UNIVERSITY OF CALGARY

The Application of Lie Theory to Markov Processes

Computation of the Maximum Likelihood Estimator of the Generator of Continuous Time Markov Processes from a Stopped Random Variable

by

Aaron Geoffrey Sheldon

A PROJECT

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE

DEGREE OF ”Masters of Science”

GRADUATE PROGRAM IN ”Statistics”

CALGARY, ALBERTA

August, 2016

Abstract

Continuous time Markov processes on finite state spaces and Lie theory have individually been highly productive fields of investigation for more than a century. However, the two fields remain ripe for cross pollination. In particular the application of results from Lie theory will yield novel computational methods for estimation problems in continuous time Markov processes on finite state spaces. In this project we derive the minimal Lie algebra that contains the generators of a continuous time Markov process on a finite state space, and then using the guarantees of algebraic and analytic closure construct a Newton-Raphson algorithm for maximum likelihood estimation of the generator of a continuous time Markov process on a finite state space from stopped random variables using Páde approximations for Taylor series expressions of the first and second order derivatives of the exponential map.

Acknowledgements

Anyone who has the patience to deal with me.

Table of Contents

List of Tables

List of Figures and Illustrations

List of Symbols, Abbreviations and Nomenclature

Symbol	Definition
U of C	University of Calgary
$A, B, C, \ldots$	Matrices, except for I which we take as identity
$a, b, c, \ldots$	Constants
$\vec{a}, \vec{b}, \vec{c}, \ldots$	Vectors
$\hat{a}, \hat{b}, \hat{c}, \ldots$	Unit vectors
$\vec{\mathbb{1}}$	Row sum vector
$\hat{\mathbb{1}}$	Normalized row sum vector
$St(\hat{\mathbb{1}})$	Stochastic Lie group with respect to $\hat{\mathbb{1}}$
$\mathfrak{st}(\hat{\mathbb{1}})$	Stochastic Lie algebra with respect to $\hat{\mathbb{1}}$
$St(\hat{\mathbb{1}}, \hat{\mathbb{1}})$	Doubly stochastic Lie group with respect to $\hat{\mathbb{1}}$
$\mathfrak{st}(\hat{\mathbb{1}}, \hat{\mathbb{1}})$	Doubly stochastic Lie algebra with respect to $\hat{\mathbb{1}}$
$\hat{a} \otimes \hat{b}$	Kronecker product of unit vectors $\hat{a}$ and $\hat{b}$
$\langle \hat{a}, \hat{b} \rangle$	Inner product of unit vectors $\hat{a}$ and $\hat{b}$
$\mathbb{P}[\Sigma]$	Probability measure of a measurable set Σ
$\mathbb{I}[\Sigma]$	Indicator function of a measurable set Σ
$\mathscr{F}_t$	Filtration of sigma algebras with respect to continuous parameter t
X_t	Stochastic process on the filtration $\mathscr{F}_t$
$\mathbb{E}[X_t]$	Expectation of X_t
$\mathbb{V}ar[X_t]$	Variance of X_t
$\mathbb{C}ov[X_t, Y_t]$	Covariance of X_t and Y_t
$\mathbb{C}or[X_t, Y_t]$	Correlation of X_t and Y_t
$Ad_X A$	Lie group Adjoint operator XAX^{-1}
$ad_X A$	Lie algebra adjoint operator $[X, A]$

$[A,B]$	Commutator bracket $AB - BA$ of matrices A and B
δ_{ij}	Dirac delta operator

Chapter 1

Introduction

1.1 Motivation and Direction

Markov process have a rich and extensive history in the statistical and probability analysis of fields such as the life sciences, operations research, queuing theory communications, natural language, finance and machine learning. Have modeled syntax of sentences, disease states, cancer survival, epidemiology and demographics, with models such as phase type, birth-death, and hidden State of the art computational methods are focused on hidden Markov Models on a finite state set, and discrete time steps; assuming all transitions are observed but are obscured with noise. This is the focus of the Baum-Welch and Viterbi algorithms, and more recently RUST models. This is because Markov models have an intuitive phenomenological interpretation.

The reason for the focus on discrete time finite state Markov processes is that in dimension greater than two the matrix exponential does not have a closed form entry wise. This is because polynomials of degree greater than two are not generally solvable. Instead computational approximations for the matrix exponential must be used. This work extends and generalizes those computational methods to include the gradient and Hessian of the matrix exponential.

The work presented in this project concerns continuous time Markov processes on a finite state space, where by definition it is not possible to observe all the transitions. Instead what is observed typical is are stopped statistics, such as first hitting times from one state to another.

The second chapter establishes the algebraic and analytic closure properties necessary for chapter three. Chapter two has a secondary role to help develop the physical intuition for the stochastic Lie group necessary to work through the derivatives and approximations of chapter three. The third chapter derives the the first and second order derivatives of the exponential map and their Padé Approximation. The fourth chapter derives the maximum likelihood estimators from first hitting times The fifth chapter concludes with summarizing remarks and a discussion of the direction for further investigation. Throughout this work we will attempt to conform to a simplified version of Lamport's guide to structuring and presenting proofs.

1.2 Background

1.2.1 Continuous Time Markov Process

1.2.2 Maximum Likelihood Estimation

1.2.3 Lie Theory

1.2.4 Padé Approximation

1.2.5 Newton-Raphson Method

Chapter 2

The Lie Algebra of the Generators of Continuous Time Markov Processes on Finite State Spaces

2.1 Stochastic Matrices

The classical Lie algebras of physics, like the infinitesimal symmetries of the special unitary algebra $\mathfrak{su}(n)$, are defined with respect to invariants of a Banach algebra, such as the matrix invariants of the determinant, trace, or norm. In contrast stochastic matrices are always characterized with respect to a specific unit vector, which we will denote $\hat{\mathbb{1}}$. In the next two sections we provide an explicit construction and characterization of the Lie algebra of stochastic matrices, building on the original the work of ??.

The common approach to stochastic matrices begins with the restriction that the matrices have non-negative entries with respect to the standard orthonormal basis for the vector space on which it acts; namely $\langle \hat{e}_i, A\hat{e}_j \rangle \geq 0$ for all i, j. In addition to allowing for singular matrices, this poses an immediate obstacle to the necessary closure with respect to matrix inversion required for matrix groups. As the inverse of a stochastic matrix need not have non-negative entries with respect to the standard orthonormal basis.

For the moment we will set aside the restriction that the entries be non-negative, and instead begin with a generalization of the concept of fixed row sums. We will show that this generalization is preserved by matrix inversion, and then develop an orthonormal basis from which specific matrices with non-negative entries, with respect to the basis, can be constructed. In essence tackling the problem from the reverse direction, starting with the more general idea of fixed row sums, and then specifying to matrices with non-negative entries with respect to a constructed orthonormal basis.

Definition 1. A matrix A is stochastic with respect to the unit vector $\hat{\mathbb{1}}$ if $A\hat{\mathbb{1}} = \hat{\mathbb{1}}$

Note that this definition does not stipulate any conditions on non-singularity, and thus includes all the matrices in the convex polytope of stochastic matrices. For an n dimensional vector space the vector $\vec{\mathbb{1}} = \sqrt{n}\hat{\mathbb{1}}$ acts as the row sum operator on matrices stochastic with respect to $\hat{\mathbb{1}}$. We will make this claim more precise after we dispense with a few more foundational definitions.

Definition 2. Let $St(\hat{\mathbb{1}})$ denote the stochastic Lie group of invertible matrices stochastic with respect to $\hat{\mathbb{1}}$

It is tempting to view the name stochastic Lie group as a bait and switch, or at least an abuse of the terminology, given we have removed the usual convex polytope of stochastic matrices and replaced it with a group of invertible matrices with a common eigenvector $\hat{\mathbb{1}}$. Previous authors ?? have denoted the convex polytope of stochastic matrices as the stochastic semi-group, and the group of invertible matrices as the pseudo-stochastic Lie group. One could even consider incorporating Markov into the name, in reference to the fact that the transition matrices of a continuous Markov process on a finite state space are by definition invertible and have common eigenvector $\hat{\mathbb{1}}$. However the suffix of Lie group in the name connotes both sufficient additional restrictions to make the name distinct, and still allows for an indication of a relationship with the original concept. Of course, this definition immediately necessitates proof of the claim embedded in the definition.

Lemma 1. *$St(\hat{\mathbb{1}})$ is a Lie group*

Proof. We proceed by working mechanistically through the Lie group axioms.

1. The identity element I is in $St(\hat{\mathbb{1}})$. Clearly I is invertible and $I\hat{\mathbb{1}} = \hat{\mathbb{1}}$.
2. If $A, B \in St(\hat{\mathbb{1}})$ then $AB \in St(\hat{\mathbb{1}})$. This follows from the computation $AB\hat{\mathbb{1}} = A\hat{\mathbb{1}} = \hat{\mathbb{1}}$.
3. If $A \in St(\hat{\mathbb{1}})$ then $A^{-1} \in St(\hat{\mathbb{1}})$. Recognize that $A\hat{\mathbb{1}} = \hat{\mathbb{1}}$ implies $\hat{\mathbb{1}} = A^{-1}A\hat{\mathbb{1}} = A^{-1}\hat{\mathbb{1}}$.
4. Associativity follows from $St(\hat{\mathbb{1}})$ being a subgroup of $GL(n)$.

5. Finally, that the matrix product $A^{-1}B$ is smooth for all $A, B \in St(\hat{\mathbb{1}})$ likewise follows from $St(\hat{\mathbb{1}}) < GL(n)$.

$\square$

That $St(\hat{\mathbb{1}})$ is a proper matrix Lie group implies that it must be infinitesimal generated by elements of a Lie algebra.

Definition 3. Let $\mathfrak{st}(\hat{\mathbb{1}})$ denote the stochastic Lie algebra of $St(\hat{\mathbb{1}})$

By infinitesimally generated we mean that every element of $St(\hat{\mathbb{1}})$ is a matrix exponential of at least one element in $\mathfrak{st}(\hat{\mathbb{1}})$. We can fully characterize this algebra as the set of matrices such that their row sums are zero with respect to $\hat{\mathbb{1}}$.

Lemma 2. *The algebra $\mathfrak{st}(\hat{\mathbb{1}})$ is exactly the set of all matrices with $\hat{\mathbb{1}}$ in their kernel.*

Proof. Working through the forward and backward inclusions we have

1. Suppose $A\hat{\mathbb{1}} = 0$ then from the definition of the matrix exponential we have:

$$\begin{aligned} \exp(A)\,\hat{\mathbb{1}} &= \sum_{n=0}^{\infty} \frac{1}{n!} A^n \hat{\mathbb{1}} \\ &= \hat{\mathbb{1}} + \sum_{n=1}^{\infty} \frac{1}{n!} 0 \\ &= \hat{\mathbb{1}} \end{aligned}$$

   Thus $\exp(A) \in St(\hat{\mathbb{1}})$ implying that $A \in \mathfrak{st}(\hat{\mathbb{1}})$

2. Now begin with the reverse assumption, that $A \in \mathfrak{st}(\hat{\mathbb{1}})$. For all $t \in \mathbb{R}$ we have $\exp(tA)\,\hat{\mathbb{1}} = \hat{\mathbb{1}}$. Differentiation with respect to t and evaluation at $t = 0$ yields

$$\begin{aligned} 0 &= \frac{d}{dt} \hat{\mathbb{1}} \Big|_{t=0} \\ &= \frac{d}{dt} \exp(tA)\,\hat{\mathbb{1}} \Big|_{t=0} \\ &= \exp(tA) A \hat{\mathbb{1}} \big|_{t=0} \\ &= A\hat{\mathbb{1}} \end{aligned}$$

$\square$

Over an n dimensional vector space, the condition on a matrix A that $A\hat{\mathbb{1}} = 0$ places n constraints on the n^2 dimensions of A. This leaves $n^2 - n$ free dimensions on $\mathfrak{st}(\hat{\mathbb{1}})$, when considered as a vector space. This hints that we can construct a generator of $\mathfrak{st}(\hat{\mathbb{1}})$ from order pairs of basis elements $\hat{e}_i$ for the vector space of $\hat{\mathbb{1}}$. To see how this is done we first construct a useful basis for the vector space to which $\hat{\mathbb{1}}$ is a member.

Lemma 3. *There exists an orthonormal basis $\hat{e}_i$ such that $\left\langle \hat{e}_i, \hat{\mathbb{1}} \right\rangle = \frac{1}{\sqrt{n}}$ for all i*

Proof. While a basis with the stipulated properties can be constructed through the Gram-Schmidt process, the proof of the existence proceeds by induction.

1. For $n = 1$ the desired basis is precisely the trivial set $\left\{ \hat{\mathbb{1}} \right\}$ which satisfies the condition that $\left\langle \hat{\mathbb{1}}, \hat{\mathbb{1}} \right\rangle = 1$.
2. Assume the claim is true for n. For $n+1$ pick a unit vector $\hat{e}_\perp$ that is orthogonal to $\hat{\mathbb{1}}$ and construct the unit vector $\hat{e}_{n+1} = \frac{1}{\sqrt{n+1}} \hat{\mathbb{1}} + \sqrt{\frac{n}{n+1}} \hat{e}_\perp$. Clearly $\hat{e}_{n+1}$ satisfies the condition $\left\langle \hat{e}_{n+1}, \hat{\mathbb{1}} \right\rangle = \frac{1}{\sqrt{n+1}}$.
3. To use the the induction assumption we construct a new row sum unit vector $\hat{\mathbb{1}}_n = \sqrt{\frac{n+1}{n}} \hat{\mathbb{1}} - \frac{1}{\sqrt{n}} \hat{e}_{n+1}$ in one dimension lower by projecting onto the subspace orthogonal to $\hat{e}_{n+1}$.
4. By the induction assumption there exists a basis $\hat{e}_i$ with $i \leq n$, such that $\left\langle \hat{e}_i, \hat{\mathbb{1}}_n \right\rangle = \frac{1}{\sqrt{n}}$.
5. Because $\hat{e}_i$ with $i \leq n$ was constructed in the space orthogonal to $\hat{e}_{n+1}$ if follows that $\left\langle \hat{e}_i, \hat{e}_j \right\rangle = \delta_{ij}$ for all $i, j \leq n+1$.
6. Then using the definitions of $\hat{e}_{n+1}$ and $\hat{\mathbb{1}}_n$ we can calculate the inner product $\left\langle \hat{e}_i, \hat{\mathbb{1}}_n \right\rangle$

for $i \leq n$

$$\begin{aligned}\frac{1}{\sqrt{n}} &= \left\langle \hat{e}_i, \hat{\mathbb{1}}_n \right\rangle \\ &= \sqrt{\frac{n+1}{n}} \left\langle \hat{e}_j, \hat{\mathbb{1}} \right\rangle - \frac{1}{\sqrt{n}} \left\langle \hat{e}_i, \hat{e}_{n+1} \right\rangle \\ &= \sqrt{\frac{n+1}{n}} \left\langle \hat{e}_j, \hat{\mathbb{1}} \right\rangle\end{aligned}$$

Inverting the fraction in the equality yields $\left\langle \hat{e}_i, \hat{\mathbb{1}} \right\rangle = \frac{1}{\sqrt{n+1}}$ for all $i \leq n+1$.

□

As a direct result of the construction of the basis vectors $\hat{e}_i$ we see that $\vec{\mathbb{1}} = \sum_{i=1}^{n} \hat{e}_i$. Thus $\vec{\mathbb{1}}$ can be interpreted as the row sum vector in basis $\hat{e}_i$.

The constructed basis leads naturally to considering the minimal non-trivial matrices $C_{ij} = \hat{e}_i \otimes \left(\hat{e}_j - \hat{e}_i\right)$ as holding significance in the structure of $\mathfrak{st}(\hat{\mathbb{1}})$. In fact this will be the central result of this chapter: that the algebraic closure of the matrices C_{ij} is the stochastic Lie algebra $\mathfrak{st}(\hat{\mathbb{1}})$. To establish this result we need a preliminary result that proves the commutators $\left[C_{ij}, C_{kl}\right]$ are linear combinations of matrices C_{ij}.

Lemma 4.

$$C_{ij}C_{kl} = \begin{cases} -C_{il} & i = k, \\ C_{il} - C_{ij} & j = k, \\ 0 & \textit{otherwise}. \end{cases}$$

Proof. We proceed in two steps; calculating the terms of the products, then simplifying the cases, always assuming $i \neq j$ and $k \neq l$.

1. Term wise computation of the Kronecker products yields

$$
\begin{aligned}
C_{ij}C_{kl} &= \hat{e}_i \otimes \left(\hat{e}_j - \hat{e}_i\right) \hat{e}_k \otimes \left(\hat{e}_l - \hat{e}_k\right) \\
&= \hat{e}_i \otimes \hat{e}_j \hat{e}_k \otimes \hat{e}_l + \hat{e}_i \otimes \hat{e}_i \hat{e}_k \otimes \hat{e}_k - \hat{e}_i \otimes \hat{e}_j \hat{e}_k \otimes \hat{e}_k - \hat{e}_i \otimes \hat{e}_i \hat{e}_k \otimes \hat{e}_l \\
&= \delta_{jk} \hat{e}_i \otimes \hat{e}_l + \delta_{ik} \hat{e}_i \otimes \hat{e}_k - \delta_{jk} \hat{e}_i \otimes \hat{e}_k - \delta_{ik} \hat{e}_i \otimes \hat{e}_l \\
&= \left(\delta_{jk} - \delta_{ik}\right) \hat{e}_i \otimes \left(\hat{e}_l - \hat{e}_k\right)
\end{aligned}
$$

2. We work through each case of $\delta_{jk} - \delta_{ik}$, starting with the case $i = k$

$$
\begin{aligned}
C_{ij}C_{il} &= \left(\delta_{jk} - \delta_{ii}\right) \hat{e}_i \otimes \left(\hat{e}_l - \hat{e}_i\right) \\
&= -\hat{e}_i \otimes \left(\hat{e}_l - \hat{e}_i\right) \\
&= -C_{il}
\end{aligned}
$$

3. When $j = k$ we have

$$
\begin{aligned}
C_{ij}C_{jl} &= \left(\delta_{jj} - \delta_{ij}\right) \hat{e}_i \otimes \left(\hat{e}_l - \hat{e}_j\right) \\
&= \hat{e}_i \otimes \left(\hat{e}_l - \hat{e}_j\right) \\
&= \hat{e}_i \otimes \left(\hat{e}_l - \hat{e}_i + \hat{e}_i - \hat{e}_j\right) \\
&= C_{il} - C_{ij}
\end{aligned}
$$

4. Finally when none of the previous conditions apply

$$
\begin{aligned}
C_{ij}C_{kl} &= \left(\delta_{jk} - \delta_{ik}\right) \hat{e}_i \otimes \left(\hat{e}_l - \hat{e}_k\right) \\
&= 0 \cdot \hat{e}_i \otimes \left(\hat{e}_l - \hat{e}_k\right) \\
&= 0
\end{aligned}
$$

□

While this result is sufficient to accomplish the central result, it is worth carrying through with the computation of the structure constants of the generators.

Corollary 1.

$$[C_{ij}, C_{kl}] = \begin{cases} C_{ij} - C_{il} & i = k, \\ C_{kj} - C_{ki} & i = l, \\ C_{il} - C_{ij} & j = k, \\ 0 & otherwise. \end{cases}$$

Proof. As in the previous lemma we work case wise through the equalities.

1. Starting with $i = k$

$$\begin{aligned} [C_{ij}, C_{il}] &= C_{ij}C_{il} - C_{il}C_{ij} \\ &= C_{ij} - C_{il} \end{aligned}$$

2. For $i = l$

$$\begin{aligned} [C_{ij}, C_{ki}] &= C_{ij}C_{ki} - C_{ki}C_{ij} \\ &= C_{kj} - C_{ki} \end{aligned}$$

3. For $j = k$

$$\begin{aligned} [C_{ij}, C_{jl}] &= C_{ij}C_{jl} - C_{jl}C_{ij} \\ &= C_{il} - C_{ij} \end{aligned}$$

4. When none of the conditions apply

$$\begin{aligned} [C_{ij}, C_{kl}] &= C_{ij}C_{kl} - C_{kl}C_{ij} \\ &= 0 \end{aligned}$$

□

We can now proceed with the central result that motivates this chapter.

Theorem 1. *The canonical generators of* $\mathfrak{st}(\hat{\mathbb{1}})$ *are* C_{ij}

Proof. The previous lemma has established that the products, and thus the commutators, of C_{ij} are linear in C_{ij}. We then have to prove that the smallest algebra that contains C_{ij} is $\mathfrak{st}(\hat{\mathbb{1}})$. As thus, it is sufficient to prove that matrices C_{ij} from a basis for $\mathfrak{st}(\hat{\mathbb{1}})$. This is because a necessary condition for an algebra to contain the matrices C_{ij} is that it must contain all sums of the matrices C_{ij}. If one could sum their way out of the algebra then it would not be an algebra.

1. That $\mathfrak{st}(\hat{\mathbb{1}})$ is an $n^2 - n$ dimensional vector space should be clear from the previous discussion. A full formal proof of this claim is found through induction on the dimension n.

2. The matrices C_{ij} are in $\mathfrak{st}(\hat{\mathbb{1}})$. From the definition of the canonical generators

$$
\begin{aligned}
C_{ij}\hat{\mathbb{1}} &= \hat{e}_i \otimes \left(\hat{e}_j - \hat{e}_i\right)\hat{\mathbb{1}} \\
&= \hat{e}_i \left(\left\langle \hat{e}_j, \hat{\mathbb{1}} \right\rangle - \left\langle \hat{e}_i, \hat{\mathbb{1}} \right\rangle\right) \\
&= \hat{e}_i \left(\frac{1}{\sqrt{n}} - \frac{1}{\sqrt{n}}\right) \\
&= 0
\end{aligned}
$$

3. C_{ij} is a set of $n^2 - n$ linear independent matrices and so must form a basis for all of $\mathfrak{st}(\hat{\mathbb{1}})$. That there are only $n^2 - n$ matrices is clear from the fact that $C_{ii} = 0$. While the formal proof of linear independence is again found through induction on the dimension n.

□

The previous theorem serves as the definition of a set of canonical generators of $\mathfrak{st}(\hat{\mathbb{1}})$. It is important to note that neither the basis $\hat{e}_i$ nor the canonical generators C_{ij} are unique. They are uniquely defined only up to rotations orthogonal to the vector $\hat{\mathbb{1}}$.

Since $C_{ij} \in \mathfrak{st}(\hat{\mathbb{1}})$ its matrix exponential must be in $St(\hat{\mathbb{1}})$, which can be summarized in the following corollary.

Corollary 2. $\exp\left(\alpha C_{ij}\right) = I + (1 - e^{-\alpha})C_{ij}$

Proof. From the definition of the matrix exponential

$$\begin{aligned}
\exp\left(\alpha C_{ij}\right) &= \sum_{n=0}^{\infty} \frac{1}{n!} \alpha^n C_{ij}^n \\
&= I + \sum_{n=1}^{\infty} \frac{1}{n!} (-1)^{n+1} \alpha^n C_{ij} \\
&= I + \left(1 - e^{-\alpha}\right) C_{ij}
\end{aligned}$$

$\square$

This last corollary admits a intuitive heuristic interpretation: that each canonical generator C_{ij} can be thought of as measuring the infinitesimal transition rate, or flow of probability, from the state represented by basis element $\hat{e}_i$ to the state represented by the basis element $\hat{e}_j$. This can be seen by considering the matrix representation of $\exp\left(\alpha C_{ij}\right)$ in the basis spanned by $\hat{e}_i$ and $\hat{e}_j$.

$$\begin{aligned}
I + \left(1 - e^{-\alpha}\right) C_{ij} &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \left(1 - e^{-\alpha}\right) \begin{pmatrix} -1 & 1 \\ 0 & 0 \end{pmatrix} \\
&= \begin{pmatrix} e^{-\alpha} & 1 - e^{-\alpha} \\ 0 & 1 \end{pmatrix}
\end{aligned}$$

Despite the fact that $\mathfrak{st}(\hat{\mathbb{1}})$ is a real vector space the limits in $St(\hat{\mathbb{1}})$ in the positive directions of C_{ij} are finite, namely, in the basis spanned by $\hat{e}_i$ and $\hat{e}_j$

$$\begin{aligned}
\lim_{\alpha \to \infty} \exp\left(\alpha C_{ij}\right) &= I + C_{ij} \\
&= \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}
\end{aligned}$$

In general, a matrix Lie group M is it's Lie algebra offset by the identity element $M = I + \mathfrak{m}$; which can be seen from the Taylor series expansion of the matrix exponential. It follows that for a $G = \sum_{ij} g_{ij} C_{ij} \in \mathfrak{st}(\hat{\mathbb{1}})$ the matrix exponential is $\exp(G) = I + \sum_{ij} h_{ij} C_{ij}$. Unfortunately, because polynomials of degree greater than two are generally unsolvable, the relationship between the coefficients g_{ij} and h_{ij} is highly non-trivial in most circumstances. We shall see that an exception to this difficulty can be found in the formulation of the distribution of first hitting times from a fixed initial state.

At this point it is worth briefly revisiting the distinction between the standard convex polytope of stochastic matrices and the stochastic Lie group, to develop some physical intuition into the relationship between the two sets of matrices which have a non-trivial and geometrical interesting intersection. Consider a general stochastic matrix $A\hat{\mathbb{1}} = \hat{\mathbb{1}}$, not necessarily invertible, on a $n = 2$ dimensional vector space, in the standard orthonormal basis

$$A = \begin{pmatrix} 1-a & a \\ b & 1-b \end{pmatrix}$$

The matrix is invertible when $\det(A) = 1 - a - b \neq 0$, and has non-negative entries in the standard orthonormal basis when $a, b \in [0,1]$. The convex polytope of 2×2 stochastic matrices in the standard basis is the convex hull of the vertexes

$$\begin{aligned} V_0 &= I + C_{21} \\ V_1 &= I + C_{12} + C_{21} \\ V_2 &= I \\ V_3 &= I + C_{12} \end{aligned}$$

Geometrically fixing a basis $\hat{e}_i$ defines a convex polytope in the form of a unit hypercube $[0,1]^{n^2-n}$ in the parameter space isomorphic to $\mathfrak{st}(\hat{\mathbb{1}})$. The vertexes of the convex polytope of stochastic matrices are the n^n sums of the limiting elements

$$
\begin{aligned}
V_k &= \sum_{i=1}^{n} \hat{e}_i \otimes \hat{e}_{j_n^i(k)} \\
&= I + \sum_{i=1}^{n} C_{i\, j_n^i(k)} \\
&= (1-n)\, I + \sum_{i=1}^{n} I + C_{i\, j_n^i(k)} \\
&= (1-n)\, I + \sum_{i=1}^{n} \lim_{\alpha \to \infty} \exp\left(\alpha C_{i\, j_n^i(k)}\right)
\end{aligned}
$$

where $j_n^i(k) = 1 + \left\lfloor \frac{k}{n^{i-1}} \right\rfloor \mod n$ is 1 plus the i digit of $0 \le k < n^n$ in base n. Analytically $St(\hat{\mathbb{1}})$ is not closed with respect to limits along the directions in the tangent space $\mathfrak{st}(\hat{\mathbb{1}})$, yet its closure contains the vertexes for the convex polytope of stochastic matrices.

The constructed basis elements $\hat{e}_i$ are a well defined enumeration of the states of a continuous Markov process on a finite state space; in that when the multipliers of the canonical generators are non-negative the matrix exponential gives a proper transition matrix for the process. The reverse is also true, up to a choice of branch of the matrix logarithm. Proofs of either direction of inclusions can be found in a number of texts, for example ??, and typically involve studying the resolvent of the matrices of $St(\hat{\mathbb{1}})$, and the derivatives of the exponentials of the matrices of $\mathfrak{st}(\hat{\mathbb{1}})$.

We have developed an interpretation of the Eigen equation $A\hat{\mathbb{1}} = \hat{\mathbb{1}}$ as a conservation of the row sums of A; likewise the Eigen equation $A^T \hat{\mathbb{1}} = \hat{\mathbb{1}}$ can be interpreted as the conservation of the column sums of A. The dual definitions for the Lie group and algebra follow natural.

Definition 4. Let $St^T(\hat{\mathbb{1}})$ denote the dual stochastic Lie group of invertible matrices whose transpose is stochastic with respect to $\hat{\mathbb{1}}$.

Definition 5. Let $\mathfrak{st}^T(\hat{\mathbb{1}})$ denote the dual stochastic Lie algebra of $St^T(\hat{\mathbb{1}})$.

Thus if C_{ij} are generators of $\mathfrak{st}(\hat{\mathbb{1}})$ then $C_{ij}^T = \left(\hat{e}_j - \hat{e}_i\right) \otimes \hat{e}_i$ are generators of $\mathfrak{st}^T(\hat{\mathbb{1}})$. That $St(\hat{\mathbb{1}}) \cap St^T(\hat{\mathbb{1}})$ is a Lie group and $\mathfrak{st}(\hat{\mathbb{1}}) \cap \mathfrak{st}^T(\hat{\mathbb{1}})$ is a Lie algebra will be used extensively in the next section.

2.2 Doubly Stochastic Matrices

Doubly stochastic matrices require row and column conservation of the vector $\hat{\mathbb{1}}$, in the sense that both $A\hat{\mathbb{1}} = \hat{\mathbb{1}}$ and $A^T\hat{\mathbb{1}} = \hat{\mathbb{1}}$ must hold. The group of invertible doubly stochastic matrices is then a subgroup of the group of stochastic matrices. The two constraints of row and column conservation leaves only $(n-1)^2$ linear degrees of freedom. This is an important clue in the construction of a canonical representation. In fact the representation can be found by choosing one additional vector $\hat{e}_n$, from the basis constructed in the previous section, to center the combinatorial construction of the generators of the algebra around. This vector plays a similar role to the diagonal in the previous construction and is used to balance the row and column sums back to zero. As in the previous section we start with a foundational definition.

Definition 6. Let $St(\hat{\mathbb{1}},\hat{\mathbb{1}})$ denote the doubly stochastic Lie group of invertible matrices A such that both A and A^T are stochastic with respect to $\hat{\mathbb{1}}$

We can immediately observe with out proof that $St(\hat{\mathbb{1}},\hat{\mathbb{1}}) = St(\hat{\mathbb{1}}) \cap St^T(\hat{\mathbb{1}})$; leading to the next definition.

Definition 7. Let $\mathfrak{st}(\hat{\mathbb{1}},\hat{\mathbb{1}})$ denote the doubly stochastic Lie algebra of $St(\hat{\mathbb{1}})$.

Again, it should be clear that $\mathfrak{st}(\hat{\mathbb{1}},\hat{\mathbb{1}}) = \mathfrak{st}(\hat{\mathbb{1}}) \cap \mathfrak{st}^T(\hat{\mathbb{1}})$. The implication being that $\mathfrak{st}(\hat{\mathbb{1}},\hat{\mathbb{1}})$ is the algebra of all matrices A such that $\hat{\mathbb{1}}$ is in the kernel of both A and A^T.

We can then find canonical generators of $\mathfrak{st}(\hat{\mathbb{1}},\hat{\mathbb{1}})$ by similar methods as in the previous section. Given a constructed basis $\hat{e}_i$ such that $\langle \hat{e}_i, \hat{\mathbb{1}} \rangle = \frac{1}{\sqrt{n}}$ we pick a single arbitrary element from the basis, say $\hat{e}_n$, the last element for example. We then balance a transition rate from i to j, with the reverse rates from i to n and n to j, and completing the circuit with any entry for n to n, yielding the matrix $C_{ijn} = C_{ij} + C_{ni} + C_{jn}$.

The matrices C_{ijn} are elements of $\mathfrak{st}(\hat{\mathbb{1}})$. Furthermore they are a closed set with respect to matrix transposition, because $C^T_{ijn} = C_{jin}$. The algebra $\mathfrak{st}(\hat{\mathbb{1}})$ is isomorphic to the space of $(n-1)\times(n-1)$ matrices, which can be seen by the relationships, for any $i, j \leq n-1$.

$$-C_{iin} = \hat{e}_i \otimes \hat{e}_i - \hat{e}_i \otimes \hat{e}_n - \hat{e}_n \otimes \hat{e}_i + \hat{e}_n \otimes \hat{e}_n$$

$$C_{ijn} - C_{iin} - C_{jjn} = \hat{e}_i \otimes \hat{e}_j - \hat{e}_i \otimes \hat{e}_n - \hat{e}_n \otimes \hat{e}_j + \hat{e}_n \otimes \hat{e}_n$$

The intuition being that any $n \times n$ matrix with fixed row and column sums can be created by starting with any matrix and appending a compensating n row and n column. These relationships are implicitly used extensively in proving the following lemma and corollary on the products, commutators, and structure constants of C_{ijn}.

Lemma 5.

$$C_{ijn}C_{kln} = \begin{cases} C_{jin} - 2C_{ijn} & i = k \text{ and } j = l, \\ -\left(C_{ijn} + C_{jin}\right) & i = l \text{ and } j = k, \\ C_{jin} - C_{iln} - C_{iin} - C_{jjn} + C_{lln} & i = k, \\ C_{jkn} - C_{iin} - C_{jjn} - C_{kkn} & i = l, \\ C_{iln} - C_{jln} - C_{ijn} & j = k, \\ C_{iin} - C_{jjn} - C_{kkn} + C_{jkn} - C_{ijn} & j = l, \\ C_{jjn} & \text{otherwise}. \end{cases}$$

Proof. We proceed by calculating the terms of the products and then simplifying the cases; assuming $i \neq j$, $k \neq l$, and $i, j, k, l \neq n$.

1. Term wise computation of the Kronecker products yields

$$\begin{aligned}
C_{ijn}C_{kln} = &\left(\hat{e}_i \otimes \hat{e}_j - \hat{e}_i \otimes \hat{e}_i + \hat{e}_n \otimes \hat{e}_i - \hat{e}_n \otimes \hat{e}_n + \hat{e}_j \otimes \hat{e}_n - \hat{e}_j \otimes \hat{e}_j\right) \\
&\cdot (\hat{e}_k \otimes \hat{e}_l - \hat{e}_k \otimes \hat{e}_k + \hat{e}_n \otimes \hat{e}_k - \hat{e}_n \otimes \hat{e}_n + \hat{e}_l \otimes \hat{e}_n - \hat{e}_l \otimes \hat{e}_l) \\
= &\delta_{jk}\hat{e}_i \otimes \hat{e}_l - \delta_{jk}\hat{e}_i \otimes \hat{e}_k + \delta_{jl}\hat{e}_i \otimes \hat{e}_n - \delta_{jl}\hat{e}_i \otimes \hat{e}_l \\
&-\delta_{ik}\hat{e}_i \otimes \hat{e}_l + \delta_{ik}\hat{e}_i \otimes \hat{e}_k - \delta_{il}\hat{e}_i \otimes \hat{e}_n + \delta_{il}\hat{e}_i \otimes \hat{e}_l \\
&\delta_{ik}\hat{e}_n \otimes \hat{e}_l - \delta_{ik}\hat{e}_n \otimes \hat{e}_k + \delta_{il}\hat{e}_n \otimes \hat{e}_n - \delta_{il}\hat{e}_n \otimes \hat{e}_l \\
&-\hat{e}_n \otimes \hat{e}_k + \hat{e}_n \otimes \hat{e}_n + \hat{e}_j \otimes \hat{e}_k - \hat{e}_j \otimes \hat{e}_n \\
&-\delta_{jk}\hat{e}_j \otimes \hat{e}_l + \delta_{jk}\hat{e}_j \otimes \hat{e}_k - \delta_{jl}\hat{e}_j \otimes \hat{e}_n + \delta_{jl}\hat{e}_j \otimes \hat{e}_l \\
= &C_{jk} - C_{jn} + C_{nk} \\
&+\delta_{ik}\left(C_{nl} - C_{il} - C_{ni}\right) - \delta_{il}\left(C_{in} + C_{ni}\right) \\
&+\delta_{jk}\left(C_{il} - C_{ij} - C_{jl}\right) + \delta_{jl}\left(C_{in} - C_{ij} - C_{jn}\right)
\end{aligned}$$

2. The cases follow from simplifying the δ functions; starting with $i = k$ and $j = l$

$$\begin{aligned}
C_{ijn}C_{ijn} &= C_{ji} - 2C_{jn} + C_{nj} - 2C_{ij} + C_{in} \\
&= C_{jin} - 2C_{ijn}
\end{aligned}$$

3. When $i = l$ and $j = k$

$$\begin{aligned}
C_{ijn}C_{jin} &= C_{nj} - C_{jn} - C_{in} - C_{ni} - C_{ij} - C_{ji} \\
&= -\left(C_{ijn} + C_{jin}\right)
\end{aligned}$$

4. When $i = k$

$$\begin{aligned}
C_{ijn}C_{iln} &= C_{ji} - C_{jn} + C_{ni} + C_{nl} - C_{il} - C_{ni} \\
&= C_{jin} - C_{iln} - C_{iin} - C_{jjn} + C_{lln}
\end{aligned}$$

5. When $i = l$

$$\begin{aligned}
C_{ijn}C_{kin} &= C_{jk} - C_{jn} + C_{nk} - C_{in} - C_{ni} \\
&= C_{jkn} - C_{iin} - C_{jjn} - C_{kkn}
\end{aligned}$$

6. When $j = k$

$$\begin{aligned} C_{ijn}C_{jln} &= C_{nj} - C_{jn} + C_{il} - C_{ij} + C_{jl} \\ &= C_{iln} - C_{jln} - C_{ijn} \end{aligned}$$

7. When $j = l$

$$\begin{aligned} C_{ijn}C_{kjn} &= C_{ij} + 2C_{jk} + C_{nj} - C_{jn} - C_{ik} \\ &= C_{iin} - C_{jjn} - C_{kkn} + C_{jkn} - C_{ijn} \end{aligned}$$

8. When none of the conditions apply

$$\begin{aligned} C_{ijn}C_{kln} &= C_{jk} - C_{jn} + C_{nk} \\ &= C_{jjn} \end{aligned}$$

□

Moving immediately to the commutators we have

Corollary 3.

$$[C_{ijn}, C_{kln}] = \begin{cases} 0 & i = k \textit{ and } j = l, \\ 0 & i = l \textit{ and } j = k, \\ C_{ijn} + C_{jin} - C_{iln} - C_{lin} - 2C_{jjn} - 2C_{lln} & i = k, \\ C_{jkn} - C_{kjn} - C_{ijn} - C_{kin} - C_{iin} - C_{jjn} - C_{kkn} & i = l, \\ C_{iln} - C_{lin} - C_{ijn} - C_{jln} - C_{iin} - C_{jjn} - C_{lln} & j = k, \\ C_{jkn} + C_{kjn} - C_{ijn} - C_{jin} + 2C_{iin} - 2C_{kkn} & j = l, \\ C_{jjn} - C_{lln} & \textit{otherwise}. \end{cases}$$

Proof. We work case wise through the equalities; assuming $i \neq j$, $k \neq l$, and $i, j, k, l \neq n$.

1. Starting with $i = k$ and $j = l$

$$\begin{aligned}\left[C_{ijn}, C_{ijn}\right] &= C_{ijn}C_{ijn} - C_{ijn}C_{ijn} \\ &= 0\end{aligned}$$

2. When $i = l$ and $j = k$

$$\begin{aligned}\left[C_{ijn}, C_{jin}\right] &= C_{ijn}C_{jin} - C_{jin}C_{ijn} \\ &= -\left(C_{ijn} + C_{jin}\right) + \left(C_{jin} + C_{ijn}\right) \\ &= 0\end{aligned}$$

3. When $i = k$

$$\begin{aligned}\left[C_{ijn}, C_{iln}\right] &= C_{ijn}C_{iln} - C_{iln}C_{ijn} \\ &= C_{ijn} + C_{jin} - C_{iln} - C_{lin} - 2C_{jjn} - 2C_{lln}\end{aligned}$$

4. When $i = l$

$$\begin{aligned}\left[C_{ijn}, C_{kin}\right] &= C_{ijn}C_{kin} - C_{kin}C_{ijn} \\ &= C_{jkn} - C_{kjn} - C_{ijn} - C_{kin} - C_{iin} - C_{jjn} - C_{kkn}\end{aligned}$$

5. When $j = k$

$$\begin{aligned}\left[C_{ijn}, C_{jln}\right] &= C_{ijn}C_{jln} - C_{jln}C_{ijn} \\ &= C_{iln} - C_{lin} - C_{ijn} - C_{jln} - C_{iin} - C_{jjn} - C_{lln}\end{aligned}$$

6. When $j = l$

$$\begin{aligned}\left[C_{ijn}, C_{kjn}\right] &= C_{ijn}C_{kjn} - C_{kjn}C_{ijn} \\ &= C_{jkn} + C_{kjn} - C_{ijn} - C_{jin} + 2C_{iin} - 2C_{kkn}\end{aligned}$$

7. When none of the conditions apply

$$\begin{aligned}\left[C_{ijn}, C_{kln}\right] &= C_{ijn}C_{kln} - C_{kln}C_{ijn} \\ &= C_{jjn} - C_{lln}\end{aligned}$$

□

Restricting to the matrices where both the row and column sums are zero, which is equivalent to demanding the conservation of the transition rates, or infinitesimal flows of probability, introduces a significant degree of complexity to the algebra. In particular demanding that all the transition rates be balanced by transitions through $\hat{e}_n$ means that only the simplest two and three state processes have easily calculable algebras. Nevertheless, the result makes the sibling theorem accessible.

Theorem 2. *C_{ijn} are canonical generators of $\mathfrak{st}(\hat{\mathbb{1}}, \hat{\mathbb{1}})$.*

Proof. The proof proceeds in the same manner as the proof of the sibling theorem in the previous section.

1. As discussed before $\mathfrak{st}(\hat{\mathbb{1}}, \hat{\mathbb{1}})$ is an $(n-1)^2$ dimensional vector space.
2. By construction there are only $(n-1)^2$ matrices C_{ijn} for a fixed choice of $\hat{e}_n$.
3. Through induction the matrices C_{ijn} are linearly independent for a fixed choice of $\hat{e}_n$.
4. Thus the matrices C_{ijn}, for a fixed choice of $\hat{e}_n$, are a basis for $\mathfrak{st}(\hat{\mathbb{1}}, \hat{\mathbb{1}})$.
5. By the previous lemma the commutators of matrices C_{ijn} are linear combinations of themselves.
6. It follows then that the smallest algebra that contains the matrices C_{ijn} is $\mathfrak{st}(\hat{\mathbb{1}}, \hat{\mathbb{1}})$.

□

As with the stochastic Lie algebra the generators of the doubly stochastic Lie algebra are not unique, not only do they depend on the choice of the basis $\hat{e}_i$ but also on the choice of the basis element $\hat{e}_n$ used to sum the rows and columns to zero.

Chapter 3

Padé Approximation of Derivatives of the Exponential Map

3.1 The Gradient

3.1.1 Algebraic Formulation

3.1.2 Algorithm

3.2 The Hessian

3.2.1 Algebraic Formulation

3.2.2 Algorithm

Chapter 4

Maximum Likelihood Estimation from First Hitting Times

4.1 Distribution of First Hitting Times

4.2 The Likelihood and Its Maximization

4.3 Newton-Raphson Maximization

4.3.1 Formulation

4.3.2 Algorithm

Chapter 5

Conclusion

5.1 Summary of Results

In chapter two we reversed the normal development of stochastic matrices; which usually starts with characterizing matrices as having non-negative entries with fixed row sums in a standard orthonormal basis $\hat{e}_i$. The line of typical development then notices that the vector $\vec{\mathbb{1}} = \sum_{i=1}^{n} \hat{e}_i$ is an Eigenvector. Instead we began by characterizing all invertible matrices A such that $A\hat{\mathbb{1}} = \hat{\mathbb{1}}$ with respect to a fixed unit vector $\hat{\mathbb{1}}$. We showed that there is always a basis $\hat{e}_i$ such that $\vec{\mathbb{1}} = \sqrt{n}\hat{\mathbb{1}}$ can be interpreted as the row sum vector, and found that this allowed us to characterize both the Lie group in which the matrices $A\hat{\mathbb{1}} = \hat{\mathbb{1}}$ reside, and the Lie algebra tangent to the Lie group. We denoted these the $St(\hat{\mathbb{1}})$ stochastic Lie group with respect to $\hat{\mathbb{1}}$ and $\mathfrak{st}(\hat{\mathbb{1}})$ the stochastic Lie group with respect to $\hat{\mathbb{1}}$. We further characterized the doubly stochastic Lie group and algebra with respect to $\hat{\mathbb{1}}$, denoted $St(\hat{\mathbb{1}}, \hat{\mathbb{1}})$ and $\mathfrak{st}(\hat{\mathbb{1}}, \hat{\mathbb{1}})$, of invertible matrices of fixed row and column sums with respect to $\hat{\mathbb{1}}$, by generalizing the stochastic Lie group and algebra to the dual stochastic Lie group and algebra, $St^T(\hat{\mathbb{1}})$ and $\mathfrak{st}^T(\hat{\mathbb{1}})$, of invertible matrices such that $A^T\hat{\mathbb{1}} = \hat{\mathbb{1}}$.

5.2 Discussion

Application of Lie Theory to the embedding problem for first hitting times... Differentiate between problem of choosing a branch of the matrix logarithm and multiple Markov models having the same first hitting time distribution. Once a principle branch of the logarithm is fixed stochastic Lie algebra can give meaning to the idea of a simplest model, the one expressed in the fewest canonical generators.

Appendix A

Julia Implementations

Code dumps of implementations of the algorithms in Julia.