On the Advancement of Optimal Experimental Design with Applications to Infectious Diseases

David Price

Thesis submitted for the degree of

Doctor of Philosophy

in

Applied Mathematics and Statistics

at

The University of Adelaide

Faculty of Engineering, Computer and Mathematical Sciences

School of Mathematical Sciences

August, 2015

Contents

Signed Statement

I, David Price, 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 for any other degree or diploma in any university of 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.

The author acknowledges that copyright of published works contained within this thesis resides with the copyright holder(s) of those works.

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.

Signature: ..... *Date*: .....

List of Tables

List of Figures

Acknowledgements

First of all, I would like to thank my supervisors, Dr Jonathan Tuke, Associate Professor Joshua Ross, and Professor Nigel Bean. From encouraging me to even consider further study in the first place, to offering me a project that perfectly tied in the elements of probability and statistics that I enjoyed, to offering me boundless help and advice – with my project or otherwise – whenever needed throughout the past few years, thank you. Thank you not only for your guidance on this project, but also in making the overall experience a fun and memorable one. I might have learned some things about optimal design and maths, but I have gained so much more knowledge from each of you regarding general research practice, teaching, and overall life skills. I can honestly say I would not have gotten to this point without the support that you have each given me. I'd also like to give a brief thank you to Dr Daniel Pagendam, for his correspondence, code and discussion regarding some of his work in optimal design, which helped get me started.

Next, I would like to sincerely thank a number of people that have made my time at University memorable. First, to the office staff at the School of Mathematical Sciences. Thank you for your amazingly friendly demeanour, and help – whether it be with organising teaching, or just getting some new stationery. Thank you to Jess, Kale, Kate, Kyle, Max and Sophie for all the great times at the beginning of my time at university. I'd like to pass on a special thanks to Ben, Mingmei, Nic, and Vincent. Each of you has made what would otherwise be the day-to-day dullness of studying, not only bearable, but a barrel of laughs. Mingmei, thank you for your

every day banter, talk about football or cricket, and your help with my (often) simple questions about code. Nic, thank you for all the memorable quotes and good times. Vincent, thanks for being a *different* pure mathematician, and providing plenty of laughs and distractions. Finally, to Ben. We have, in your words, "dominated stats teaching" together for a while now. With all of the banter and jokes, I can honestly say... it would not have been the same without you.

To my family, and in particular my parents. Thank you for your support in every form, throughout not only my time at university, but the whole 24 years of my life. Finally, thank you to my amazing girlfriend, Claire. For putting up with me when I'm stressed, grumpy, tired, busy, or all of the above; you have been amazing. Thank you for all of your love and support.

Abstract

In this thesis, we investigate the optimal experimental design of some common biological experiments. The theory of optimal experimental design is a statistical tool that allows us to determine the optimal experimental protocol to gain the most information about a particular process, given constraints on resources. We focus on determining the optimal design for experiments where the underlying model is a Markov chain — a particularly useful stochastic model.

Markov chains are commonly used to represent a range of biological systems, for example: the evolution and spread of populations and disease, competition between species, and evolutionary genetics. There has been little research into the optimal experimental design of systems where the underlying process is modelled as a Markov chain, which is surprising given their suitability for representing the random behaviour of many natural processes. While the first paper to consider the optimal experimental design of a system where the underlying process was modelled as a Markov chain was published in the mid 1980's, this research area has only recently started to receive significant attention.

Current methods of evaluating the optimal experimental design within a Bayesian framework can be computationally inefficient, or infeasible. This is due to the need for many evaluations of the posterior distribution, and thus, the model likelihood — which is computationally intensive for most non-linear stochastic processes. We implement an existing method for determining the optimal Bayesian experimental design to a common epidemic model, which has not been considered in a Bayesian

framework previously. This method avoids computationally costly likelihood evaluations by implementing a likelihood-free approach to obtain the posterior distribution, known as Approximate Bayesian Computation (ABC). ABC is a class of methods which uses model simulations to estimate the posterior distribution. While this approach to optimal Bayesian experimental design has some advantages, we also note some disadvantages in its implementation.

Having noted some drawbacks associated with the current approach to optimal Bayesian experimental design, we propose a new method – called ABCdE – which is more efficient, and easier to implement. ABCdE uses ABC methods to calculate the utility of all designs in a specified region of the design space. For problems with a low-dimensional design space, it evaluates the optimal design in significantly less computation time than the existing methods. We apply ABCdE to some common epidemic models, and compare the optimal Bayesian experimental designs to those published in the literature using existing methods. We present a comparison of how well the designs – obtained from each of the different methods – performs when used for statistical inference. In each case, the optimal designs obtained via ABCdE are similar to those obtained via existing methods, and the statistical performance is indistinguishable.

The main applications we consider are concerned with group dose-response challenge experiments. A group dose-response challenge experiment is an experiment in which we expose subjects to a range of doses of an infectious agent or bacteria (or drug), and measure the number that are infected (or, the response) at each dose. These experiments are routinely used to quantify the infectivity or harmful (or safe) levels of an infectious agent or bacteria (e.g., minimum dose required to infect 50% of the population), or the efficacy of a drug. We focus particularly on the introduction of the bacteria *Campylobacter jejuni* to chickens. *C. jejuni* can be spread from animals to humans, and is the species most commonly associated with enteric (intestinal) disease in humans. By quantifying the dose-response relationship of the bacteria in chickens – via group dose-response challenge experiments – we can determine the

safe levels of bacteria in chickens with the aim to minimise, or eradicate, the risk of transmission amongst the flock, and thus, to humans. Thus, accurate estimates of the dose-response relationship are crucial – and can be obtained efficiently by considering the optimal experimental design. However, the statistical analysis of most dose-response experiments assume that the subjects are independent. Chickens engage in copraphagic activity (oral ingestion of faecal matter), and are social animals meaning they must be housed in groups. Thus, oral-faecal transmission of the bacteria may be present in these experiments, violating the independence assumption and altering the measured dose-response relationship. We use a Markov chain model to represent the dynamics of these experiments, accounting for the latency period of the bacteria, and the transmission between chickens. We determine the optimal experimental design for a range of models, and describe the relationship between different model aspects and the resulting designs.