Inference for epidemics on networks

Brock Hermans

Thesis submitted for the degree of

Masters of Philosophy

in

Statistics and Applied Mathematics

at

The University of Adelaide

(Faculty of Engineering, Computer and Mathematical Sciences)

School of Mathematical Sciences

May 14, 2016

Contents

List of Tables

List of Figures

List of Algorithms

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 University's 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

I would like to begin by thanking my supervisors, Dr. Jonathan Tuke and Dr. Joshua Ross, and not just for their help over the past two years. Prior to commencing a Masters of Philosophy, Dr. Tuke and Dr. Ross had given me a wealth of support and guidance in my academic studies. Without fostering my interest in epidemiology I might never have undertaken a Masters of Philosophy, and so for that I am grateful.

I would also like to acknowledge Associate Professor Garique Glonek for his assistance in helping my understanding of multivariate adaptive regression splines, and how I could use them in my thesis.

Finally I would like to acknowledge my family for always supporting me when I needed it: my father and mother, Brian and Terri, for always believing I can do anything I set my mind to; my brother Luke, for always taking a keen interest in what I was doing; and my sister Amy for always trying her hardest to understand what in the world I was talking about. Special mention goes to my partner, Tessa Longstaff, for putting up with me in times of stress, especially over the last two months before submitting my thesis.

Abstract

One of the motivating questions for many epidemiologists is “how quickly or widely will a particular infection spread?" To answer this question, often epidemic models are used to model the spread of a disease, with different epidemic models making different assumptions about the development of the disease; for example, two similar epidemic models might differ in whether they assume that people develop immunity after recovering from the disease. The advantage of these epidemic models is that they can be used to quickly estimate the epidemic model parameters, given an observed outbreak of a disease.

One assumption of most standard epidemic models is that an infectious individual has an equal probability of spreading the disease to any susceptible person in a population. The classical Susceptible-Infective-Recovered (SIR) model is an epidemic model with this assumption, which says that any pair of individuals has an equal probability of having an interaction [1]; when this interaction is between a susceptible and infectious individual, we call this interaction *adequate contact* [1] if the interaction results in a susceptible individual contracting the disease.

This assumption of equal probability of interaction (called homogeneous-mixing) can be a restrictive and unreasonable assumption in situations where within a population there are some pairs of individuals that never interact or some pairs of individuals that interact with higher probability. A more general model considers networks, in which nodes represent people and edges represent a possible path of infection; that is, if Node A (infectious) and Node B (susceptible) don’t share an edge then Node A cannot directly infect Node B. However without this homogeneous-mixing assumption, inference for the epidemic model parameters can be computationally intensive.

This thesis will answer two questions:

1. given observed properties of the network and the final epidemic size(s), can we efficiently estimate the epidemic model parameters; and,

2. given observed properties of the network and known times and types of events, can we efficiently estimate the epidemic model parameters?

We will answer these questions by assuming the data came from the homogeneous-mixing SIR model, and then estimating the epidemic model parameters. We then use a linear model to adjust these estimates, which provides a fast way to estimate epidemic model parameters. We will also show that the error of the models we present are, on average, never greater than 0.2; that is, the estimated epidemic model parameters are on average within 0.2 of the true epidemic model parameters.