeneous Markov chains that are applicable to population modelling. In contrast to other applications, symmetries of these systems are not a guide to the appropriate Lie algebra to use in solution of population models; a certain amount of trial and error is necessary. The focus of this paper is therefore on three examples in population modelling where it is possible to define an appropriate Lie algebra, and a discussion of the potential benefits of doing so.

2. Methodology

2.1. Lie algebras

In general, a Lie algebra over a field F is an F -vector space V , together with a bilinear map called the *Lie bracket*. Elements of the vector space are written $u, v, w \in V$. The Lie bracket

Received 5 July 2011; revision received 23 November 2011.

* Postal address: Mathematics Institute, University of Warwick, Coventry, CV4 7AL, UK.

Email address: t.a.house@warwick.ac.uk

is written $[u, v] \in V$ and obeys

$$[u, u] = 0, \quad [u, [v, w]] + [v, [w, u]] + [w, [u, v]] = 0. \tag{1}$$

We will be interested in the vector space $GL(n, \mathbb{R})$, i.e. the set of real-valued $n \times n$ matrices, and will define the Lie bracket for $X, Y \in GL(n, \mathbb{R})$ through the *commutator*

$$[X, Y] := XY - YX, \tag{2}$$

which can be readily seen to satisfy (1). It is often convenient to define an adjoint endomorphism operator, ad , to represent the Lie bracket:

$$(\text{ad}X)Y := [X, Y]; \tag{3}$$

so that multiple applications of the Lie bracket can be concisely written as, e.g. $(\text{ad}X)^2Y = [X, [X, Y]]$.

2.2. Time-inhomogeneous Markov processes

Suppose that $p(t)$ is a probability vector, i.e. a vector with values $p_n(t) \geq 0$ summing to unity that represent the probability that an integer stochastic variable takes the value n at time t . We consider models in which the evolution of these probabilities over time is given by

$$\frac{dp}{dt} = H(t)p(t), \tag{4}$$

where $H(t)$ is a time-dependent matrix such that at any time t its off-diagonal elements are positive and its column sums are 0. This defines a time-inhomogeneous continuous-time Markov chain. For some special cases, analytic solutions can be obtained. But, in general, where the state space of the Markov chain is finite, numerical algorithms exist that calculate $p(t)$ by making use of expansions such as

$$p(t + \delta t) = H(t)p(t)\delta t + O(\delta t^2), \tag{5}$$

and accumulating a sufficient number of δt steps to reach $p(t)$ from $p(0)$. Methods based on this direct integration will therefore calculate $p(t)$ in $O(t)$ operations.

2.3. The method of Wei and Norman

In this subsection we recall the methodology of Wei and Norman [14], which is applicable to equations of the form (4). The first step is to look for a decomposition of $H(t)$,

$$H(t) = \sum_{i=1}^m a_i(t)H_i, \tag{6}$$

where the H_i are linearly independent matrices obeying

$$[H_i, H_j] = H_i H_j - H_j H_i = \sum_k \xi_{ij}^k H_k \tag{7}$$

for (in our case real-valued) scalars ξ_{ij}^k . Given such matrices, we can form a vector space

$$V = \text{span}\{H_i\}_{i=1}^m \subseteq GL(n, \mathbb{R})$$

such that, for a Lie bracket as defined in (2), we have $[X, Y] \in V$ for all $X, Y \in V$. This closure under the action of the Lie bracket can be used to look for solutions of the form

$$p(t) = e^{g_1(t)H_1} \dots e^{g_m(t)H_m} p(0) =: U(t)p(0), \tag{8}$$

where matrix exponentiation is defined through the power series in the standard way. Using the ad operator as defined in (3), the Baker–Campbell–Hausdorff formula is

$$e^X Y e^{-X} = e^{(\text{ad}X)} Y, \tag{9}$$

which will enable us to derive the solution form advertised. Substituting (6) and (8) into (4) then gives

$$\frac{dp}{dt} = \sum_{i=1}^m a_i(t) H_i U(t) p(0) = \sum_{i=1}^m \dot{g}_i(t) \left(\prod_{j=1}^{i-1} e^{g_j(t)H_j} \right) H_i \left(\prod_{j=i}^m e^{g_j(t)H_j} \right) p(0).$$

Since this expression holds for any $p(0)$, we can equate the operators acting on $p(0)$, postmultiply by U^{-1} , and repeatedly apply (9) to obtain

$$\sum_{i=1}^m a_i(t) H_i = \sum_{i=1}^m \dot{g}_i(t) \left(\prod_{j=1}^{i-1} e^{g_j(t)(\text{ad}H_j)} \right) H_i. \tag{10}$$

The precise solution to this equation will depend on the constants ξ in (7); however, since the H_i are chosen to be linearly independent, the terms in (10) in front of the same basis matrix can be equated, leading to a set of ordinary differential equations (ODEs) for the $g_i(t)$.

The usefulness of this method therefore depends on whether appropriate H_i matrices can be defined, so that the equations that must be solved for $g_i(t)$ are not excessively complex. But in the event that $g_i(t)$ can be calculated in $O(1)$ rather than, say, $O(t)$ —which will typically be the case if an analytic result is obtained—then the computation of $p(t)$ can be achieved in $O(1)$ rather than $O(t)$ through the numerical calculation of the matrix exponentials in (8). Such enhanced computational tractability of stochastic models clearly has benefits in the application of probability theory to statistical inference, where the speed of evaluation of likelihoods is highly important.

3. Examples

The primary difficulty in applying the method above to population models is finding an appropriate expansion of the form (6), since the systems involved are not obviously symmetric. We now turn to three examples where an appropriate expansion can be found. In two cases, special initial conditions give analytic results that can be checked against other methods to confirm the soundness of the approach; and in the remaining case a significant numerical benefit is observed compared to direct integration.

3.1. A birth–death process

Suppose that we have a time-inhomogeneous birth–death process, characterised by a stochastic variable $N(t) \geq 0$, taking integer values n , and transition rates

$$n \rightarrow n + 1 \quad \text{at rate } b(t), \quad n \rightarrow n - 1 \quad \text{at rate } nd(t). \tag{11}$$

A biological interpretation of this process would be the survival of juvenile animals, introduced to an inhospitable region by seasonal breeding happening at another site, and dying at a rate that depends on the climate. Defining components of a vector $p_n(t) = \Pr(N(t) = n)$, the Kolmogorov equation for this process is

$$\frac{dp}{dt} = (b(t)(R - \mathbb{I}) + d(t)(L - M))p. \tag{12}$$

The matrices involved are countably infinite in dimension, and are defined implicitly by (11) and (12). It is also possible to write explicit definitions in terms of the Kronecker delta:

$$(\mathbb{I})_{n,k} = \delta_{n,k}, \quad (R)_{n,k} = \delta_{n,k+1}, \quad (L)_{n,k} = (k - 1)\delta_{n,k-1}, \quad (M)_{n,k} = k\delta_{n,k}.$$

Clearly, the identity matrix commutes with everything (i.e. $[\mathbb{I}, X] = 0$ for any X) while the other matrices obey

$$[L, R] = \mathbb{I}, \quad [M, R] = R, \quad [L, M] = L. \tag{13}$$

We then look for solutions of the form

$$p(t) = e^{g_1(t)\mathbb{I}} e^{g_2(t)R} e^{g_3(t)L} e^{g_4(t)M} p(0), \tag{14}$$

noting that $e^{g_1(t)\mathbb{I}} = e^{g_1(t)}$. Making use of result (10) together with the algebra in (13) gives

$$g_1(t) = -e^{\mathcal{D}(t)} \int_0^t b(u)e^{\mathcal{D}(u)} du, \tag{15}$$

$$g_2(t) = e^{\mathcal{D}(t)} \int_0^t b(u)e^{\mathcal{D}(u)} du, \tag{16}$$

$$g_3(t) = e^{\mathcal{D}(t)} - 1,$$

$$g_4(t) = -\mathcal{D}(t),$$

where

$$\mathcal{D}(t) := \int_0^t d(u) du.$$

This provides a solution to the original model, but one that is much simpler if we assume the initial condition $N(0) = 0$, in which case

$$p_n(t) = e^{g_1(t)} \frac{(g_2(t))^n}{n!} \tag{17}$$

for g_1 and g_2 as in (15) and (16). In this way an infinite-dimensional time-inhomogeneous Markov chain is reduced to carrying out the two integrals in (15) and (16). It is worth comparing this to the ‘textbook’ method for dealing with time-inhomogeneous Markov chains, which is to derive an expression for the probability generating function (PGF) [4]. This is done by writing down the Kolmogorov equation (12) in component form and substituting in the definition of the PGF, $G(s, t) := \sum_n s^n p_n(t)$. This gives a partial differential equation for the PGF of

$$\frac{\partial G}{\partial t} = (b(t)s - d(t))(s - 1) \frac{\partial G}{\partial s},$$

which, given the initial condition $N(0) = 0$, gives

$$G(s, t) = \exp\left[(s - 1)e^{-\mathcal{D}(t)} \int_0^t b(u)e^{\mathcal{D}(u)} du\right]. \tag{18}$$

Equation (18) yields the same solution as (17) through the standard relation

$$p_n = \frac{1}{n!} \left. \frac{\partial^n G}{\partial s^n} \right|_{s=0}.$$

The effort in deriving solution (17) through the PGF and Lie algebraic methods is therefore roughly similar; however, the intermediate results obtained in each method are likely to be useful in different contexts. For example, the PGF in (18) is likely to be the easiest way to derive moments of the process; while the matrix exponential form (14) is likely to be useful if the derivative of the solution with respect to a parameter of the model is required [15].

3.2. Epidemic surveillance

Consider the following situation. An epidemic is in progress in a population such that individuals are either susceptible to infection, infectious, or recovered and immune. Surveillance of the epidemic is carried out by recruitment of individuals at random from the general population (or more realistically through recruitment of individuals in contact with the healthcare system due to noninfectious illness) who are tested and determined to be either susceptible, infectious, or recovered. The epidemic is characterised by a force of infection $\lambda(t)$, which is the rate at which susceptible individuals become infectious and for which there are various parametric forms [7], and also by a recovery rate $\gamma(t)$, which is the rate at which infectious individuals recover. A plausible explicit choice for these functions is to hold $\gamma(t)$ constant, and to take $\lambda(t) = \lambda_0 e^{rt}$, representing the early exponential growth phase that is common to many different epidemics.

As other authors have found, manipulation of more complex Markov chains is simplified by the use of Dirac notation [2], [5], [6], [13]. In this formalism, the probability vector for the model described above is written

$$|p(t)\rangle = \sum_{S,I} \Pr(S, I | N, t) |S, I\rangle,$$

where $\Pr(S, I | N, t)$ is the probability that from a cohort of size N a time t after the start of the epidemic a number S of the cohort is susceptible and a number I is infectious (leaving $N - S - I$ recovered individuals), and $|S, I\rangle$ is a basis vector, linearly independent of any other basis vector with different susceptible and infectious counts. Operators, marked using a hat, $\hat{\mathcal{O}}$, act on basis vectors to give linear combinations of basis vectors. For this system, we need the following operators:

$$\begin{aligned} \hat{S}|S, I\rangle &= S|S, I\rangle, \\ \hat{I}|S, I\rangle &= I|S, I\rangle, \\ \hat{\Delta}|S, I\rangle &= S|S - 1, I\rangle, \\ \hat{\rho}|S, I\rangle &= I|S, I - 1\rangle, \\ \hat{\tau}|S, I\rangle &= S|S - 1, I + 1\rangle. \end{aligned}$$

The action of these operators can be described computationally as follows: \hat{S} returns the number of susceptibles; \hat{I} returns the number of infectives; $\hat{\Delta}$ returns the number of susceptibles and

TABLE 1: Values of $[\hat{X}, \hat{Y}]$ for the epidemic model.

\hat{X}	$[\hat{X}, \hat{S}]$	$[\hat{X}, \hat{I}]$	$[\hat{X}, \hat{\Delta}]$	$[\hat{X}, \hat{\rho}]$	$[\hat{X}, \hat{\tau}]$
\hat{S}	0	0	$-\hat{\Delta}$	0	$-\hat{\tau}$
\hat{I}	0	0	0	$-\hat{\rho}$	$\hat{\tau}$
$\hat{\Delta}$	$\hat{\Delta}$	0	0	0	0
$\hat{\rho}$	0	$\hat{\rho}$	0	0	$\hat{\Delta}$
$\hat{\tau}$	$\hat{\tau}$	$-\hat{\tau}$	0	$-\hat{\Delta}$	0

TABLE 2: Values of $e^{x(\text{ad}\hat{X})}\hat{Y}$ for the epidemic model, for scalar x .

\hat{X}	$e^{x(\text{ad}\hat{X})}\hat{S}$	$e^{x(\text{ad}\hat{X})}\hat{I}$	$e^{x(\text{ad}\hat{X})}\hat{\Delta}$	$e^{x(\text{ad}\hat{X})}\hat{\rho}$	$e^{x(\text{ad}\hat{X})}\hat{\tau}$
\hat{S}	\hat{S}	\hat{I}	$e^{-x}\hat{\Delta}$	$\hat{\rho}$	$e^{-x}\hat{\tau}$
\hat{I}	\hat{S}	\hat{I}	$\hat{\Delta}$	$e^{-x}\hat{\rho}$	$e^x\hat{\tau}$
$\hat{\Delta}$	$\hat{S} + x\hat{\Delta}$	\hat{I}	$\hat{\Delta}$	$\hat{\rho}$	$\hat{\tau}$
$\hat{\rho}$	\hat{S}	$\hat{I} + x\hat{\rho}$	$\hat{\Delta}$	$\hat{\rho}$	$\hat{\tau} + x\hat{\Delta}$
$\hat{\tau}$	$\hat{S} + x\hat{\tau}$	$\hat{I} - x\hat{\tau}$	$\hat{\Delta}$	$\hat{\rho} - x\hat{\Delta}$	$\hat{\tau}$

depletes these by one; $\hat{\rho}$ returns the number of infectives and depletes these by one; and $\hat{\tau}$ returns the number of susceptibles, depletes the susceptible population by one, and increases the infectious population by one. The dynamical model is then

$$\frac{d}{dt}|p(t)\rangle = (\gamma(t)(\hat{\rho} - \hat{I}) + \lambda(t)(\hat{\tau} - \hat{S}))|p(t)\rangle. \tag{19}$$

Note that while (19) does not make use of the operator $\hat{\Delta}$, it is necessary to include this to have an algebra that is closed under the action of the Lie bracket. The full set of Lie brackets is shown in Table 1, and the action of the exponential adjoint endomorphism is shown in Table 2. We then look for a solution of the form

$$|p(t)\rangle = e^{g_1(t)\hat{\Delta}}e^{g_2(t)\hat{\tau}}e^{g_3(t)\hat{S}}e^{g_4(t)\hat{\rho}}e^{g_5(t)\hat{I}}|p(0)\rangle. \tag{20}$$

Going through the same procedure as before gives the solution

$$g_1(t) = e^{\Lambda(t)}\left(1 - e^{-\Lambda(t)} - \int_0^t \lambda(u)e^{-\Lambda(u)}e^{\Gamma(u)-\Gamma(t)} du\right), \tag{21a}$$

$$g_2(t) = e^{\Lambda(t)}\int_0^t \lambda(u)e^{-\Lambda(u)}e^{\Gamma(u)-\Gamma(t)} du, \tag{21b}$$

$$g_3(t) = -\Lambda(t), \tag{21c}$$

$$g_4(t) = 1 - e^{-\Gamma(t)}, \tag{21d}$$

$$g_5(t) = -\Gamma(t), \tag{21e}$$

where

$$\Lambda(t) := \int_0^t \lambda(u) du \quad \text{and} \quad \Gamma(t) := \int_0^t \gamma(u) du.$$

To check this result, assuming that $|p(0)\rangle = |N, 0\rangle$ and substituting (21) into (20) gives

$$\Pr(S, I \mid N, t) = \binom{N!}{S! I! (N - S - I)!} (\pi_1)^S (\pi_2)^I (1 - \pi_1 - \pi_2)^{N-S-I}, \tag{22}$$

where

$$\pi_1 = e^{-\Lambda(t)}, \quad \pi_2 = \int_0^t \lambda(u) e^{-\Lambda(u)} e^{\Gamma(u) - \Gamma(t)} du. \tag{23}$$

This is the form we would expect for this solution; considering each individual’s probability of remaining susceptible to be π_1 and being infections to be π_2 , these should obey

$$\frac{d}{dt} \begin{pmatrix} \pi_1 \\ \pi_2 \end{pmatrix} = \begin{pmatrix} -\lambda(t) & 0 \\ \lambda(t) & -\gamma(t) \end{pmatrix} \begin{pmatrix} \pi_1 \\ \pi_2 \end{pmatrix},$$

which has solution (23), and the independence of each individual leads to the multinomial distribution (22) for the cohort as a whole. As for the birth–death process, the analytic result obtained through Lie algebraic methods can be obtained otherwise, and which method is preferable will depend on which further calculations one wishes to undertake.

3.3. A pure birth process

Now suppose that we have a pure birth process, characterised by a stochastic variable $N(t)$ and transition rate

$$n \rightarrow n + 1 \quad \text{at rate } a(t) + nb(t). \tag{24}$$

Special forms of $a(t)$ and $b(t)$ have been used to model the formation of social contacts [1]. Defining components of a vector $p_n = \Pr(N(t) = n)$, and assuming that there is a maximum count of interest m , such that we keep track of only $\Pr(N(t) > m)$ as the final component of p , we write the Kolmogorov equation for this process as

$$\frac{dp}{dt} = (a(t)P_1 + b(t)Q)p. \tag{25}$$

The generating matrices P_1 and Q are implicitly defined by (24); however, an additional set of matrices are required to produce a vector space that is closed under the Lie bracket. These matrices take a similar form to P_1 and are indexed by integer i . The definition of the matrices used, in terms of the Kronecker delta, is

$$(Q)_{n,k} = k(\delta_{n,k+1} - \delta_{n,k}), \quad (P_i)_{n,k} = \delta_{n,k+i} - \delta_{n,k+i-1}.$$

These matrices satisfy the commutation relations

$$[P_i, P_j] = 0, \quad [P_i, Q] = -iP_{i+1} + (i - 1)P_i. \tag{26}$$

Note that it is here that the assumption of a finite state space (i.e. the introduction of a maximum count of interest above) allows the solution method to work, since otherwise a countably infinite number of matrices would be needed. We then look for a solution of the form

$$p(t) = e^{f_1(t)P_1} \dots e^{f_m(t)P_m} e^{g(t)Q} p(0),$$

and make use of (10) to give the following system of equations:

$$\dot{f}_1(t) = a(t), \quad \dot{g}(t) = b(t), \quad \dot{f}_{i>1}(t) = (i - 1)\dot{g}(t)[f_{i-1}(t) - f_i(t)].$$

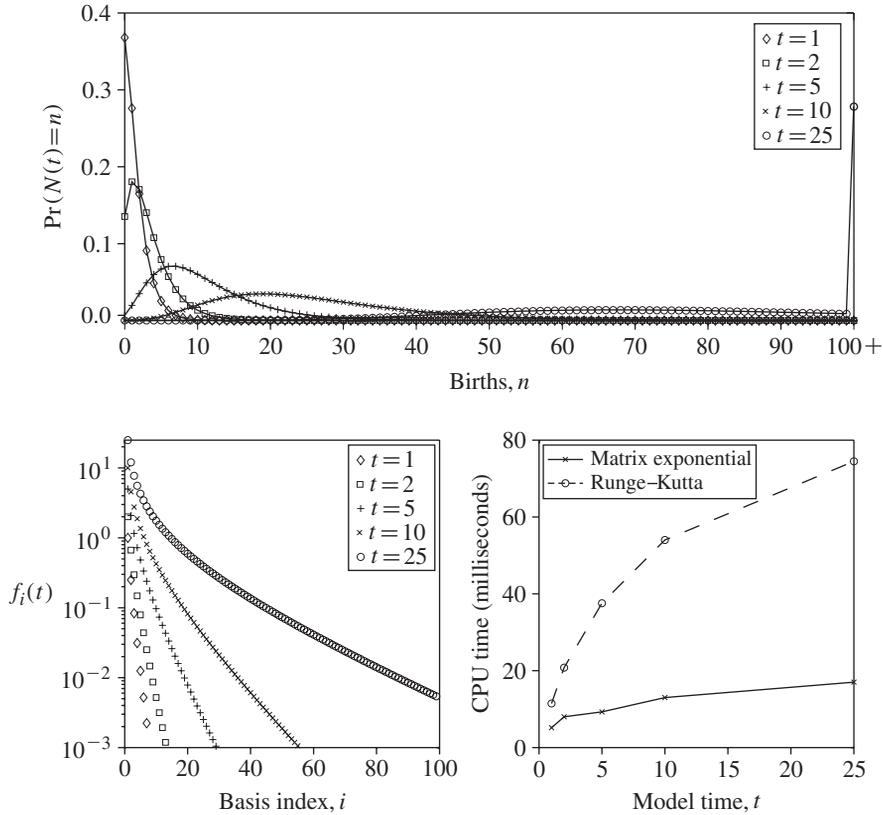


FIGURE 1: Numerical results for a time-inhomogeneous pure birth process. *Top*: probability distributions at different times. Matrix exponential solutions are shown with markers and direct integration via Runge–Kutta is shown with solid lines—clearly, these are numerically indistinguishable. *Bottom left*: values of the quantities f_i over time. *Bottom right*: CPU time needed to run each method as a function of the model time.

These equations have the solution

$$\begin{aligned}
 f_1(t) &= \mathcal{A}(t) = \int_0^t a(u) \, du, \\
 g(t) &= \mathcal{B}(t) = \int_0^t b(u) \, du, \\
 f_{i>1}(t) &= e^{-(i-1)\mathcal{B}(t)} \int_0^t a(u)(e^{\mathcal{B}(t)} - e^{\mathcal{B}(u)})^{i-1} \, du.
 \end{aligned}$$

Figure 1 shows numerical results for this system for the simple choice

$$a(t) = 1, \quad b(t) = \frac{1}{1+t}, \quad m = 100, \quad N(0) = 0.$$

Note that, for this initial condition, and due to relations (26), we can write the solution as

$$p_n(t) = \left(\exp \left[\sum_i f_i(t) P_i \right] \right)_{n,0},$$

the evaluation of which can be seen in Figure 1 to give a significant computational advantage, as implemented in EXPOKIT [11], compared to direct integration of (25) through Runge–Kutta, as implemented in the MATLAB[®] function `ode45`, at large times. Perhaps unexpectedly, this is seen despite the relatively large value of m . These plots show that, while the ODE solver uses a more sophisticated relationship than (5) to obtain better than $O(t)$ performance, it is still much more sensitive to model time t than the matrix exponential method.

There will, of course, be more complex systems where numerical integration via Runge–Kutta is impractical, but analytic integration is simple (e.g. if either $a(t)$ or $b(t)$ is a square wave rapidly oscillating between 0 and 1); for these systems, the matrix exponential method will further outperform Runge–Kutta. But there will also be systems where direct numerical integration is straightforward and there is no simply obtained form for $f_i(t)$ and $g(t)$, meaning that the matrix exponential solution is not useful.

4. Discussion

In this paper we have considered the solutions of time-inhomogeneous Markov chains in population modelling through the use of matrix exponentials. This was done using the method of Lie algebras originally developed for applications in physical sciences [14]. In contrast to physical applications, population models are often insufficiently symmetric to write down a well-studied Lie algebra. In even the relatively simple pure birth process considered, a large number of basis matrices were needed to derive a matrix exponential solution; but despite this the exponential solution is useful if a derivative with respect to a model parameter is required [15], and perhaps more importantly often has a significant numerical advantage over direct integration of the ODE system [8]. Given the popularity of computationally intensive inference in modern population models [3], any such improvement in numerical efficiency of likelihood evaluation can have important practical benefits.

Acknowledgements

This work was funded by the UK Engineering and Physical Sciences Research Council. The author would like to thank Josh Ross and Jeremy Sumner, in addition to the editor and anonymous referee, for helpful comments on this work.

References

- [1] DANON, L., HOUSE, T., KEELING, M. J. AND READ, J. M. (2012). Social encounter networks: collective properties and disease transmission. Submitted.
- [2] DODD, P. J. AND FERGUSON, N. M. (2009). A many-body field theory approach to stochastic models in population biology. *PLoS ONE* **4**, e6855, 9pp.
- [3] GILKS, W. R., RICHARDSON, S. AND SPIEGELHALTER, D. J. (1995). *Markov Chain Monte Carlo in Practice*. Chapman and Hall/CRC, Boca Raton, FL.
- [4] GRIMMETT, G. R. AND STIRZAKER, D. R. (2001). *Probability and Random Processes*, 3rd edn. Oxford University Press, New York.
- [5] JARVIS, P. D., BASHFORD, J. D. AND SUMNER, J. G. (2005). Path integral formulation and Feynman rules for phylogenetic branching models. *J. Phys. A* **38**, 9621–9647.
- [6] JOHNSON, J. E. (1985). Markov-type Lie groups in $GL(n, \mathbb{R})$. *J. Math. Phys.* **26**, 252–257.
- [7] KEELING, M. J. AND ROHANI, P. (2008). *Modeling Infectious Diseases in Humans and Animals*. Princeton University Press.
- [8] KEELING, M. J. AND ROSS, J. V. (2008). On methods for studying stochastic disease dynamics. *J. R. Soc. Interface* **5**, 171–181.
- [9] MOURAD, B. (2004). On a Lie-theoretic approach to generalized doubly stochastic matrices and applications. *Linear Multilinear Algebra* **52**, 99–113.

- [10] ROSS, J. V. (2010). Computationally exact methods for stochastic periodic dynamics spatiotemporal dispersal and temporally forced transmission. *J. Theoret. Biol.* **262**, 14–22.
- [11] SIDJE, R. B. (1998). Expokit: a software package for computing matrix exponentials. *ACM Trans. Math. Software* **24**, 130–156.
- [12] SUMNER, J., FERNANDEZ-SANCHEZ, J. AND JARVIS, P. (2012). Lie Markov models. *J. Theoret. Biol.* **298**, 16–32.
- [13] SUMNER, J. G., HOLLAND, B. R. AND JARVIS, P. D. (2011). The algebra of the general Markov model on phylogenetic trees and networks. *Bull. Math. Biol.* 17pp.
- [14] WEI, J. AND NORMAN, E. (1963). Lie algebraic solution of linear differential equations. *J. Math. Phys.* **4**, 575–581.
- [15] WILCOX, R. M. (1967). Exponential operators and parameter differentiation in quantum physics. *J. Math. Phys.* **8**, 962–982.