On the Efficient Determination of Optimal Bayesian Experimental Designs using ABC: A Case Study in Optimal Observation of Epidemics

David J. Price[a], Nigel G. Bean[a,b], Joshua V. Ross[a], Jonathan Tuke[a]

[a] *School of Mathematical Sciences, University of Adelaide, SA 5005, Australia*
[b] *ARC Centre of Excellence for Mathematical & Statistical Frontiers*

Abstract

We present a new method for determining optimal Bayesian experimental designs, which we refer to as ABCdE. ABCdE uses Approximate Bayesian Computation to calculate the utility of possible designs. For problems with a low-dimensional design space, it evaluates the designs' utility in less computation time compared to existing methods. We apply ABCdE to stochastic epidemic models. Optimal designs evaluated using ABCdE are compared to those evaluated using existing methods for the stochastic death and susceptible-infectious (SI) models. We present the Bayesian optimal experimental designs for the susceptible-infectious-susceptible (SIS) model using ABCdE.

Keywords: Bayesian optimal design, Stochastic epidemic models, Approximate Bayesian Computation.

Email address: david.j.price@adelaide.edu.au (David J. Price)

1. Introduction

Optimising the design of experiments is an important consideration in many areas of science, including but not limited to: biology (Faller et al. (2003)), chemical engineering (Telen et al. (2012)), clinical trials (Berry (2004)) and epidemiology (Pagendam and Pollett (2013)). 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 model parameters, given constraints on resources.

The aim of this paper is to introduce a new, efficient method of determining optimal Bayesian experimental designs, which we call ABCdE, that uses only simulations from the model. As a demonstration, we provide a comparison of this new method to existing methods. The improvement in efficiency of our method comes about when searching across a low-dimensional design space.

The particular problem we address is when to observe an epidemic process in order to gain the most information about the model parameters. We consider a death process and a susceptible-infectious (SI) epidemic model, previously considered in a Bayesian framework by Cook et al. (2008) and Drovandi and Pettitt (2013), and also a susceptible-infectious-susceptible (SIS) epidemic model, previously considered in the frequentist framework by Pagendam and Pollett (2013). In these examples, a design is considered to be a vector of observation times of length n, where n is the number of observation times, constrained by resources.

Review of Related Work

Let $U(\boldsymbol{\theta}, \boldsymbol{x}, d) \in [0, \infty)$ be a measure of information one would obtain if the experiment were conducted under design d, where $\boldsymbol{\theta}$ is the model parameters and data $\boldsymbol{x}$ is observed. A sensible choice of design d, is one that maximises the expected utility of the experiment $\boldsymbol{E}_{\boldsymbol{\theta},\boldsymbol{x}}[U(\boldsymbol{\theta}, \boldsymbol{x}, d)]$. When the utility $U(\boldsymbol{\theta}, \boldsymbol{x}, d)$ is a function of the posterior distribution in some way – as is the case in this paper – we call this Bayesian optimal experimental design (for a review of Bayesian experimental design theory, see Chaloner and Verdinelli (1995)). To evaluate this expected utility, Müller (1999) proposed treating the expected utility function as an unnormalised, marginal probability density function, by placing a joint distribution on $(\boldsymbol{\theta}, \boldsymbol{x}, d)$. An MCMC scheme was then employed to sample from the design space proportional to the utility function. The optimal design is then the mode of the sampled distribution. Determining the mode of this (possibly) multivariate distribution is complex. Drovandi and Pettitt (2013) for example, chose to use nonparametric techniques to evaluate the mode, however, they note that their approach may not extend well to higher dimensional designs.

The utility $U(\boldsymbol{\theta}, \boldsymbol{x}, d)$ should quantify the information contained in the posterior distribution of the model parameters. One issue that arises in evaluating the expected utility is that we require evaluation of the likelihood in determining the posterior distribution. For partially-observed, non-linear stochastic processes – such as the epidemic models considered in this paper – evaluating the likelihood is often computationally intensive. Even for models where the likelihood is not computationally intensive, we require evaluation of the likelihood at every iteration of the MCMC scheme. Hence, timely

evaluation of these designs quickly becomes infeasible.

Recent work has aimed to avoid the time-consuming evaluation of the exact likelihood function. Cook et al. (2008) employed the MCMC algorithm proposed by Müller (1999), coupled with a moment-closure approximation of the likelihood, allowing a closed-form for – and thus timely evaluation of – the approximate model likelihood. Alternatively, within the algorithm of Müller (1999), Drovandi and Pettitt (2013) looked to avoid likelihood evaluations by using only model simulations to evaluate the posterior distribution, and thus the utility, using Approximate Bayesian Computation (ABC) methods (for an introduction to ABC methods, see Marjoram et al. (2003)). Alternatively, Ryan et al. (2014) utilised indirect inference methods to approximate the posterior distribution within the algorithm of Müller (1999). In each of these methods, the optimal design is determined as the empirical mode of the sampled distribution. The method of Hainy et al. (2013b) also avoids likelihood evaluations, suggesting evaluation of the utility at every design across a grid on the design space. The posterior distribution is once again replaced by the approximate posterior distribution, determined by an ABC method. The expected utility is then approximated using Monte-Carlo integration. The optimal design in this algorithm is then the design corresponding to the largest expected utility.

The algorithm of Müller (1999) is the current standard search algorithm for Bayesian optimal experimental designs, with variations to evaluating the utility (for example, Cook et al. (2008), Drovandi and Pettitt (2013), Ryan et al. (2014), Hainy et al. (2013a)). However, there are some drawbacks to this methodology. The standard issues that plague a Metropolis-Hastings

algorithm also affect the MCMC algorithm here. For example, one must decide on a suitable proposal density for designs, which will govern the rate of convergence to the target density and hence the amount of time the algorithm will take to complete. There is the question of how many samples (designs) are to be accepted in order to determine the utility surface accurately enough, and similarly, the "curse-of-dimensionality", which suggests the chain should run for significantly longer as the number of design parameters increases, in order to ensure the design space has been properly explored. Once a suitable number of samples has indeed been accepted in the Metropolis-Hastings algorithm across the design space, one must then determine the mode of a (possibly) high-dimensional distribution from an approximation: which is not a trivial problem (see Drovandi and Pettitt (2013)).

Our Algorithm

We present a new method of determining Bayesian optimal experimental designs. Our method is similar to Hainy et al. (2013b), however we use our simulation effort more efficiently, thus simultaneously improving on computational efficiency and accuracy. We pre-simulate a large number of realisations N_{pre}, corresponding to parameters sampled from the prior distribution of $\boldsymbol{\theta}$, from the model at each design over a gridded design space. We then use an ABC method with each of the N_{pre} simulated datum under a particular design as our 'observed' data to evaluate the utility. We take an average of these N_{pre} evaluations of the utility as our estimate of the expected utility for that design. This process is repeated for every design. The optimal design is then the design that returns the maximum expected utility. Hence, we are using Approximate Bayesian Computation methods to evaluate the utility

for all designs efficiently, and hence, we refer to this algorithm as ABCdE. The small 'd' is deliberately chosen to represent the efficiency with respect to small design spaces, as we will discuss later.

A particularly attractive feature of the algorithm is that, unlike an MCMC algorithm, it does not rely on previous iterations of the algorithm. This means that ABCdE can easily be implemented in parallel (e.g., using `parfor` rather than `for` in `MATLAB`). Note, we provide `MATLAB` code in the Supplementary Materials to implement the ABCdE method for the Markovian death model, as specified in this paper.

2. Methodology

In this section, we begin by providing some general background to Bayesian optimal experimental design and then detail the current methods. Next, we propose a new method of determining Bayesian optimal experimental designs in an efficient manner, utilising Approximate Bayesian Computational (ABC) methods, which we refer to as ABCdE.

The aim of optimal experimental design is to determine the best experimental setup in order to maximise some utility of the experiment. To achieve this aim, we specify a utility function $U(\boldsymbol{\theta}, \boldsymbol{x}, d)$ representing how we 'value' the experimental design d, chosen from the set of all designs $\mathcal{D}$, where $\boldsymbol{\theta}$ is the model parameters and $\boldsymbol{x}$ is the data. We are interested in the expected utility of using design d, over the unknown model parameters and data. That is, we wish to evaluate,

$$u(d) = E_{\boldsymbol{\theta},\boldsymbol{x}}[U(\boldsymbol{\theta}, \boldsymbol{x}, d)] = \int_{\boldsymbol{x}} \int_{\boldsymbol{\theta}} U(\boldsymbol{\theta}, \boldsymbol{x}, d) p(\boldsymbol{x} \mid \boldsymbol{\theta}, d) p(\boldsymbol{\theta}) d\boldsymbol{\theta} d\boldsymbol{x}, \qquad (1)$$

where $p(\boldsymbol{x} \mid \boldsymbol{\theta}, d)$ is the likelihood function of the unobserved data, under design d, and $p(\boldsymbol{\theta})$ is the prior distribution of the model parameters. The optimal design d^* maximises the expected utility over the design space $\mathcal{D}$, $d^* = \text{argmax}_{d \in \mathcal{D}} u(d)$.

The utility function we use throughout this work is the Kullback-Leibler divergence (Kullback and Leibler (1951)) from the prior distribution to the posterior distribution,

$$U(\boldsymbol{x}, d) = \int_{\boldsymbol{\theta}} \log \left(\frac{p(\boldsymbol{\theta} \mid \boldsymbol{x}, d)}{p(\boldsymbol{\theta})} \right) p(\boldsymbol{\theta} \mid \boldsymbol{x}, d) d\boldsymbol{\theta}. \tag{2}$$

The choice of utility – the Kullback-Leibler divergence – is one such example of a utility function $U(\boldsymbol{\theta}, \boldsymbol{x}, d)$. However, due to the integration over all values of $\boldsymbol{\theta}$, this utility is independent of $\boldsymbol{\theta}$; hence, we denote the utility $U(\boldsymbol{x}, d)$.

Substituting equation (2) into equation (1), noting that equation (2) is independent of $\boldsymbol{\theta}$, and through repeated use of the law of total probability, we can write $u(d)$ as,

$$\begin{aligned}
u(d) &= \int_{\boldsymbol{x}} \int_{\boldsymbol{\theta}} \left\{ \int_{\boldsymbol{\theta}} \log \left(\frac{p(\boldsymbol{\theta} \mid \boldsymbol{x}, d)}{p(\boldsymbol{\theta})} \right) p(\boldsymbol{\theta} \mid \boldsymbol{x}, d) d\boldsymbol{\theta} \right\} p(\boldsymbol{x} \mid \boldsymbol{\theta}, d) p(\boldsymbol{\theta}) d\boldsymbol{\theta} d\boldsymbol{x} \\
&= \int_{\boldsymbol{x}} \left\{ \int_{\boldsymbol{\theta}} \log \left(\frac{p(\boldsymbol{\theta} \mid \boldsymbol{x}, d)}{p(\boldsymbol{\theta})} \right) p(\boldsymbol{\theta} \mid \boldsymbol{x}, d) d\boldsymbol{\theta} \right\} \int_{\boldsymbol{\theta}} p(\boldsymbol{\theta}, \boldsymbol{x} \mid d) d\boldsymbol{\theta} d\boldsymbol{x} \\
&= \int_{\boldsymbol{x}} \left\{ \int_{\boldsymbol{\theta}} \log \left(\frac{p(\boldsymbol{\theta} \mid \boldsymbol{x}, d)}{p(\boldsymbol{\theta})} \right) p(\boldsymbol{\theta} \mid \boldsymbol{x}, d) d\boldsymbol{\theta} \right\} p(\boldsymbol{x} \mid d) d\boldsymbol{x} \\
&= \int_{\boldsymbol{x}} \int_{\boldsymbol{\theta}} \log \left(\frac{p(\boldsymbol{\theta} \mid \boldsymbol{x}, d)}{p(\boldsymbol{\theta})} \right) p(\boldsymbol{\theta} \mid \boldsymbol{x}, d) p(\boldsymbol{x} \mid d) d\boldsymbol{\theta} d\boldsymbol{x} \\
&= \int_{\boldsymbol{x}} \int_{\boldsymbol{\theta}} \log \left(\frac{p(\boldsymbol{\theta} \mid \boldsymbol{x}, d)}{p(\boldsymbol{\theta})} \right) p(\boldsymbol{x} \mid \boldsymbol{\theta}, d) p(\boldsymbol{\theta}) d\boldsymbol{\theta} d\boldsymbol{x}.
\end{aligned} \tag{3}$$

Unfortunately, analytic evaluation of the expected utility function $u(d)$ can rarely be achieved. Müller (1999) proposed an MCMC sampling scheme

from the joint probability distribution, $h(\boldsymbol{\theta}, \boldsymbol{x}, d) \propto U(\boldsymbol{\theta}, \boldsymbol{x}, d)p(\boldsymbol{x} \mid \boldsymbol{\theta}, d)p(\boldsymbol{\theta})$. Sampling from $h(\boldsymbol{\theta}, \boldsymbol{x}, d)$ in this way allows us to obtain samples from a distribution that is proportional to $u(d)$ by considering the marginal of $h(d, \boldsymbol{\theta}, \boldsymbol{x})$ in d. The approximate optimal experimental design is thus obtained as the mode of the function proportional to $u(d)$, as determined by the samples from the MCMC sampling scheme. The MCMC sampling scheme defined by Müller (1999) is outlined in Appendix A, Algorithm 3. The optimal design is the mode of the sampled distribution. For further details and comments on Algorithm 3, see Müller (1999).

Some utility surfaces can be relatively flat in the region of the mode. To manage this issue, Müller (1999) proposed an alternative algorithm. This algorithm exaggerates the mode of the distribution, thus making identification of the optimal design easier. The proposed algorithm alters Steps 2 and 5 of Algorithm 3 to instead simulate J parameters $\boldsymbol{\theta}^i_j, j = 1, \ldots, J$, and corresponding data $\boldsymbol{x}^i_j, j = 1, \ldots, J$. The utility at the i^{th} iteration is then evaluated as $u^i = \prod_{j=1}^{J} U(\boldsymbol{\theta}^i_j, \boldsymbol{x}^i_j, d^i)$, meaning we sample from a "powered-up" version of $h(\boldsymbol{\theta}, \boldsymbol{x}, d)$.

The standard version of Algorithm 3 requires one set of simulated data at every iteration. Thus, for m iterations, we require m simulations. However, due to the relative flatness of most utility surfaces, the powered-up version is typically employed to exaggerate the mode of the sampled distribution. Hence, a total of $m \times J$ simulations would be required.

If evaluation of the model likelihood, $p(\boldsymbol{x} \mid \boldsymbol{\theta}, d)$, is computationally intensive, or intractable, then the MCMC sampling scheme detailed in Algorithm 3 will be computationally intensive, or impossible. This is a result of the

utility being a function of the posterior distribution, $p(\boldsymbol{\theta} \mid \boldsymbol{x}, d)$, which must then be evaluated in every iteration of the algorithm.

As an alternative, Cook et al. (2008) proposed a moment-closure approximation to the likelihood for one of the models we consider. Details of the moment closure approximation can be found in Krishnarajah et al. (2005). This approximation gives a closed-form for the likelihood, allowing it to be evaluated reasonably quickly.

Another approach to avoid likelihood evaluations was proposed simultaneously by Hainy et al. (2013a) and Drovandi and Pettitt (2013). We focus on the implementation of Drovandi and Pettitt (2013). They take advantage of Approximate Bayesian Computation (ABC) methods to determine the posterior distribution of the model parameters, thus avoiding the need to evaluate the likelihood function. ABC is a simulation based method that avoids evaluation of the likelihood by simulating data from the model with suitably chosen parameters (typically sampled from $p(\boldsymbol{\theta})$), and accepting the parameter value as a sample from the posterior distribution if the simulated data is "close" to the "observed data".

Algorithm 4 (Appendix A) details the ABC algorithm employed in Drovandi and Pettitt (2013) to obtain the ABC posterior distribution, and evaluate the utility required at Steps 2 and 5 of Algorithm 3. We define the discrepancy function $\rho(\boldsymbol{y}, \boldsymbol{x})$ to be some measure of difference between the observed data $\boldsymbol{x}$ and simulated data $\boldsymbol{y}$, and ϵ to be some tolerance that controls how "close" the observed and simulated data need to be in order to accept the corresponding parameter. In Step 3 of Algorithm 4, p is chosen such that the number of samples used to determine the ABC posterior is $\lfloor pN_{pre} \rfloor$ (where $\lfloor \cdot \rfloor$

denotes the floor function). A trade-off exists between accuracy of the posterior sample and the acceptance rate. For further details of ABC methods, the reader is directed to Fearnhead and Prangle (2012).

Drovandi and Pettitt (2013) exploit the typical ABC rejection algorithm by sampling N_{pre} prior parameter values $\boldsymbol{\theta}$, and simulating data $\boldsymbol{y}$ for each parameter value across all designs on a grid across the design space, prior to running Algorithm 3. The pre-simulated data is then stored, and called on when required to evaluate the utility at Steps 2 and 5 of Algorithm 3. This greatly reduces the simulation effort required, at the expense of being memory intensive. Thus, a total of $N_{pre} \times |\mathcal{D}|$ simulations are performed and stored prior to starting the algorithm, and a further $m \times J$ simulations are performed during the MCMC scheme (a total of $N_{pre} \times |\mathcal{D}| + m \times J$ simulations).

Hainy et al. (2013b) proposed a method of determining the Bayesian optimum experimental design, using ABC methods without MCMC. Their method considers every design on a grid – each time simulating a number of 'observed' data, and comparing to another, independent set of 'simulated' data in order to determine a series of posterior distributions to evaluate the utility at that design. Evaluation of the utility is done using Monte Carlo integration. Their algorithm is detailed in Appendix A, Algorithm 5.

For each design d^i, they simulate G sets of data to be used as observed data. To evaluate the ABC posterior distribution, a further H sets of data are simulated for each of the G data sets. Hence, a total of $|\mathcal{D}| \times G \times H$ simulations are performed in Algorithm 5.

The ABCdE Algorithm

We propose a similar approach to finding the optimal design to that of both Hainy et al. (2013b) and Drovandi and Pettitt (2013). For every design d^i across a grid, we sample N_{pre} parameters $\boldsymbol{\theta}^i$ from $p(\boldsymbol{\theta})$, and pre-simulate N_{pre} corresponding data sets $\{\boldsymbol{x}^i \mid \boldsymbol{\theta}^i, i = 1, \ldots, N_{pre}\}$ from $p(\boldsymbol{x} \mid \boldsymbol{\theta}^i, d)$ across that grid.

Our method differs from Drovandi and Pettitt (2013) in that rather than simulating a design, parameter value and corresponding datum at each stage of an MCMC algorithm, we now use only this $N_{pre} \times |\mathcal{D}|$ matrix to evaluate our expected utility across the gridded design space. This also differs from the approach of Hainy et al. (2013b), as we do not simulate *new* data to evaluate our posterior distribution. Instead, we use our $N_{pre} \times |\mathcal{D}|$ matrix of data as both our observed and simulated data – this ensures we save on simulation effort, whilst making sure we obtain the most information from what we have simulated.

Similar to Drovandi and Pettitt (2013) and Hainy et al. (2013b), we use an approximate ABC posterior distribution to evaluate our utility function. However, we choose to use an alternative, more efficient approach to evaluating the posterior distribution to Algorithm 4. Namely, our approach to evaluating the posterior distribution does not require sorting the data in order to find a fixed proportion of samples. The approach taken is detailed in Algorithm 1 (Marjoram et al. (2003)).

For each design, we use each set of the pre-simulated data as the "observed datum" one-by-one, and evaluate the utility using all the N_{pre} data as "simulated data". This creates a set of posterior samples having ob-

Algorithm 1 ABC Algorithm: Fixed tolerance

Input: Observed data $\boldsymbol{x}$, simulated data $\boldsymbol{y} = (\boldsymbol{y}^1, \ldots, \boldsymbol{y}^N)$, corresponding parameter values $\boldsymbol{\theta}^i, i = 1, \ldots, N$, and tolerance ϵ.

1: Evaluate discrepancies $\rho^i = \rho(\boldsymbol{x}, \boldsymbol{y}^i)$, creating particles $\{\boldsymbol{\theta}^i, \rho^i\}$ for $i = 1, \ldots, N$.

2: Using the posterior sample of parameters $\boldsymbol{\theta}^i$ such that $\rho^i < \epsilon$, evaluate utility.

Output: Utility for current design, having observed $\boldsymbol{x}$, $U(d, \boldsymbol{x})$.

served every set of simulated data for a particular design. That is, for simulated data $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_{N_{pre}}$ under design d, we determine ABC posteriors $[\hat{p}(\boldsymbol{\theta} \mid \boldsymbol{x}_1, d), \hat{p}(\boldsymbol{\theta} \mid \boldsymbol{x}_2, d), \ldots, \hat{p}(\boldsymbol{\theta} \mid \boldsymbol{x}_{N_{pre}}, d)]$ using Algorithm 1. Similar to Drovandi and Pettitt (2013), we pre-simulate data across all designs and thus we can pass pre-simulated data and corresponding parameter values to Algorithm 1. This increases memory requirements, but saves on simulation effort, as we do not simulate new parameter values and data each time, as would typically be done in an ABC rejection-algorithm, or as used in Algorithm 5. We evaluate the utility using each of these N_{pre} posterior distributions under a particular design, and take the average of these N_{pre} values to be our measure of the expected utility for that design. The optimal design is then the design that returns the largest expected utility. The full algorithm is outlined in Algorithm 2.

Algorithm 2 ABCdE Algorithm

1: Choose grid over the parameter space for the discrete estimate of the utility, number of simulations N_{pre}, and tolerance ϵ.

2: Sample N_{pre} parameters $\boldsymbol{\theta}$ from $p(\boldsymbol{\theta})$.

3: For each of the N_{pre} parameters, and under every design d in the design space $\mathcal{D}$, simulate process and store $X_{N_{pre}\times|\mathcal{D}|}(\boldsymbol{\theta}, d)$.

4: **for** $i = 1$ to $|\mathcal{D}|$ **do**

5: Consider the unique rows of data $Y(\boldsymbol{\theta}, d^i) =$ unique$(X(\boldsymbol{\theta}, d^i))$.
Note: We let K^i be the number of such unique data, and n_{k^i} be the number of repetitions of the $k^{i^{th}}$ unique data, for $k^i = 1, \ldots, K^i$.

6: **for** $k^i = 1$ to K^i **do**

7: Pass 'observed data' $\boldsymbol{y}^{k^i} = [Y(\boldsymbol{\theta}, d^i)]_{k^i}$, 'simulated data' $X(\boldsymbol{\theta}, d^i)$, N_{pre} sampled parameters, and tolerance ϵ to Algorithm 1, and return contribution $U(\boldsymbol{y}^{k^i}, d^i)$ to the expected utility, for $k^{i^{th}}$ unique datum ('observed data') and i^{th} design.

8: **end for**

9: Store $u(d^i) = \frac{1}{N_{pre}} \sum_{k^i} n_{k^i} U(\boldsymbol{y}^{k^i}, d^i)$; the average utility over all parameters and data for design d^i.

10: **end for**

Output: The optimal design $d^* = \underset{d\in\mathcal{D}}{\text{argmax}}(u(d))$.

A total of $N_{pre} \times |\mathcal{D}|$ simulations are used for the ABCdE algorithm. In order to obtain the same level of accuracy as the ABCdE algorithm, we would need to set $G = H = N_{pre}$ in the ABCD algorithm of Hainy *et al.* [2013b]. Hence, a total of $N_{pre}^2 \times |\mathcal{D}|$ simulations would be required, and hence the run time would significantly increase.

We propose that the number of simulations N_{pre} and ABC tolerance ϵ be chosen in the same way as one would choose the number of simulations and tolerance when using ABC for inference. That is, perform a number of pilot studies prior to running the ABCdE algorithm in order to determine a sensible tolerance level (see, for example, McKinley et al. (2009)).

As our method is based on the evaluation of the ABC posterior distribution, we are required to sample parameter values from the prior distribution. Having obtained these parameter values, we are inherently left with a discrete parameter space (as the prior distribution is discrete). Thus, to evaluate the utility (equation (2)), we evaluate the ratio of the approximate posterior – the accepted parameter values from the ABC scheme – to the sampled prior. Hence, we represent the posterior distribution as a histogram of the accepted parameter values with bins centred at the grid points of the parameter space, and employ discrete Monte-Carlo integration to evaluate the utility.

By considering only the unique data sets at Step 5, we avoid evaluating the same posterior distributions multiple times. For any given set of observed data (e.g., x), the parameters corresponding to the same sets of simulated data will form the ABC posterior (e.g., all θ^i corresponding to y^i s.t. $\rho(x, y^i) < \epsilon$). Hence, we can evaluate one such posterior distribution for each unique data set and re-use this posterior distribution n_{k^i} times. This can greatly reduce the number of calculations required, hence speeding up the algorithm considerably. For example, consider the death model with $N = 50$, and $N_{pre} = 100,000$. Evaluating the posterior distribution for only the unique data will result in the calculation of at most 51 posterior distributions (having observed $0, 1, \ldots, 50$ infectious individuals) – significantly less

calculations than if we were to evaluate the posterior distribution for each of the N_{pre} simulations. This approach does allow one extra sample in each posterior distribution (the value that created the observed data). However, we do not consider this to be an issue as we simulate a large amount of data, and so this does not noticeably alter the resulting posterior distribution. We have implemented our ABCdE algorithm by creating the posterior distribution for each data set having removed the parameter value that created it, and noted there were negligible differences in the resulting optimal designs, but a much greater computation time. Hence, we chose to proceed with the more efficient algorithm. We note that this advantage may only hold for discrete data. While we have not investigated this avenue, it may be possible to discretise continuous data in a sensible way – perhaps taking advantage of the ABC metric – in order to still obtain some improvement in computational efficiency.

Finally, we note that in contrast to the Metropolis-Hastings approach of Müller (1999), Cook *et al.* [2008], and Drovandi and Pettitt (2013), the ABCdE algorithm is not dependent on previous iterations of the algorithm. Hence, we have what is known as an *embarrassingly parallel* problem. That is, it takes little-to-no effort to run the algorithm in parallel. With the recent work into parallel computing, and the introduction of multi-core CPUs, and graphical processing units (GPUs) for parallel computing, current efforts to make such tools more widely accessible to programmers will lead to significant improvements in the efficiency of this algorithm in the near future.

3. Examples

To demonstrate the methodology, we consider three examples concerning stochastic epidemic models. The first two have been considered by Cook et al. (2008) and Drovandi and Pettitt (2013). These will allow us to directly compare the resulting optimal designs and their ability to recover the true model parameters when each is employed. The final model we consider is the Markovian SIS epidemic model. We use a continuous-time Markov chain to model each of the processes, with state space given by the possible numbers of 'infectious' individuals in the system: $\mathcal{S} = \{i : i = 0, 1, 2, \ldots, N\}$. We also note that optimal designs are dependent on the choice of prior distribution, and thus the examples considered here are simply illustrative rather than comprehensive. The approximate frequentist optimal designs considered by Pagendam and Pollett (2013) are the only example of optimal designs for the SIS epidemic model.

Markovian Death Model

Consider the Markovian death model as defined by Cook et al. (2008). We have N individuals in a population. Independently, individuals move to an infectious class I, at constant rate b_1 (e.g., from an environmental source). The number of individuals in the infectious and susceptible classes at time t are given by $I(t)$ and $S(t)$, respectively, with $S(t) = N - I(t)$. The transition rate of the Markov chain is given by, $q_{i,i+1} = b_1(N - i)$ for $i = 0, \ldots, N - 1$. The prior distribution we consider is $b_1 \sim \log N(-0.005, 0.01)$, chosen such that the mean lifetime of individuals in the population is 1, with an approximate variance of 0.01 (as per Cook et al. (2008)).

Markovian SI Epidemic Model

In the Markovian SI epidemic model, the transition rate accounts for the contagious/transmissible nature of infectious diseases. Specifically, b_1 represents the rate at which individuals are exposed via the environmental source, as in the death model, but now we also have transmission between susceptible and infectious individuals at rate b_2. Thus, the transition rate of the Markov chain is given by, $q_{i,i+1} = (b_1 + b_2 i)(N - i)$ for $i = 0, \ldots, N - 1$. The rate b_1 per susceptible, can be thought of as the rate of infection occurring from an external source, and $b_2 i$ the rate of infections per susceptible occurring due to the infectious population.

Prior distributions considered are $b_1 \sim \log N(-3.6, 0.1024)$ and $b_2 \sim \log N(-4.5, 0.16)$ (again, as per Cook et al. (2008)).

Markovian SIS Epidemic Model

Consider now that there is no external source of infection, and that infectious individuals can recover from the infection without immunity, and transition back to the susceptible class. The transition rates for the Markov chain are thus,

$$\begin{aligned} q_{i,i+1} &= \beta \frac{i(N-i)}{N}, & i &= 0, \ldots, N-1, \\ q_{i,i-1} &= \mu i, & i &= 1, \ldots, N, \end{aligned}$$

where β is the effective transmission rate of infection, and μ is the rate of recovery per infectious individual. Due to the high level of correlation between β and μ in the SIS epidemic model (Pagendam and Pollett (2013)), when performing inference we consider estimation of α and ρ, where $\alpha = \beta - \mu$ and $\rho = \mu/\beta$.

We consider independent truncated-normal prior distributions for (α, ρ) with mean $(3, 0.25)$ and variance $(0.0625, 0.0025)$. The parameter spaces are truncated to $\alpha \in (0, 20)$, and $\rho \in (0, 1)$. The lower limits on the parameter space are to ensure non-negativity of the transition rates. The upper limit for ρ is to ensure that the transmission rate β is greater than the recovery rate μ, so that there is a non-zero probability of a major outbreak occurring (Ludwig (1975)). The optimal observation schedule for the SIS epidemic model has only been considered previously in a frequentist framework, by Pagendam and Pollett (2013).

As discussed in Pagendam and Pollett (2013), the SIS epidemic model can be categorised into two main phases: (1) an initial period of drift towards a *quasi-equilibrium* (provided the initial number of infectious individuals differs sufficiently from the expected quasi-equilibrium number of infectious individuals), and (2) fluctuations about this quasi-equilibrium. The rate at which the process drifts towards the quasi-equilibrium (phase (1)), is governed by α, whereas the position of the quasi-equilibrium (phase (2)) is determined by ρ. Hence, an observation during the initial drift phase will provide information predominantly about α, while observation during the quasi-equilibrium phase will provide information predominantly about ρ.

4. Results

The following section provides a comparison of the methods of Cook et al. (2008), Drovandi and Pettitt (2013) and ABCdE, when applied to the death and SI models described in Section 3. We begin by providing the optimal observation schedules determined by each of Cook et al. (2008) and Drovandi

and Pettitt (2013), and compare these designs to those determined using ABCdE. A naïve design is also considered in order to demonstrate the gain in using an optimal design determined by one of the three methods. The naïve designs are chosen by placing equally spaced observation times across the pre-specified design region.

We initially consider up to four observation times for the death model. However, evaluating the optimal experimental design for four observation times using the ABCdE method is inefficient. The ABCdE method performs significantly slower than the existing method of Drovandi and Pettitt (2013). However, we consider the amount of information obtained by making each of one, two, three and four observations, and note that there is not a significant increase in the amount of information obtained by considering four observations, rather than three. Hence, we consider only three observation times for the remaining examples, and the analysis of results. Finally, we provide the Bayesian optimal experimental designs for the SIS epidemic using the ABCdE method, when one, two or three observations are permitted.

We compare the performance of the optimal designs in terms of how well each recovers known model parameters from simulated data, observed at each observation schedule. For the death model, we use an exact posterior distribution, evaluated via a Metropolis-Hastings algorithm (each with a burn-in of 5000, and 20000 accepted samples). The posterior distributions evaluated for the SI and SIS models are evaluated using exact ABC (that is, $\epsilon = 0$), with 2 million and 5 million prior samples, respectively. In each case, the data $\boldsymbol{x}$ is the observed number of infectious individuals at the corresponding design.

For each posterior distribution – arising under each method and each number of observation times – we record parameter estimates (maximum *a posteriori* estimates (MAP)), variances and covariances (where applicable) of the corresponding posterior samples. The MAP is evaluated using `kde` and `kde2d` (Botev et al. (2010)) for the one- and two-parameter models, respectively. The variance and covariances are evaluated directly from the posterior samples.

We note that in each case, the optimal designs from ABCdE are similar to those previously published under the alternative methodologies, and perform just as well as the others in terms of both the MAP and posterior variance.

The ABCdE method requires the design space to be gridded. In order to provide solutions to a similar accuracy to those of Cook et al. (2008) and Drovandi and Pettitt (2013), we choose to use a grid spacing of 0.1, and allow each observation time to be in the range [0.1,6] for the death model, [1,15] for the SI model, and [0.5,10] for the SIS model.

The utility employed by both Cook et al. (2008) and ABCdE is the Kullback-Leibler divergence (equation (2)), whereas Drovandi and Pettitt (2013) use the inverse of the determinant of the posterior covariance matrix. For gridded parameter values $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_l$, we estimate the Kullback-Leibler divergence between the prior distribution and posterior distribution having observed data $\boldsymbol{x}$ under design d as:

$$U(\boldsymbol{x}, d) = \sum_{j=1}^{l} \log \left(\frac{\hat{p}(\boldsymbol{\theta}_j \mid \boldsymbol{x}, d)}{p(\boldsymbol{\theta}_j)} \right) \hat{p}(\boldsymbol{\theta}_j \mid \boldsymbol{x}, d), \tag{4}$$

where $\hat{p}(\boldsymbol{\theta}_j \mid \boldsymbol{x}, d)$ and $p(\boldsymbol{\theta}_j)$ are the ABC posterior probability, and prior probability associated with gridded parameter value $\boldsymbol{\theta}_j$, respectively. The

Expected Kullback-Leibler divergence is then estimated by summing these values over all simulated data $\boldsymbol{x}$.

We employ the same discrepancy function as that of Drovandi and Pettitt (2013), when evaluating the ABC posterior distribution in the ABCdE algorithm. That is, for observed data $\boldsymbol{x} = (x_1, \ldots, x_n)$ and simulated data $\boldsymbol{y} = (y_1, \ldots, y_n)$, under design d – which in these examples corresponds to observation schedule $(t_1, \ldots, t_n)$ – the discrepancy is,

$$\rho(\boldsymbol{x}, \boldsymbol{y} \mid d) = \sum_{i=1}^{n} \frac{|x_i - y_i|}{\text{std}(y_i \mid t_i)},$$

where $\text{std}(y_i \mid t_i)$ is the standard deviation of the simulated data y_i at observation time t_i. Given we pre-simulate all of the data in Algorithm 2, we are able to evaluate the standard deviation of the number of infectious individuals at each observation time prior to running the algorithm (similar to the approach of Drovandi and Pettitt (2013)).

4.1. Death Model

4.1.1. Optimal Designs & their Performance

Table 1 provides the optimal observation schedules as determined by Cook et al. (2008), Drovandi and Pettitt (2013), ABCdE and the naïve designs, for the death process.

For the death model, Cook et al. (2008) used the exact model likelihood. This provides a 'gold-standard' comparison, as no approximations are required (other than the Monte-Carlo error in the posterior of model parameters), and thus should be the closest indication of the true Bayesian optimal experimental designs ability to accurately recover model parameters. The ABCdE algorithm was run with $N_{pre} = 50,000$ simulations, and a tolerance

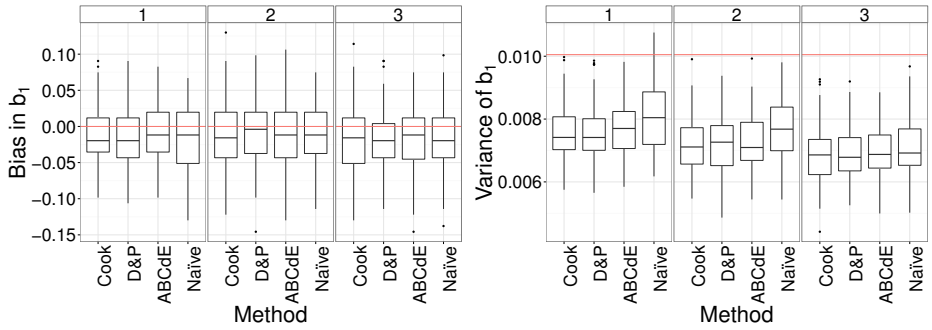
$\epsilon = 0.25, 0.50$ and 0.75 for 1, 2 and 3 observations, respectively. A tolerance of 0.25 corresponds to the data matching exactly, as the largest standard deviation at any observation time is < 4 (i.e., $1/\text{std}(y_i \mid t_i) > 0.25$, $\forall t_i$). The increasing tolerance as the number of observations increases were chosen to account for the change in dimension of the data.

Table 1: Comparison of the optimal observation times for the death process, from Cook et al. (2008), Drovandi and Pettitt (2013) and our ABCdE method. $|t|$ is the pre-determined number of observation times, and i is the i^{th} time.

		Design Method			
$\|t\|$	i	Cook, Gilligan & Gibson	Drovandi & Pettitt	ABCdE	Naïve
1	1	1.70	1.60	1.30	3.15
2	1	0.90	1.15	0.80	2.2
-	2	2.40	3.05	2.80	4.1
3	1	0.70	0.75	0.40	1.725
-	2	1.50	1.90	1.30	3.15
-	3	2.90	3.90	2.60	4.575

Figure 1 demonstrates the fitness of the optimal observation schedules in terms of recovering the true parameter value, using an exact inference method (i.e., a Metropolis-Hastings algorithm with 5000 burn-in and 20000 accepted samples.). We evaluate 100 realisations of the Markov process, under the true (known) parameter value. We evaluate the posterior distributions at each design, for $|t| = \{1, 2, 3\}$, and record the maximum *a posteriori* estimate (MAP) and variance of each posterior sample.

Figure 1a shows boxplots of the bias – the difference between our estimator (MAP) and the true value $b_1 = 1$ – and variances of the posterior distributions recorded for each of the four methods for the death model. There appears to be minimal bias in our estimate of b_1 for each method. Each method appears to have reasonably similar posterior variances (Figure 1b), but as one would expect, the variance decreases (on average) as the number of observations increases. The posterior variances under the naïve design are perhaps marginally worse than the other designs. Finally, note that the posterior variance for each method, and each number of observations, is less than the prior variance of 0.01 (indicated by the red line in Figure 1b). This indicates an improvement in the knowledge about b_1 having conducted the experiment.



(a) Bias in estimate of b_1.

(b) Variance in estimate of b_1.

Figure 1: Bias (a) and variance (b) in estimates for b_1 in the death model. Posterior distributions were evaluated for 100 realisations of the death process, observed at each methods' respective optimal observation schedules, when one, two and three observations were permitted (banner above each subfigure indicates number of observations). The red line in (a) represents zero bias, and in (b) represents the prior variance.

4.1.2. Comparison of Computation Time

Here, we provide a demonstration of the improved efficiency of ABCdE at determining optimal Bayesian experimental designs for problems with a low-dimensional design space. There is no mention of computational time in Cook et al. (2008), while Drovandi and Pettitt (2013) mention that they were able to run their code *"on a high-end desktop PC in a feasible amount of time"*. We run the code supplied by Drovandi and Pettitt (2013) (as is) on the same computer as we have run our ABCdE method, and provide computation times as a comparison/indication of the speed-up in performance of ABCdE. Note that the method of Drovandi and Pettitt (2013) is a MCMC algorithm over the design space with $m = 100,000$ iterations, with no thinning or burn-in.

The ABCdE algorithm was implemented in `MATLAB R2013b`, with the evaluation of the discrepancy coded in a `MEX` function. However, for the purpose of comparing the run time to the method of Drovandi and Pettitt (2013), we present the results when the discrepancy was not coded in a `MEX` function. Timings are recorded from a Macbook Pro, running OSX10.10, with a 2.7GHz Quad-core Intel Core i7 processor, Turbo Boost up to 3.7GHz, and 16GB 1600MHz DDR3L SDRAM.

Table 2 demonstrates the massive gain in efficiency for ABCdE when the design space is relatively small — in this case, less than four observation times. There is a large increase in run-time between three and four observations for ABCdE. Hence, we note that the efficiency of ABCdE is lost when the design space increases – which occurs either by considering a wider-range of designs, or increasing the fineness of the grid over which we

Table 2: Illustration of run-times for Drovandi & Pettitt [2013] algorithm compared to ABCdE.

	Computation Time	
$\|t\|$	Drovandi & Pettitt	ABCdE
1	4.2 hours	0.6 secs
2	10.2 hours	85 secs
3	15.5 hours	3 hours
4	21.3 hours	190 hours

search. However, there are considerable gains in efficiency for one, two and three observation times. Note that the majority of the increase in time can be attributed to the combinatorial nature of the number of designs. Changing the grid spacing will dramatically reduce the computation time.

We note, however, that we do not believe that being restricted to optimal experimental design for small design spaces is a significant drawback in this case. Consider implementing each of the optimal designs for one, two, three or four observations of the Markovian death model, 100 times. For each simulation and each design, we evaluate the utility (Kullback-Leibler divergence) having utilised that experimental procedure. Figure 2 demonstrates the distribution of the utility under each design.

As the number of observations increases, the utility appears to rapidly converge to the maximum information that can be obtained (i.e., that which one would obtain via continuous observation). We performed *multiple comparisons* (using the `agricolae` package (de Mendiburu (2014)), in the statis-

tical software package R (R Core Team (2014)), after performing the relevant transformation of the data), and established that there was no significant increase in the amount of information obtained from four observations, compared to three at the 5% significance level (p-value=1, with Bonferroni correction).

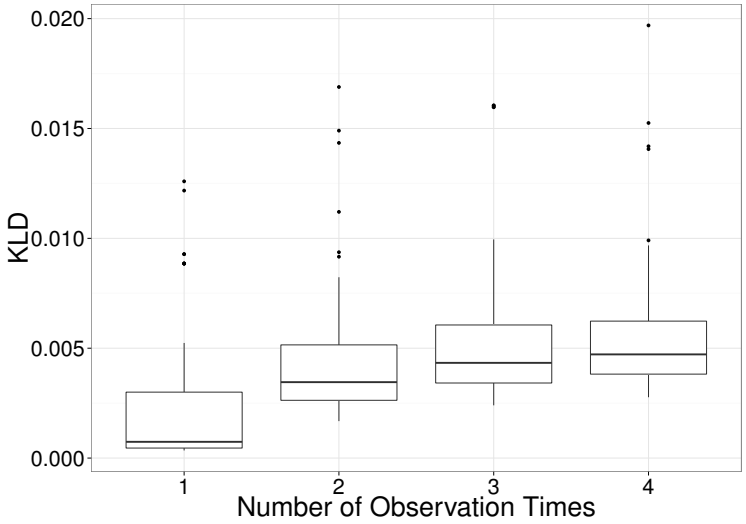


Figure 2: Distribution of Kullback-Leibler divergence for 100 simulations of the Markovian death model, observed at the optimal observation schedule for one, two, three and four observations.

4.2. SI Model

4.2.1. Optimal Designs & their Performance

Table 3 provides the optimal observation schedules as determined by Cook et al. (2008), Drovandi and Pettitt (2013), and those determined by ABCdE as well as the naïve designs, for the SI epidemic process.

For this model, Cook et al. (2008) use the moment-closure approximation to the model likelihood. Hence, our comparisons of the optimal designs contrast those for the death model, as each method is now employing an approximation. That is, there is no 'gold-standard' approach with which to directly compare our results. The ABCdE algorithm used the same tolerances as used for the death model, but with $N_{pre} = 100,000$. More simulations were used to account for the extra model parameter.

Table 3: Comparison of the optimal observation times for the SI epidemic process, from Cook et al. (2008), Drovandi and Pettitt (2013) and our ABCdE method. $|t|$ is the pre-determined number of observation times, and i is the i^{th} time.

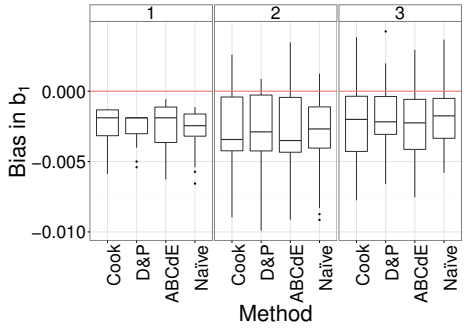
		Design Method			
$\|t\|$	i	Cook, Gilligan & Gibson	Drovandi & Pettitt	ABCdE	Naïve
1	1	9.2	12.1	8.8	8
2	1	4.1	4.6	3.8	5.6
-	2	9.6	12.1	8.8	10.3
3	1	2.9	3.7	1.5	4.5
-	2	7.2	8.7	3.6	8
-	3	10.9	15	9.3	11.5

Figure 3 demonstrates the fitness of the optimal observation schedules in terms of recovering the true parameter value, using an exact ABC (that is, Algorithm 1 with $\epsilon = 0$) with 2×10^6 prior simulations. We evaluate 100 realisations of the Markov process, under the true (known) parameter values. We evaluate the posterior distributions at each design, for $|t| = \{1, 2, 3\}$, and

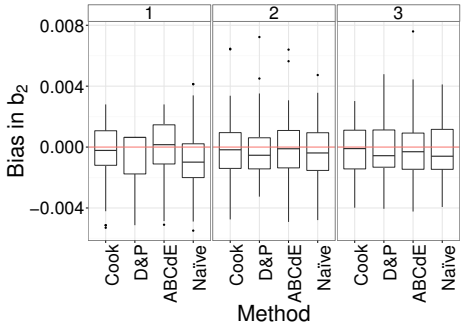
record the MAP, variance and covariance of the parameters (b_1, b_2), for each posterior sample.

Figure 3 shows boxplots of the bias, log of variances and log of covariance of the posterior samples recorded for each of the methods for the SI epidemic model, where $(b_1, b_2) = (0.02875, 0.01203)$. The bias is the difference between our estimator – the MAP estimate – and the true parameter values.

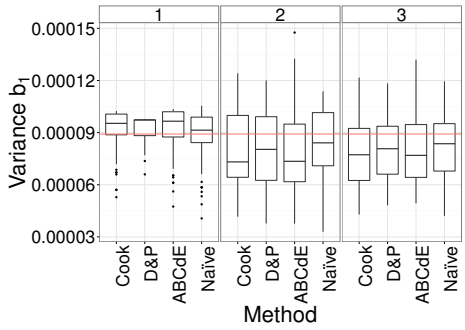
There is an overall negative bias in the MAP estimates of b_1, using each method (Figure 3a). The bias in the MAP estimates of b_2 appear to be roughly centred about zero,for all methods, indicating the correct values are recovered, on average (Figure 3b). The variances of the posterior distribution of b_1 corresponding to each method are all similar in this instance, and lower than the prior variance (on average), for more than two observations. One observation of the SI model appears to result in greater uncertainty about the parameter b_1. The posterior variance for b_2 is less than the prior variance for all numbers of observations, with a decreasing trend as more observations are made. The distribution of the variance of b_2 evaluated at the design for Drovandi and Pettitt (2013) appears to be heavily negatively-skewed, with a median quite close to the prior variance. Conversely, the distribution of the variance of b_2 evaluated at the designs of Cook et al. (2008) and ABCdE appear to be heavily positively-skewed, however the variance is on average considerably lower than the prior variance. The naïve design for one observation appears to perform the best in terms of variances. The posterior covariance is slightly negative for all methods.



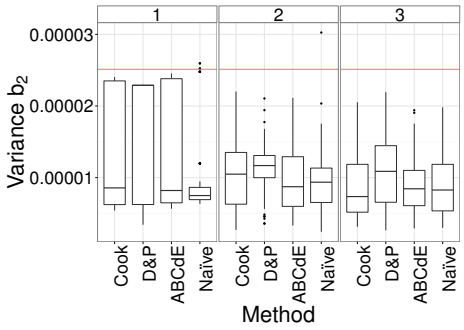
(a) Bias in estimate of b_1.



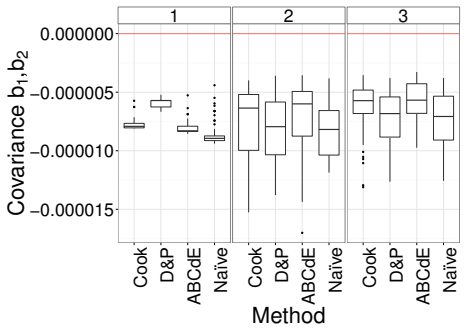
(b) Bias in estimate of b_2.



(c) Variance of estimate of b_1.



(d) Variance of estimate of b_2.



(e) Covariance of estimate of b_1 and b_2.

Figure 3: Bias in estimates of b_1 (a) and b_2 (b), variance of b_1 (c) and b_2 (d), and covariance of b_1 and b_2 (e), of the joint posterior distribution of (b_1, b_2) for the SI model. Posterior distributions were evaluated for 100 realisations of the Markovian SI process, observed at each methods' respective optimal observation schedules, when one, two and three observations were permitted (banner above each subfigure indicates number of observations). The red lines represent zero bias, the prior variance, and zero covariance where appropriate.

4.3. SIS Model

4.3.1. Optimal Designs & their Performance

Table 4 provides the optimal observation schedules for the SIS epidemic process using ABCdE, and a naïve approach. The ABCdE algorithm uses the same tolerances as previous, and the same number of simulations as used for the SI model ($N_{pre} = 100,000$).

Table 4: Optimal observation times for the SIS process, from the ABCdE method and a naïve, equidistant approach. $|t|$ is the pre-determined number of observation times, and i is the i^{th} time

		Method	
$\|t\|$	i	ABCdE	Naïve
1	1	7.2	5.25
2	1	6.0	3.67
-	2	9.3	6.83
3	1	2.3	2.875
-	2	6.0	5.25
-	3	10.0	7.625

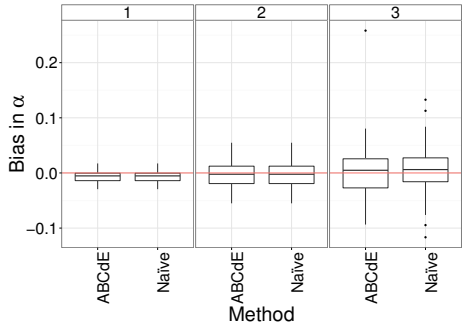
Figure 4 demonstrates the ability of our optimal designs to recover the true model parameters, using an exact ABC (that is, Algorithm 1 with $\epsilon = 0$) with 5×10^6 prior simulations. We simulate the SIS epidemic model 100 times under true (know) parameter values. We compare our optimal design to a naïve design. We evaluate the posterior distributions for each method, and for each $|t| = \{1, 2, 3\}$, and record the MAP, variance and covariance of the

parameters (α, ρ), for each posterior sample, and compare these estimates to the known parameter values.

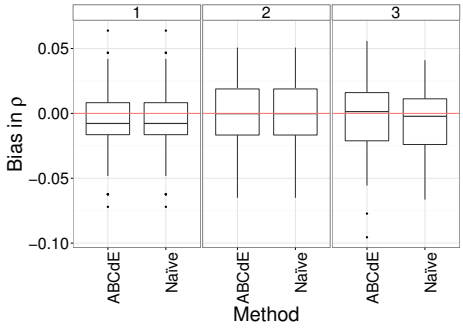
The bias in the MAP estimates of α and ρ appear to be centred about zero for all observation times (Figures 4a and 4b). The variances of the estimate of α appears to be roughly the same (a slight decrease in the median) for each observation time, with a slight increase in variability as the number of observation times increases (Figures 4c and 4d). The variance of α are marginally lower than the prior variance (on average), indicating the relative difficulty of obtaining information about α when there is uncertainty in the model parameters. The variance in the estimates of ρ decrease significantly as the number of observations increases, for both the ABCdE design and the naïve design. The variance is also significantly lower than the prior variance, indicating a significant gain in information about the model parameter ρ.

5. Discussion

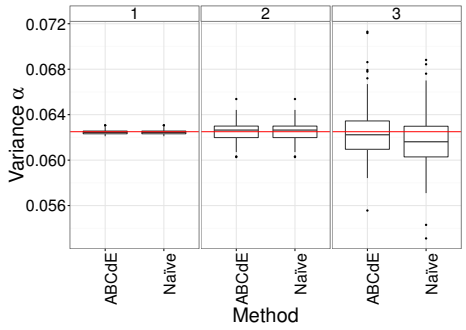
The results of Cook et al. (2008) for the death model are determined using the algorithm of Müller (1999), with the exact model likelihood. This allows a 'gold-standard' comparison, as there are no approximations to the model likelihood used. We can see that the corresponding times for the death model (Table 1) for ABCdE follow the same trend as those determined by the other two methods. The most notable difference being that in each case ($|t| = \{1, 2, 3\}$), the times determined by ABCdE are typically earlier than those of the other methods. We believe this difference to be a result of the use of ABC in conjunction with non-identifiability issues at larger times. ABC methods rely directly on the difference in simulated data (from the



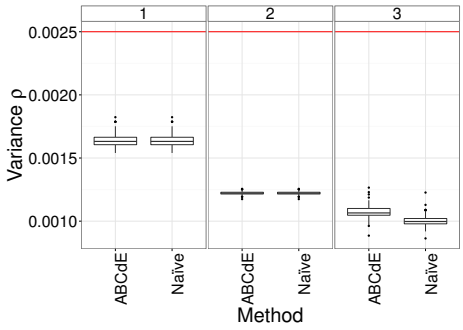
(a) Bias in estimate of α.



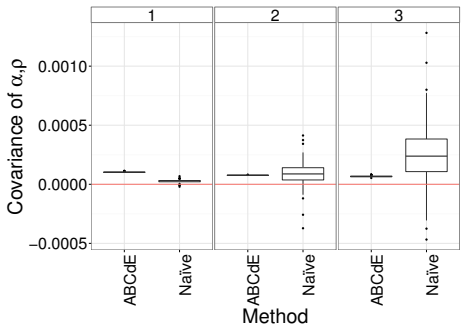
(b) Bias in estimate of ρ.



(c) Variance of estimate of α.



(d) Variance of estimate of ρ.



(e) Covariance of estimate of α and ρ.

Figure 4: Bias in estimates of α (a), and ρ (b), variance of α (c) and ρ (d), and covariance between estimates of α and ρ (e), of the joint posterior distribution of (α, ρ) for the SIS model. Posterior distributions were evaluated for 100 realisations of the Markovian SIS process, observed at each methods' respective optimal observation schedules, when one, two and three observations were permitted (banner above each subfigure indicates number of observations). The red lines represent zero bias, the prior variance, and zero covariance where appropriate.

observed data), in order to build the posterior. As such, later observation times are not as useful, as it becomes more difficult to identify differences in simulated data between different parameter values, for a fixed tolerance. That is, if we observe the process too late, there is a high probability that all individuals have already become infectious, from a wide range of parameter values. Whilst this issue is relevant to all methods, it is more significant in the ABC algorithm.

For the death model, each of the optimally determined designs appears to recover the true parameter value quite well at their respective optimal observation times (Figure 1a), while the naïve design may perform marginally worse. Similarly for the variances of the posterior distributions (Figure 1b). It is important to note that the variance under each method is still significantly lower than the prior variance (≈ 0.01).

The gain in efficiency when determining Bayesian optimal designs via ABCdE comes about when the design space is low-dimensional. The size of the design space is a function of both the number of design parameters being considered, and also the size and resolution of the grid across which you wish to search for the optimal design. If we consider a large design space, our method suffers from the curse-of-dimensionality, worse than the algorithm of Müller (1999). Hence, problems with a low-dimensional design space (that is, a small grid and/or a coarse resolution over which to search for the optimal design), and problems with a small number of unique data sets to consider under each design (that is, a small population size, or simulations that do not vary significantly), will be the most suitable for the ABCdE algorithm. In such cases, massive reductions in computation time

will be achieved, as evidenced by the fractional running times of the ABCdE algorithm in comparison to Drovandi and Pettitt (2013) for the death model (Table 2). We noted however, that in this example, performing more than three observations does not provide significantly more information.

In the examples we consider in this paper, we choose to use the same grid coarseness as in Cook et al. (2008) and Drovandi and Pettitt (2013) to ensure a comparable level of accuracy. We have only presented results for up to three observation times (with the exception of the death model). More observation times would be simple to consider; no alteration to the method needs to be made other than considering a larger number of designs. For the SI and SIS models (Sections 4.2 and 4.3, respectively), evaluating the optimal design for one and two observations was computationally efficient. However, due to the combinatorial increase in the number of designs that must be considered as the number of observation times increases, evaluating the optimal design over a wide grid with the same grid spacing quickly becomes inefficient to evaluate. In order to evaluate the optimal designs for these scenarios in a more computationally efficient manner, a coarser grid may need to be considered. We note that in determining optimal designs in a practical setting, one must take into account the feasibility of the sampling times, and the time-scale of the model. For example, if it is possible to only sample at one time during a day, there is no benefit in specifying a grid so fine we consider the possibility of observing the process at any hour of the day.

The optimal designs for the SI epidemic process are obtained using three different approximations to the model likelihood – the moment-closure ap-

proximation of Cook et al. (2008), the ABC algorithm of Drovandi and Pettitt (2013) and the ABC algorithm detailed in Algorithm 1. We note that the resulting optimal designs obtained via ABCdE follow the same trend as those of the other two methods. Notably, increasing from one to two observation times appears to simply introduce a new observation early on, while keeping the second observation time the same for all three approaches. Once again, we note that the observation times obtained via ABCdE are all earlier than the corresponding observation times determined by the other two methods, for the same reasons as stated previously.

The SI model parameter b_1 (Figure 3a) may be more difficult to estimate, due to the infection events being dominated by transmission (b_2), rather than external infection (b_1), once the process has reached a reasonable number of infectious individuals. This difficulty is also apparent in the moderate improvement observed for the posterior variance of b_1 compared to the prior variance (Figure 3c). Each design appears to perform comparably with regards to bias in estimates of the parameters b_1 and b_2. The average lower posterior variance for b_2 at the one-observation designs of Cook et al. (2008), ABCdE and the naïve approach, are perhaps a result of the significantly earlier observation time compared to the design of Drovandi and Pettitt (2013) – the later observation not allowing identifiability of the parameter when all, or close to all, individuals in the population are already infected by that time. The trade-off between the two sources of infection to balance a 'net infection rate' is apparent in the negative covariance estimate of b_1 and b_2.

Perhaps surprisingly, it appears as though the naïve design performs quite well in this case. However, if we consider the observation schedules in Ta-

ble 3, the naïve designs, which were chosen in an uninformed manner, are reasonably similar to the optimal designs determined using each of the three established methods. Hence, we should not expect them to perform significantly worse in this instance.

Consider now the SIS epidemic model. Note that the utility surface for this model is quite flat. For example, for two observation times, roughly 70% of the considered observation schedules on our grid contained at least 95% of the information (Expected Kullback-Leibler divergence) that was contained in the optimal observation schedule (and 50% of designs contained at least 97.4% of the information). Hence, any observation schedules which lie on the flat surface are going to all perform reasonably well. This is the case with the naïve design used here. Thus, we do not expect to see a large difference in the performance of the naïve design compared to the ABCdE design.

As noted earlier, observations during the early drift phase of the SIS epidemic provide information predominantly about the parameter α, while later observations during the quasi-equilibrium predominantly provide information about ρ. This was discussed in Pagendam and Pollett (2013), when considering frequentist optimal designs. In a frequentist framework, we specify the model parameters that we wish to determine the optimal design for, and so the trajectories of simulated events are reasonably similar. However, as we have a prior distribution on the model parameters, the initial drift phase has a wide range of trajectories it can follow, depending on which parameters (α, ρ) were used to simulate the process. Thus, choosing an observation time early enough to catch the drift phase of *all* simulated epidemics is difficult. Hence, we note that the optimal observation times are much later than the

corresponding frequentist designs would be, if evaluated at the mode of the prior distributions. The difficulty in obtaining information about α is demonstrated in the relatively slight improvement in the variance of α, compared to the improvement seen for ρ (Figures 4c and 4d).

Besides the huge gains in efficiency for low-dimensional design problems, ABCdE has some other attractive features. First, the use of an ABC posterior distribution means it avoids the cumbersome likelihood evaluations. That each design can be considered independently of the others means that ABCdE can be implemented using parallel computing with ease (e.g., using `parfor`, rather than `for`, in `MATLAB`), whereas MCMC techniques are reliant on the previous iteration. Furthermore, there is no need to evaluate possibly high-dimensional multivariate modes of sampling distributions; an issue that was flagged in Drovandi and Pettitt (2013). As ABCdE does not require an MCMC algorithm, there is no issue of convergence, or choosing a suitable proposal density, and similarly, no need to decide a suitable point to define the "burn-in" phase. Also, considering non-uniformly spaced times across the design space does not require any extra effort, as there is no need to specify a proposal distribution across the design space. Finally, by evaluating the utility for all designs, *post-hoc* decisions can be made about which designs to implement. For example, the optimal design may provide only marginally more information than a sub-optimal design, but the sub-optimal design could perhaps be implemented at a fraction of the cost.

Future work is to increase the efficiency of the ABCdE method for problems with large design spaces. One approach is to develop an iterative method, whereby we define a coarse grid over which to first search for a

viable region in which the optimal design resides. Then, a new, finer grid is placed about this region, and the ABCdE algorithm is run to determine a more precise optimal design. This process will be repeated until a suitable level of accuracy is obtained. Furthermore, we are looking at implementing this ABCdE algorithm for sequential designs, where the optimal design is updated after each observation as new information is obtained.

Supplementary Materials

Code to implement ABCdE for the Markovian death model is supplied as supplementary material. The algorithm is supplied as implemented in this paper. The code to simulate the death model was supplied by Drovandi and Pettitt (2013), and was used to ensure consistent simulation effort in the timings.

Acknowledgements

JVR acknowledges the support of the Australian Research Council (Future Fellowship, FT130100254) and the National Health and Medical Research Council (NHMRC Centre of Research Excellence PRISM^2, APP1078068). The work of NGB and JVR is supported by ARC Discovery Project Funding (DP110101929). The authors wish to sincerely thank the reviewers for their insightful feedback and comments.

Appendix A. Existing Algorithms

Algorithm 3 details the MCMC algorithm for determining Bayesian optimal designs proposed by Muller [1999].

Algorithm 3 MCMC with stationary distribution $h(d, \boldsymbol{\theta}, \boldsymbol{x})$, Muller [1999]

Input: Number of samples m, prior distribution of model parameters $p(\boldsymbol{\theta})$, and proposal density $q(\cdot)$.

Choose, or simulate an initial design, d^1.

Sample $\boldsymbol{\theta}^1 \sim p(\boldsymbol{\theta})$, simulate $\boldsymbol{x}^1 \sim p(\boldsymbol{x} \mid \boldsymbol{\theta}^1, d^1)$, and evaluate $u^1 = U(\boldsymbol{\theta}^1, \boldsymbol{x}^1, d^1)$.

for $i = 1 : m$ **do**

    Generate a candidate design, $\tilde{d}$, from a proposal density $q(\tilde{d} \mid d^i)$.

    Sample $\tilde{\boldsymbol{\theta}} \sim p(\boldsymbol{\theta})$, simulate $\tilde{\boldsymbol{x}} \sim p(\boldsymbol{x} \mid \tilde{\boldsymbol{\theta}}, \tilde{d})$, and evaluate $\tilde{u} = U(\tilde{\boldsymbol{\theta}}, \tilde{\boldsymbol{x}}, \tilde{d})$.

    Calculate,

$$\alpha = \min \left\{ 1, \frac{\tilde{u}\ q(d^i \mid \tilde{d})}{u^i\ q(\tilde{d} \mid d^i)} \right\}.$$

    Generate $a \sim U(0, 1)$

    if $a < \alpha$ **then**

        Set $(d^{i+1}, u^{i+1}) = (\tilde{d}, \tilde{u})$

    else

        Set $(d^{i+1}, u^{i+1}) = (d^i, u^i)$

    end if

end for

Output: Sample of m designs, d.

Algorithm 4 details the ABC algorithm for determining the approximate Bayesian posterior distribution for a fixed (minimum) number of samples, detailed in Drovandi & Pettitt [2013].

Algorithm 4 ABC algorithm: Fixed (minimum) number of samples

Input: Observed data $\boldsymbol{x}$, simulated data $\boldsymbol{y} = (\boldsymbol{y}^1, \ldots, \boldsymbol{y}^{N_{pre}})$, corresponding parameters $\boldsymbol{\theta}$, and (minimum) proportion of points to accept p.

1: Evaluate discrepancies $\rho^i = \rho(\boldsymbol{x}, \boldsymbol{y}^i)$, creating particles $\{\boldsymbol{\theta}^i, \rho^i\}$ for $i = 1, \ldots, N_{pre}$.

2: Sort the particles according to the discrepancies ρ^i (such that $\rho^1 \leq \rho^2 \leq \cdots \leq \rho^{N_{pre}}$).

3: Calculate tolerance $\epsilon = \rho^{\lfloor pN_{pre} \rfloor}$.

4: Use the posterior sample of parameters $\boldsymbol{\theta}^i$ such that $\rho^i \leq \epsilon$, to evaluate the utility.

Output: Return utility evaluated for design d, with observed data $\boldsymbol{x}$, $U(d, \boldsymbol{x})$.

Algorithm 5 details the ABCD scheme proposed by Hainy *et al.* [2013b].

Algorithm 5 ABCD Algorithm

Input: Set of designs $\mathcal{D}$, number of posterior distributions to evaluate for each design G, number of samples generated for ABC posterior H, tolerance ϵ controlling the points accepted into posterior distribution.

for $i = 1$ to $|\mathcal{D}|$ **do**
  for $k = 1$ to G **do**
    Sample $\boldsymbol{\theta}^k$ from the prior distribution $p(\boldsymbol{\theta})$.
    Generate an observed datum $\boldsymbol{x}^k$ from $p(\boldsymbol{x} \mid \boldsymbol{\theta}, d^i)$.
    Sample $\{\boldsymbol{y}^j, \boldsymbol{\theta}^j, j = 1, \ldots, H\}$ from $p(\boldsymbol{\theta}, \boldsymbol{x} \mid d^i)$.
    Let $J_\epsilon(k) = \{j : \rho(\boldsymbol{x}^k, \boldsymbol{y}^j) < \epsilon\}$.
    Evaluate the utility for the k^{th} observed datum as,

$$U(\boldsymbol{x}^k, d^i) = \frac{1}{|J_\epsilon(k)|} \sum_{j \in J_\epsilon(k)} U(\boldsymbol{\theta}^j, \boldsymbol{x}^k, d^i).$$

  end for
  Evaluate utility for design d^i as,

$$u(d^i) = \frac{1}{G} \sum_{k=1}^{G} U(\boldsymbol{x}^k, d^i).$$

end for

Output: The optimal design $d^* = \underset{i}{\text{argmax}}(u(d^i))$.

References

D. A. Berry. Bayesian statistics and the efficiency and ethics of clinical trials. *Statistical Science*, 19:175–187, 2004.

Z. I. Botev, J. F. Grotowski, and D. P. Kroese. Kernel density estimation via diffusion. *Annals of Statistics*, 38:2916–2957, 2010.

K. Chaloner and I. Verdinelli. Bayesian experimental design: A review. *Statistical Science*, 10(3):273–304, 1995.

A. R. Cook, G. J. Gibson, and C. A. Gilligan. Optimal observation times in experimental epidemic processes. *Biometrics*, 64(3):860–868, 2008.

F. de Mendiburu. *agricolae: Statistical Procedures for Agricultural Research*, 2014. URL `http://CRAN.R-project.org/package=agricolae`.

C.C. Drovandi and A.N. Pettitt. Bayesian experimental design for models with intractable likelihoods. *Biometrics*, 69:937–948, 2013.

D. Faller, U. Klingmüller, and J. Timmer. Simulation methods for optimal experimental design in systems biology. *Simulation*, 79:717–725, 2003.

P. Fearnhead and D. Prangle. Constructing summary statistics for approximate Bayesian computation: semi-automatic approximate Bayesian computation. *Journal of the Royal Statistical Society*, 2012.

M. Hainy, W. G. Müller, and H. Wagner. Likelihood-free simulation-based optimal design. Technical report, Johannes Kepler University, Linz, 2013a.

M. Hainy, W. G. Müller, and H. P. Wynn. *Approximate Bayesian Computation Design (ABCD), an Introduction*. Springer International Publishing Switzerland, 2013b.

I. Krishnarajah, A. Cook, G. Marion, and G. Gibson. Novel moment closure approximations in stochastic epidemics. *Bulletin of Mathematical Biology*, 67:855–873, 2005.

S. Kullback and R. A. Leibler. On information and sufficiency. *The Annals of Mathematical Statistics*, 22:79–86, 1951.

D. Ludwig. Qualitative behavior of stochastic epidemics. *Mathematical Biosciences*, 23:47–73, 1975.

P. Marjoram, J. Molitor, V. Plagnol, and S. Tavaré. Markov chain Monte Carlo without likelihoods. *PNAS*, 100(26):15324–15328, 2003.

T. McKinley, A. R. Cook, and R. Deardon. Inference in epidemic models without likelihoods. *The International Journal of Biostatistics*, 5, 2009.

P. Müller. Simulation based optimal design. In J.M. Bernardo, editor, *Bayesian Statistics*, pages 459–474. Oxford University Press, 1999.

D. E. Pagendam and P. K. Pollett. Optimal design of experimental epidemics. *Journal of Statistical Planning and Inference*, 143(3):563–572, 2013.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2014. URL `http://www.R-project.org/`.

C. Ryan, C. Drovandi, and A. Pettitt. Bayesian Experimental Design for Models with Intractable Likelihoods using Indirect Inference. 2014. URL `http://eprints.qut.edu.au/73409/`.

D. Telen, F. Logist, E. Van Derlinden, I. Tack, and J. Van Impe. Optimal experiment design for dynamic bioprocesses: A multi-objective approach. *Chemical Engineering Science*, 78:82–97, 2012.