

# Inference Methods for First Few Hundred Studies

James Nicholas Walker

*Thesis submitted for the degree of*

*Master of Philosophy*

*in*

*Applied Mathematics and Statistics*

*at*

*The University of Adelaide*

*(Faculty of Engineering, Computer and Mathematical Sciences)*

School of Mathematical Sciences



November, 2015

# Contents

|  |            |
|--|------------|
| <b>Abstract</b>                              | <b>vii</b> |
| <b>Signed Statement</b>                      | <b>ix</b>  |
| <b>Acknowledgements</b>                      | <b>x</b>   |
| <b>1 Introduction</b>                        | <b>1</b>   |
| 1.1 Epidemic Models and Thresholds . . . . . | 2          |
| 1.2 Inference for Epidemics . . . . .        | 4          |
| 1.3 Thesis Structure . . . . .               | 6          |
| <b>2 Technical Background</b>                | <b>7</b>   |
| 2.1 Markov Chain Theory . . . . .            | 7          |
| 2.2 Stochastic Epidemic Models . . . . .     | 9          |
| 2.2.1 SIR Model . . . . .                    | 9          |
| 2.2.2 SEIR Model . . . . .                   | 10         |
| 2.2.3 Household Epidemic Models . . . . .    | 10         |
| 2.3 Statistical Inference . . . . .          | 13         |
| 2.4 Poisson Process . . . . .                | 15         |
| 2.5 Path Integrals . . . . .                 | 17         |
| 2.6 Laplace Transforms . . . . .             | 17         |
| <b>3 Inference for the SIR Model</b>         | <b>20</b>  |
| 3.1 General Framework . . . . .              | 20         |
| 3.2 Expectation Approximation . . . . .      | 28         |
| 3.3 Distribution Approximation . . . . .     | 38         |
| 3.3.1 Specific Path Probabilities . . . . .  | 41         |

|          |  |           |
|----------|--|-----------|
| 3.3.2    | Laplace Transform Corresponding to a Specific Path . . . . .     | 43        |
| 3.3.3    | Path Integral Distribution Calculations . . . . .                | 47        |
| <b>4</b> | <b>Inference for the SEIR Model</b>                              | <b>56</b> |
| 4.1      | Expectation Approximation . . . . .                              | 63        |
| 4.2      | Distribution Approximation . . . . .                             | 71        |
| 4.2.1    | Newly Infectious Household Path Integral Distribution . . . . .  | 72        |
| 4.2.2    | Likelihood Calculations . . . . .                                | 76        |
| <b>5</b> | <b>Discussion</b>  | <b>80</b> |
| <b>A</b> | <b>Proofs</b>  | <b>86</b> |
| A.1      | Poisson and Binomial Mixture . . . . .                           | 86        |
| A.2      | Convolution of Independent Poisson<br>Random Variables . . . . . | 86        |
|          | <b>Bibliography</b>  | <b>88</b> |

# List of Tables

|     |  |    |
|-----|--|----|
| 2.1 | Transitions and Rates for the SIR Model . . . . .            | 10 |
| 2.2 | Transitions and Rates for the SEIR Model . . . . .           | 10 |
| 2.3 | Transitions and Rates for the SIR Household Model . . . . .  | 11 |
| 2.4 | Transitions and Rates for the SEIR Household Model . . . . . | 12 |

# List of Figures

|      |   |    |
|------|---|----|
| 3.1  | An Example of the SIR Infection Process . . . . .   | 21 |
| 3.2  | Example of Data Grouped into $W_t^i$ . . . . .  | 23 |
| 3.3  | Example of Infection over a Day . . . . .   | 25 |
| 3.4  | SIR State Space Grouped According to the Number of Infections Observed  | 29 |
| 3.5  | A Comparison of SIR Sample Means and Estimated Means . . . . .  | 33 |
| 3.6  | Expected Value SIR MAP Estimates . . . . .  | 36 |
| 3.7  | Posterior Distributions of Various Orders of Expectation Approximations   | 37 |
| 3.8  | Sample Paths of $I(X_t W_{(0:1)} = (1, 3))$ and Corresponding $R_1$ . . . . .   | 39 |
| 3.9  | State Space Modification . . . . .  | 42 |
| 3.10 | Cumulative Distribution Functions of Path Integrals for Specific Paths  | 47 |
| 3.11 | Cumulative Distribution Functions of Path Integrals of the Number of<br>Infectives for Origin Destination Pairs . . . . . | 48 |
| 3.12 | Cumulative Distribution Functions of Path Integrals of the Number of<br>Infectives Within Households . . . . .            | 49 |
| 3.13 | Probability Density Function of the Path Integral of the Number of<br>Infectives in the Population Over Time . . . . .    | 51 |
| 3.14 | Probability Density Function of the Path Integral of the Number of<br>Infectives in the Population Over Time . . . . .    | 52 |
| 3.15 | Distribution Approximation SIR MAP Estimates . . . . .  | 53 |
| 3.16 | Posterior Distributions for Various Orders of Distribution Approximations   | 54 |
| 3.17 | A Comparison of $0^{th}$ Order Posterior Distributions for Each Method . .  | 55 |
| 4.1  | SEIR Household State Space . . . . .  | 57 |
| 4.2  | SEIR Example of Infection over a Day . . . . .  | 58 |
| 4.3  | SEIR Example of Chain of Infectiousness over a Day . . . . .  | 59 |

|      |  |    |
|------|--|----|
| 4.4  | SEIR State Space Grouped According to the Number of Infectious Individuals Observed . . . . .                    | 65 |
| 4.5  | A Comparison of SEIR Sample Means and Estimated Means . . . . .  | 67 |
| 4.6  | Expected Value SEIR Variance and MAP Estimates . . . . .   | 69 |
| 4.7  | Posterior Distributions for Various Orders of SEIR Expectation Approximations . . . . .                          | 70 |
| 4.8  | Distributions of the Path Integral of the Number of Infectious Individuals within a Household . . . . .          | 72 |
| 4.9  | SEIR Distribution of Infection within Newly Infected Households with Uniform Initial Time Distribution . . . . . | 75 |
| 4.10 | SEIR Distribution of Infection within Newly Infected Households with Beta Initial Time Distribution . . . . .    | 75 |
| 4.11 | SEIR Distribution Likelihoods For Varying $D$ . . . . .  | 77 |
| 4.12 | Comparison of SEIR Expected Value and Distribution Methods . . . . .   | 79 |

# Abstract

Infectious diseases are a threat to the health of populations around the world. During the early stages of a novel infectious disease outbreak, intensive data collection may be conducted on the first few hundred symptomatic individuals in what is referred to as a First Few Hundred (FF100) Study. This data is used to assess the potential impact of a pandemic in terms of its transmissibility and clinical severity; these are used to inform a response that is proportionate to the level of risk posed by the disease. Transmissibility is not only determined by the infectiousness of a disease; it is also determined by how individuals interact with one another in a population. Household epidemic models allow for interactions at a household level and interactions at a population level. Hence, they account for some of the inherent structure in a population. Furthermore, data in FF100 studies are collected from all individuals within an infectious individuals' household, hence these models are appropriate when considering FF100 Study data.

This thesis develops new statistical methods, based upon continuous-time Markov chain epidemic models, to estimate the transmissibility of a disease in a population of households. In particular, we use Bayesian inference to estimate posterior distributions for the rate of infection *between* households. The large state space of household models means that conventional methods for calculating likelihoods for Markov chain models are infeasible for this problem and hence motivates the development of the new methods presented in this thesis. These are based on the assumption that between household infections only occur between an infectious household and a household of susceptible individuals. This is a reasonable assumption for the beginning of an outbreak in a large population and allows us to consider the dynamics of infection within households, following the import of infection, independently of each other.

Under the assumption of no secondary introductions into households, the force of infection of newly infected households and hence the likelihood for the number of newly infected households over a day can be calculated as a convolution of the force of infection within each household. Two types of methods for evaluating the likelihood are developed: the first method calculates the expected force of infection; and, the second method calculates the distribution of the force of infection. The expectation method utilises matrix exponential methods on the small, household process. The distribution method utilises recursive methods to calculate the Laplace transform of the force of infection within households, and then numerically inverts the Laplace transform of the distribution corresponding to the product of individual households forces of infection. These methods are compared for accuracy – in terms of point estimates and posterior distributions – and numerical efficiency.

# Signed Statement

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission in my name for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint award of this degree.

I give consent to this copy of my thesis, when deposited in the University Library, being made available for loan and photocopying, subject to the provisions of the Copyright Act 1968.

I also give permission for the digital version of my thesis to be made available on the web, via the Universitys digital research repository, the Library Search and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

SIGNED: ..... DATE: .....

# Acknowledgements

First and foremost I would like to thank my supervisors, Dr. Joshua Ross and Dr. Andrew Black, for the time and effort they put into assisting me with the completion of this thesis. Secondly I would like to thank my friends and family for their support over the last two years. Lastly I would like to thank you, the reader, for spending your time on this thesis; whether you read this thesis in its entirety or exclusively these acknowledgements.