UNIVERSITY OF CALGARY

The Application of Lie Algebras 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 and Lie algebras 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 algebra theory will fruitfully yield novel computational methods for estimation problems in continuous time Markov processes. In this project we derive the minimal Lie algebra that contains the generators of a continuous time Markov process, and then using the guarantees of algebraic closure construct a Newton-Raphson algorithm for maximum likelihood estimation of the generator of a continuous time Markov process 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

Chapter 1

Background

1.1 Continuous Time Markov Process

1.1.1 Stopped Processes

1.1.2 Phase Type Distributions and Birth Death Processes

1.1.3 Baum-Welch and Viterbi Algorithms

1.2 Lie Algebras and the Exponential Map

1.2.1 Lie Groups

1.2.2 Lie Algebras

1.2.3 The Exponential Map

Chapter 2

The Lie Algebra of the Generators of Continuous Time Markov Processes

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 subsections we provide an explicit construction and characterization building on the original the work of ??.

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

In 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 we a few more foundational definitions.

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

This definition immediately necessitates proof of the claim 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 Lie algebra of $St(\hat{\mathbb{1}})$

We can immediately fully characterize this algebra as the set of matrices such 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:

   Thus $A \in \mathfrak{st}(\hat{\mathbb{1}})$

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 $A\hat{\mathbb{1}} = 0$

$\square$

Over an n dimensional vector space, the condition on the 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 of $\hat{\mathbb{1}}$.

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

Proof. Invoke Gram-Schmidt orthogonalization $\square$

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

Theorem 1. *The canonical generators of* $\mathfrak{st}(\hat{\mathbb{1}})$ *are* $C_{ij} = \frac{1}{\sqrt{2}}\hat{e}_i \otimes \left(\hat{e}_j - \hat{e}_i\right)$

Proof. Show that the smallest containing algebra is the whole algebra. □

The previous theorem serves as the definition of the canonical generators of $\mathfrak{st}(\hat{\mathbb{1}})$. The first result from this theorem is the ability to calculate the structure constants of the generators of the algebra. We proceed by studying the products of the generators.

Corollary 1. $C_{ij}C_{kl} = \delta$

Corollary 2. $\exp\left(C_{ij}\right) = e$

2.2 Doubly Stochastic Matrices

Doubly stochastic matrices require double conservation of the vector $\hat{\mathbb{1}}$, leaving 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 $\hat{e}_n$ to "omit". 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. $St(\hat{\mathbb{1}}, \hat{\mathbb{1}})$ and $\mathfrak{st}(\hat{\mathbb{1}}, \hat{\mathbb{1}})$

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

Appendix A

Julia Implementations

Code dumps of implementations of the algorithms in Julia.