Bayesian Analysis (2016) **11**, Number 3, pp. 857–883

Optimal Bayesian Experimental Design for Models with Intractable Likelihoods Using Indirect Inference Applied to Biological Process Models

Caitríona M. Ryan*, Christopher C. Drovandi†,‡, and Anthony N. Pettitt§,¶

Abstract. This paper addresses the problem of determining optimal designs for biological process models with intractable likelihoods, with the goal of parameter inference. The Bayesian approach is to choose a design that maximises the mean of a utility, and the utility is a function of the posterior distribution. Therefore, its estimation requires likelihood evaluations. However, many problems in experimental design involve models with intractable likelihoods, that is, likelihoods that are neither analytic nor can be computed in a reasonable amount of time. We propose a novel solution using indirect inference (II), a well established method in the literature, and the Markov chain Monte Carlo (MCMC) algorithm of Müller et al. (2004). Indirect inference employs an auxiliary model with a tractable likelihood in conjunction with the generative model, the assumed true model of interest, which has an intractable likelihood. Our approach is to estimate a map between the parameters of the generative and auxiliary models, using simulations from the generative model. An II posterior distribution is formed to expedite utility estimation. We also present a modification to the utility that allows the Müller algorithm to sample from a substantially sharpened utility surface, with little computational effort. Unlike competing methods, the II approach can handle complex design problems for models with intractable likelihoods on a continuous design space, with possible extension to many observations. The methodology is demonstrated using two stochastic models; a simple tractable death process used to validate the approach, and a motivating stochastic model for the population evolution of macroparasites.

Keywords: approximate Bayesian computation, auxiliary model, Bayesian experimental design, indirect inference, Markov chain Monte Carlo, Markov processes.

1 Introduction

Experimental design is fundamental to a wide range of research activities. Optimal experimental design allows statistical inference with the least experimental cost. A utility

*Mathematical Sciences, Queensland University of Technology, Brisbane, 4001, Australia
†Mathematical Sciences, Queensland University of Technology, Brisbane, 4001, Australia, c.drovandi@qut.edu.au
‡ARC Centre of Excellence for Mathematical and Statistical Frontiers (ACEMS)
§Mathematical Sciences, Queensland University of Technology, Brisbane, 4001, Australia
¶ARC Centre of Excellence for Mathematical and Statistical Frontiers (ACEMS)

 DOI: 10.1214/15-BA977

function is specified reflecting the purpose of the experiment, for example, parameter inference, prediction, or model choice. The choice of optimal design is a decision problem to select a design by maximising the expected utility. Bayesian experimental design (Raiffa and Schlaifer, 1961; Lindley, 1972; Chaloner and Larntz, 1989; Chaloner and Verdinelli, 1995; Bernardo and Smith, 2000; Clyde, 2001) is a flexible framework incorporating prior knowledge. This paper develops Bayesian experimental design methodology for stochastic models with intractable likelihoods. In our examples, the experimental design problem is what values of controllable covariates to use in order to optimally estimate the model parameters.

Much of the experimental design literature requires a tractable likelihood as the utility is a functional of the posterior. In many problems, the likelihood is intractable. An approximation to the likelihood for stochastic process models has been proposed to address this problem (Cook et al., 2008), but this approach is inadequate in the case of our motivating stochastic model for the population evolution of macroparasites (Riley et al., 2003). An approximate Bayesian computation (ABC) approach has been applied, specifying the utility as a functional of the simulation based ABC posterior (Drovandi and Pettitt, 2013). However, this approach is restricted to a discrete design space, to designing for a small number of observations and to available computer memory (see the discussion in Section 5 for more details). Our approach using indirect inference (II) is free of these restrictions.

II is a simulation-based method where a tractable auxiliary model is used in conjunction with the intractable generative model to make inferences about the generative model, the assumed true model of interest. The method assumes that a useful auxiliary model can be found easily. It may be known from the literature or found using statistical methods of data analysis. This and other assumptions are clarified in Section 3. II is strongly established as an inferential approach in the frequentist literature (Gourieroux et al., 1993; Smith, 1993; Gallant and Tauchen, 1996; Heggland and Frigessi, 2004) and more recently in the Bayesian literature (see Drovandi et al. (2015) for a review). The use of II in this paper is similar to the Bayesian approach of Gallant and McCulloch (2009) and Reeves and Pettitt (2005), where the auxiliary model likelihood is used as a replacement for the intractable generative model likelihood. In our problem, this replacement expedites utility estimation for models with intractable likelihoods.

Our approach is as follows. Having found a suitable auxiliary model, the relationship between the auxiliary model parameters and the generative model parameters is estimated using simulations from the generative model and maximum likelihood estimation. This results in precomputed auxiliary parameter values and corresponding generative model parameter draws from the prior. Using importance sampling, an II posterior is formed. The utility is a function of this auxiliary posterior. We note that this approximation of the utility can be used within any search algorithm to determine an optimal design. For illustrative purposes, we use the Metropolis–Hastings algorithm of Müller et al. (2004) (Müller algorithm hereafter).

In this paper, we demonstrate that the use of Bayesian II and the Müller algorithm can provide useful optimal designs for models with intractable likelihoods. It is a novel approach and, unlike existing techniques, can be carried out on a continuous

design space. Moreover, it can be extended to complex design problems as the storage requirements of the precomputation scale well with an increase in the dimension of the experimental design and the number of observations to design for. We demonstrate its effectiveness in choosing optimal Bayesian designs for precise parameter estimation in biological process models. The motivating example for this work is designing for a stochastic model for the population evolution of macroparasites (Michael et al., 1998; Riley et al., 2003). Macroparasites are transmitted by mosquitoes, causing lymphatic filariasis, a painful disfiguring disease in an estimated 120 million people worldwide (Ottesen, 2006). This model results in an intractable likelihood and therefore an intractable utility function. A simple tractable death model (Renshaw, 1991) validates the approach and algorithm settings.

The paper describes the Bayesian approach to optimal experimental design in Section 2. Section 3 explains our novel II approach to Bayesian optimal design. Section 4 illustrates the methodology using the simple tractable death process and the motivating stochastic model, which has an intractable likelihood. The paper concludes in Section 5 with a discussion of the novelty and advantages of our approach and related open questions in Bayesian design, involving complex designs and a substantial number of observations, where our approach should be useful.

2 Bayesian experimental design

2.1 Introduction

To address experimental design in the Bayesian paradigm, a utility function $u(d, \mathbf{y}, \boldsymbol{\theta})$ is specified to reflect the goals of the experiment, where $\mathbf{y}$ are the data that may be observed when experimental design d is used and where $\boldsymbol{\theta}$ is the vector of model parameters. The prior distribution $p(\boldsymbol{\theta})$, can be elicited from experts or based on previous experiments. The optimal design d^* in the design space D corresponds to the maximum expected utility,

$$d^* = \underset{d \in D}{\operatorname{argmax}}\ u(d), \tag{1}$$

where $u(d)$ is the expected utility function over the data $\mathbf{y}$ and the model parameters $\boldsymbol{\theta}$,

$$u(d) = \mathbf{E}_{\boldsymbol{\theta},\mathbf{y}}[u(d, \mathbf{y}, \boldsymbol{\theta})] = \int_{\mathbf{y}} \int_{\boldsymbol{\theta}} u(d, \mathbf{y}, \boldsymbol{\theta}) p(\mathbf{y}|\boldsymbol{\theta}, d) p(\boldsymbol{\theta}) d\boldsymbol{\theta} d\mathbf{y}, \tag{2}$$

where $p(\mathbf{y}|\boldsymbol{\theta}, d)$ is the likelihood of the generative model, the assumed true model of interest. Integration across all possible values of the data $\mathbf{y}$ that are yet to be observed and model parameters $\boldsymbol{\theta}$ can be carried out using Monte Carlo simulation. However, in Bayesian experimental design, the utility $u(d, \mathbf{y})$ is a functional of the posterior distribution (thus $\boldsymbol{\theta}$ is integrated out of the general utility $u(d, \mathbf{y}, \boldsymbol{\theta})$) and is therefore intractable for models with intractable likelihoods. Section 3 demonstrates how the utility is approximated using II, avoiding generative model likelihood evaluations.

A pragmatic approach taken by Müller et al. (2004) is to sample from an augmented joint probability distribution

$$h(d, \boldsymbol{\theta}, \mathbf{y}) \propto u(d, \mathbf{y}, \boldsymbol{\theta}) p(\mathbf{y}|\boldsymbol{\theta}, d) p(\boldsymbol{\theta}). \tag{3}$$

Note that the normalising constant required for this distribution for fixed d is the expected utility $u(d)$ of (2). Thus the margin over $\mathbf{y}$ and $\boldsymbol{\theta}$ of the target distribution $h(d, \boldsymbol{\theta}, \mathbf{y})$ is proportional to $u(d)$ and shares the same mode, that is, the optimal design. Müller et al. (2004) borrow ideas from simulated annealing (Van Laarhoven and Aarts, 1987), tempering the distribution by a single temperature J. This exaggerates the peaks of the distribution and enables an easier search for the mode. The tempered target distribution is

$$h_J(d, \boldsymbol{\theta}_{1:J}, \mathbf{y}_{1:J}) \propto \prod_{j=1}^{J} u(d, \mathbf{y}_j, \boldsymbol{\theta}_j) p(\mathbf{y}_j | d, \boldsymbol{\theta}_j) p(\boldsymbol{\theta}_j), \tag{4}$$

where $\boldsymbol{\theta}_{1:J}$ and $\mathbf{y}_{1:J}$ are J independent draws from the prior predictive distribution conditional on d. Increasing the value of J tightens the distribution further at its mode, but incurs a computational cost. An alternative approach is provided below in Section 2.2 that exaggerates the peaks of the expected utility surface with little computational cost. To use the Müller algorithm, the utility function must be bounded below by 0.

In order to sample from the augmented target in (4), Müller et al. (2004) employ a Metropolis–Hastings algorithm that requires the specification of a joint proposal distribution, $q(d^*, \boldsymbol{\theta}^*_{1:J}, \mathbf{y}^*_{1:J} | d, \boldsymbol{\theta}_{1:J}, \mathbf{y}_{1:J})$, that defines how new values in the Markov chain are proposed, where $(d^*, \boldsymbol{\theta}^*_{1:J}, \mathbf{y}^*_{1:J})$ denotes the jointly proposed values. Here we use $q(d^*, \boldsymbol{\theta}^*_{1:J}, \mathbf{y}^*_{1:J} | d, \boldsymbol{\theta}_{1:J}, \mathbf{y}_{1:J}) = \prod_{j=1}^{J} p(\mathbf{y}^*_j | d^*, \boldsymbol{\theta}^*_j) p(\boldsymbol{\theta}^*_j) q(d^* | d)$, where $q(d^* | d)$ is user-specified (we provide more details on specific examples later) while $(\boldsymbol{\theta}^*_{1:J}, \mathbf{y}^*_{1:J})$ are generated from the prior predictive distribution conditional on d^*. Note that the choice to draw $\mathbf{y}^*_{1:J}$ from the likelihood is necessary since it ensures that all intractable likelihood components cancel in the Metropolis–Hastings ratio. With this choice of proposal, the Metropolis–Hastings acceptance probability, α, becomes

$$\alpha = \min \left(1, \frac{q(d|d^*) \prod_{j=1}^{J} u(d^*, \mathbf{y}^*_j, \boldsymbol{\theta}^*_j)}{q(d^*|d) \prod_{j=1}^{J} u(d, \mathbf{y}_j, \boldsymbol{\theta}_j)} \right).$$

Driven by the goal of parameter inference, we use the Bayesian D-posterior precision utility (Drovandi et al., 2013), which is the inverse of the determinant of the posterior covariance matrix of the model parameters $\boldsymbol{\theta}$,

$$u(d, \mathbf{y}) = 1/\det(\mathrm{Var}(\boldsymbol{\theta} | \mathbf{y}, d)), \tag{5}$$

which does not depend on $\boldsymbol{\theta}$. The utility can also be defined for a subset of $\boldsymbol{\theta}$.

2.2 Sharpening the utility surface

In order to sharpen the utility surface, a large value of J, greater than 100, can be used in the Müller algorithm. This is highly computationally intensive. Instead, we take advantage of an assumption of our method (see Assumption (2), Section 3) that an informative prior is available. We consider a gain in utility from the data over and above the utility found from the informative prior,

$$\tilde{u}(d, \mathbf{y}) = \max\{(u(d, \mathbf{y}) - s u_p), 0\}, \tag{6}$$

where u_p is the utility calculated from the prior distribution,

$$u_p = 1/\det(\mathrm{Var}(\boldsymbol{\theta})), \tag{7}$$

and where $s \leq 1$ is a user-defined scaling factor on u_p.

Using this modified utility, we found that the utility surface was substantially sharpened in our examples using $J = 10$ in the Müller algorithm. The idea is demonstrated in Figure 1(a) for one observation of the macroparasite model, for which details are given in Section 4.2. The estimated expected utility is plotted for various observation times. In the left hand plot, $\hat{u}(d)$ is plotted and the dashed line is the prior utility, u_p. In the right hand plot, $\hat{\tilde{u}}(d)$ is plotted. This curve has much greater curvature near the mode than the left hand plot. The curvature of the normalised $\tilde{u}(d)$ is increased by a factor of about 5 compared with the normalised $u(d)$. We surmise that this is equivalent to increasing J by a factor of 5.

In our examples, $s \approx (0.7, 0.8)$ in (6) ensured $\tilde{u}(d, \mathbf{y}) - s u_p$ was usually greater than 0. This approach is effective as the Müller algorithm treats the target $\tilde{u}(d, \mathbf{y}) p(\mathbf{y}|\boldsymbol{\theta}, d) p(\boldsymbol{\theta})$ as a normalised density with respect to $(d, \mathbf{y}, \boldsymbol{\theta})$. The normalised target marginal density $\tilde{u}(d)/\tilde{u}$, where $\tilde{u}$ is the normalising constant for $\tilde{u}(d)$ with respect to d, has greater curvature near the mode than the normalised density $u(d)/u$, where u is the normalising constant for $u(d)$ with respect to d. So draws of d using $\tilde{u}(d, \mathbf{y})$ are more concentrated about the mode of $u(d)$ than draws using $u(d, \mathbf{y})$. As currently defined, if the prior is uninformative, then u_p could be 0. However, in the general case, any approximate lower bound for $u(d, \mathbf{y})$ will sharpen the utility surface. The mode of $\tilde{u}(d, \mathbf{y})$ and correspondingly the optimal design, is not changed from that of $u(d, \mathbf{y})$ by the choice of s. The maximum is introduced to ensure that when $u(d, \mathbf{y})$ is estimated, $\tilde{u}(d, \mathbf{y})$ remains non-negative and the Müller algorithm can be used.

3 Bayesian experimental design using indirect inference

II employs an auxiliary model with a tractable likelihood in conjunction with the generative model, the assumed true model of interest, to aid inference about the generative model. For a full review of the recent advances in the Bayesian II literature see Drovandi et al. (2015). Our II methodology is similar to Reeves and Pettitt (2005) and Gallant and McCulloch (2009), that is, a simulation based method, where an auxiliary model likelihood is used as a replacement for the intractable generative model likelihood. This is useful when a generative model for the observed data is derived from underlying scientific theory and when the following assumptions hold (see Gallant and McCulloch (2009)): (i) the likelihood is not available; (ii) an informative prior is available; (iii) the model can be simulated from; and (iv) an adequate auxiliary statistical model for the data is available. The auxiliary model may be known from the literature or chosen from a generally flexible and richly parameterised set of models. Statistical methods of data analysis are used, for example, building a well fitted regression model. Assumption (iv) can be verified by assessing goodness-of-fit for observed and simulated data. Assumption (iv) is important and it may not be possible or straightforward to find a suitable auxiliary model.

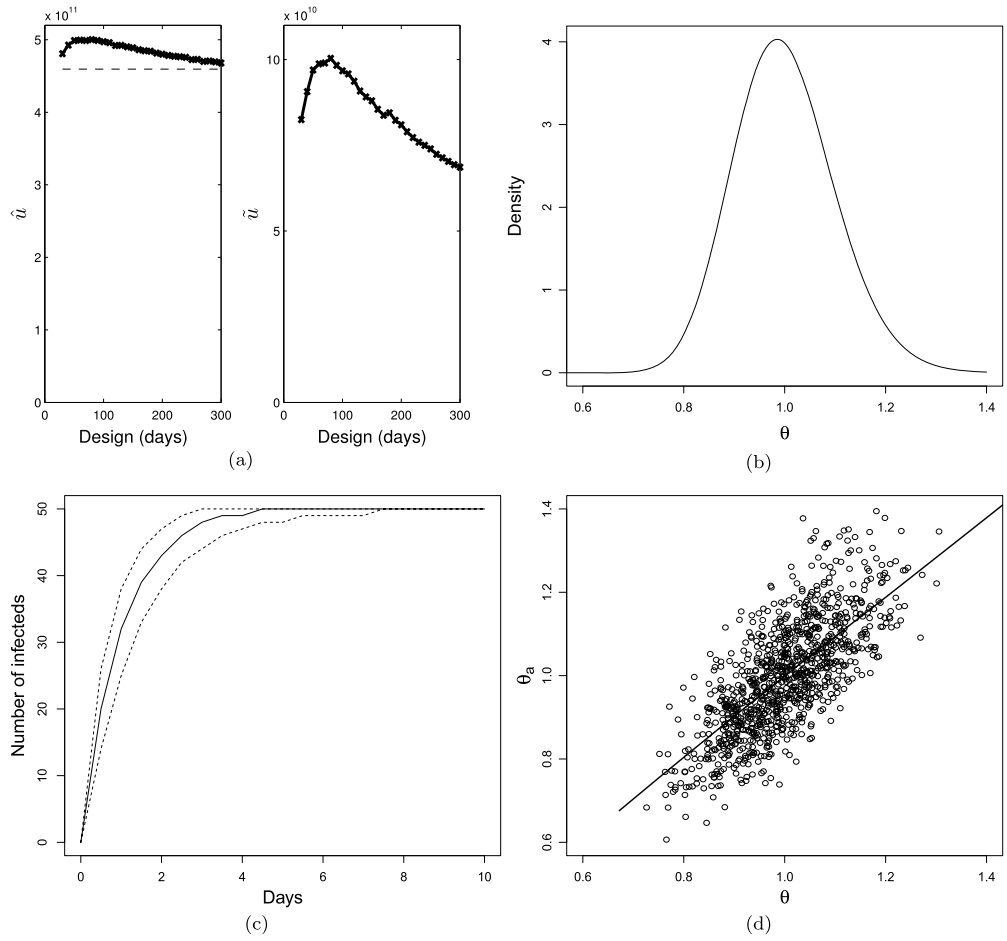


Figure 1: Plot (a) displays the mean utility of (12) for 1 observation of the macroparasite population evolution model, calculated using $1,000$ simulations over a fixed grid of designs. In the left hand plot, the expected utility estimate is $\hat{u}(d)$ and the dashed line is the prior utility u_p of (7). The right hand plot is of the modified expected utility estimate $\tilde{\hat{u}}(d)$. Plot (b) displays the prior distribution for θ, the generative model parameter of the death model. Plot (c) displays the prior predictive median (solid line) and 95% prior prediction intervals (dashed lines) of $100,000$ simulated counts of infected individuals from the death model. Plot (d) displays the marginal relationship between the auxiliary parameter $\hat{\theta}_a$ and the generative model parameter θ for the death model, where the curve is a *lowess* smooth.

Given assumption (iv), we can construct g, a map or binding function $\boldsymbol{\theta} \to \boldsymbol{\phi}$, from the parameters of the generative model to those of the auxiliary model such that the generative model likelihood is approximated, up to proportionality, by the auxiliary

model likelihood. To facilitate this, the number of parameters of the auxiliary model should be at least that of the generative model.

The fundamental assumption of the Bayesian II approach that we use is that there exists an auxiliary model with a fully tractable likelihood $p_a(\mathbf{y}|\boldsymbol{\phi}, d)$, so that approximately

$$p(\mathbf{y}|\boldsymbol{\theta}, d) \propto p_a(\mathbf{y}|\boldsymbol{\phi} = g(\boldsymbol{\theta}), d), \tag{8}$$

where the subscript a denotes the auxiliary model. If $g(\boldsymbol{\theta})$ were known, the problem becomes tractable up to the approximation. For intractable problems, $g(\boldsymbol{\theta})$ must be estimated by $\hat{g}(\boldsymbol{\theta})$ such that

$$p(\mathbf{y}|\boldsymbol{\theta}, d) \dot{\propto} p_a(\mathbf{y}|\boldsymbol{\phi} = \hat{g}(\boldsymbol{\theta}), d). \tag{9}$$

The posterior distribution $p(\boldsymbol{\theta}|\mathbf{y}, d)$ of the generative model is approximated by the II posterior,

$$p_a(\boldsymbol{\theta}|\mathbf{y}, d) \propto p_a(\mathbf{y}|\phi = \hat{g}(\boldsymbol{\theta}), d)p(\boldsymbol{\theta}). \tag{10}$$

An estimate of the auxiliary parameter $\boldsymbol{\phi}$ is found by

$$\hat{\boldsymbol{\phi}} = \hat{g}(\boldsymbol{\theta}) = \underset{\boldsymbol{\phi}\in\Phi}{\operatorname{argmax}}\ p_a(\mathbf{x}|\boldsymbol{\phi}, d_T), \quad \mathbf{x} \sim p(\mathbf{x}|\boldsymbol{\theta}, d_T), \tag{11}$$

for auxiliary parameter space Φ, $\boldsymbol{\theta} \sim p(\boldsymbol{\theta})$ and training design d_T. The training design could be of the same structure as in previous experiments, but the user is free to choose this design (more information is provided in the discussion in Section 5). Drovandi et al. (2015) demonstrate that if the auxiliary model likelihood acts as a good approximation to the generative model likelihood for regions of the parameter space of non-negligible posterior density, then a useful posterior approximation can be obtained. They demonstrate this for a number of examples including the macroparasite population evolution model that motivates this work. Such an approximation may be appropriate for design purposes since an important process in Bayesian experimental design is ranking the utilities of the designs and accurate posterior estimation is less important.

Algorithm 1 is employed to establish a noisy estimate $\hat{g}(\boldsymbol{\theta})$ of $g(\boldsymbol{\theta})$. For $\boldsymbol{\theta}^i$, the ith value of $\boldsymbol{\theta}$ generated from the prior, a dataset $\mathbf{x}$ is simulated from the generative model based on the training design d_T. The auxiliary parameters are estimated using (11). Repeating this process for a collection of n parameter values drawn from the prior produces the noisy mapping $\{\boldsymbol{\theta}^i, \hat{\boldsymbol{\phi}}^i\}_{i=1}^n$. The idea of precomputing a mapping function in II has been considered by Moores et al. (2015).

In an attempt to reduce the variability of $\hat{g}(\boldsymbol{\theta})$, one can simulate m independent replicates of the data, denoted $\mathbf{x}_{1:m} = (\mathbf{x}_1, \ldots, \mathbf{x}_m)$ to estimate $\hat{\boldsymbol{\phi}}$. This is implemented as lines 3 and 4 of Algorithm 1. Under the assumption that the auxiliary estimator $\hat{g}(\boldsymbol{\theta})$ is consistent, the true mapping $g(\boldsymbol{\theta})$, for a particular value of $\boldsymbol{\theta}$, can be recovered as $m \to \infty$. Cox (1961) gives some theory concerning maximum likelihood estimation under the wrong model, which is relevant to II and estimation of the map. In our notation, it is assumed that $\hat{g}(\boldsymbol{\theta})$ based on m replications converges in probability to $g(\boldsymbol{\theta})$ as m tends to ∞. For example, (25) of Cox (1961) gives the definition of $g(\boldsymbol{\theta})$. In

Algorithm 1: Estimation of $g(\boldsymbol{\theta})$.

INPUT: Training design d_T, number of prior simulations n, number of replicates m;

1 **for** $i = 1, \ldots, n$ **do**

2 Draw $\boldsymbol{\theta}^i \sim p(\boldsymbol{\theta})$;

3 Simulate $\mathbf{x}_j \overset{iid}{\sim} p(\mathbf{y}|\boldsymbol{\theta}^i, d_T), \quad j = 1, \ldots, m$;

4 Find $\hat{\boldsymbol{\phi}}^i = \underset{\boldsymbol{\phi} \in \Phi}{\operatorname{argmax}} \prod_{j=1}^{m} p_a(\mathbf{x}_j|\boldsymbol{\phi}, d_T)$;

5 **end**

OUTPUT: Parameter values $\{\boldsymbol{\theta}^i, \hat{\boldsymbol{\phi}}^i\}_{i=1}^n$;

Section 9, Cox (1961) considers, in our terms, the generative model to be exponential with one parameter while the auxiliary model is log normal with two parameters and gives results about the asymptotic distribution of $\hat{g}(\boldsymbol{\theta})$.

The choice of m and n in Algorithm 1 is discussed in Section 4. Increasing n allows the mapping to be approximated for more points across the prior space. For the purposes of auxiliary parameter estimation we recommend maximum likelihood estimation in line 4, Algorithm 1, since it generally leads to more statistically efficient estimators compared with other techniques. We note that other methods could be used such as the method of moments or estimating equations (Heggland and Frigessi, 2004). An important aspect of Algorithm 1 is that storage of the simulated data $\mathbf{x}_{1:m}$ is not required. It is only necessary to store the values $\{\boldsymbol{\theta}^i, \hat{\boldsymbol{\phi}}^i\}_{i=1}^n$, whose size does not necessarily grow with an increase in the number of experimental observations and the number of design variables. In the macroparsite example in Section 4.2, we demonstrate how an extra design variable can be accommodated.

Having established a noisy mapping $\hat{g}(\boldsymbol{\theta})$ using Algorithm 1, the resulting values $\{\boldsymbol{\theta}^i, \hat{\boldsymbol{\phi}}^i\}_{i=1}^n$ are used to estimate the utility function, which is required in the Müller algorithm to sample from the tempered target distribution in (4). The utility function is estimated by forming the II posterior, $p_a(\boldsymbol{\theta}|\mathbf{y}, d)$, given in (10). The Bayesian D-posterior precision utility in (5) is respecified as a function of the II posterior and is defined by

$$u(d, \mathbf{y}) = 1/\det(\widehat{\operatorname{Var}}_a(\boldsymbol{\theta}|\mathbf{y}, d)), \tag{12}$$

where $\operatorname{Var}_a(\boldsymbol{\theta}|\mathbf{y}, d)$ is the covariance matrix of $\boldsymbol{\theta}$ under the auxiliary posterior $p_a(\boldsymbol{\theta}|\mathbf{y}, d)$. The process for calculating this utility is shown in Algorithm 2. The II posterior for $\boldsymbol{\theta}$ is formed using importance sampling, where the prior $p(\boldsymbol{\theta})$ is the importance distribution and the ith importance weight, W^i, is proportional to $p_a(\mathbf{y}|\hat{\boldsymbol{\phi}}^i, d)$ for $i = 1, \ldots, n$. Then $\{W^i, \boldsymbol{\theta}^i\}_{i=1}^n$ forms an importance sampling approximation of the II posterior. This weighted sample is then used to estimate the utility $u(d, \mathbf{y})$. The Monte Carlo error of the estimated utility is associated with the effective sample size (ESS) of the importance sampling approximation, which can be estimated by $1/\sum_{i=1}^n (W^i)^2$. The

value of n needs to be set large enough so that the ESS is reasonably large (e.g. above 100) throughout the optimal design search algorithm. We provide more discussion on the importance sampling approximation in Section 5. We refer to the general approach of using the II posterior to estimate the utility as II design.

Algorithm 2: Calculation of the utility $u(d, \mathbf{y})$ using an importance sampling approximation of the II posterior distribution.

INPUT: Simulated data $\mathbf{y}$, precomputed values $\{\boldsymbol{\theta}^i, \hat{\boldsymbol{\phi}}^i\}_{i=1}^n$, design d;

1 Calculate importance weights $W^i \propto p_a(\mathbf{y}|\hat{\boldsymbol{\phi}}^i, d)$ for $i = 1, \ldots, n$, $\sum_{i=1}^n W^i = 1$;

2 Estimate $\mathrm{Var}_a(\boldsymbol{\theta}|\mathbf{y}, d)$ using the importance sampling approximation $\{W^i, \boldsymbol{\theta}^i\}_{i=1}^n$ of the II posterior distribution;

3 Set $u(d, \mathbf{y}) = 1/\det(\widehat{\mathrm{Var}}_a(\boldsymbol{\theta}|\mathbf{y}, d))$;

OUTPUT: Utility $u(d, \mathbf{y})$;

The utility calculation in line 3, Algorithm 2, is deterministic given the values $\{\boldsymbol{\theta}^i, \hat{\boldsymbol{\phi}}^i\}_{i=1}^n$ that are precomputed in Algorithm 1. The same set of values $\{\boldsymbol{\theta}^i, \hat{\boldsymbol{\phi}}^i\}_{i=1}^n$ is used for different values of d. This ensures fully reversible Metropolis–Hastings moves in the Müller algorithm, satisfying detailed balance.

The II design approach is illustrated in Section 4 using the simple tractable death process and the motivating macroparasite evolution model, which has an intractable likelihood.

4 Results

II design is demonstrated using two examples. The first is the death model (Renshaw, 1991), which is used as an illustrative problem since the likelihood for the model is easy to compute. For this model, comparisons can be made between the Müller algorithm using II, using the exact likelihood and using the likelihood approximation approach (Cook et al., 2008). We seek times when to observe the stochastic death process in order to optimally estimate the model parameters.

The second example motivates the work of this paper and concerns the population evolution of macroparasites in host cats (Michael et al., 1998; Riley et al., 2003). We seek the set of optimal times to sacrifice each cat and the optimal initial injection of juveniles to use for each new cat, whose observation is assumed to be drawn from a stochastic process. This stochastic process model is of scientific interest and satisfies the four assumptions of the II methodology given in Section 3: (i) the likelihood is computationally intractable; (ii) an informative prior is available from a previous experiment based on 212 observations; (iii) it is straightforward to simulate from this model using the algorithm of Gillespie (1977); and (iv) a useful auxiliary model is known from the literature (Drovandi et al., 2015).

Our approach is to use Nelder–Mead optimisation (Nelder and Mead, 1965) in line 4 of Algorithm 1, which is robust and works well for our examples. The target distribution

was tempered using $J = 10$, which sufficiently sharpened the utility surface for our examples using the modified utility of (6) with $s \approx (0.7, 0.8)$. Increasing J exaggerates the peaks of the marginal densities of the designs, but does not change their locations. Using a value of s above 0 allows us not to need a large value of J, which saves substantially on computation. For example, in the case of designing for one observation, the use of $s = 0$, requires $J = 40$ to achieve similar results in the same number of iterations of the Müller algorithm as the use of $s = 0.7$ with $J = 10$. As discussed in Section 2.2, these results can be explained as the Müller algorithm treats the target $\tilde{u}(d, \mathbf{y})p(\mathbf{y}|\boldsymbol{\theta}, d)p(\boldsymbol{\theta})$ as a normalised density in $(d, \mathbf{y}, \boldsymbol{\theta})$ and the curvature about its mode is larger than the mode of the target using $u(d, \mathbf{y})$ rather than $\tilde{u}(d, \mathbf{y})$.

4.1 Death model example

Generative model

The simple death process is used to illustrate the methodology and to validate the algorithm settings. At time t, with $S(t) = i$ susceptibles, the probability that an infection occurs in the next infinitesimal time period Δt is given by

$$p\left(S(t + \Delta t) = i - 1 | S(t) = i\right) = \theta i \Delta t + o(\Delta t),$$

where θ is the per-capita infection rate, the generative model parameter of interest. The observable data are susceptible, time pairs $(t_1, S(t_1)), \ldots (t_T, S(t_T))$, where $S(t_j)$ follows a conditional binomial distribution,

$$S(t_j)|S(t_{j-1}) \sim \text{binomial}\left(S(t_{j-1}), \mathbf{e}^{-\theta(t_j - t_{j-1})}\right), \quad j = 1, \ldots, T. \tag{13}$$

The process is initialised with one infected and the number of susceptibles at time t_0 is $S(t_0)$. The prior distribution for θ follows Cook et al. (2008), where $\log\theta \sim \text{normal}(-0.005, 0.01)$ and is displayed in Figure 1(b). The initial number of susceptibles is 50. Figure 1(c) displays the prior predictive median (solid line) and 95% prior prediction intervals (dashed lines) of 100,000 data simulations from the death model.

Auxiliary model

To demonstrate the methodology, a normal distribution is used as an auxiliary model,

$$\begin{aligned} &S(t_j)|S(t_{j-1}) \\ &\quad \sim \text{normal}\left(S(t_{j-1})\mathbf{e}^{-\theta_a(t_j^\gamma - t_{j-1}^\gamma)}, \sigma^2 S(t_{j-1})\mathbf{e}^{-\theta_a(t_j^\gamma - t_{j-1}^\gamma)}\left(1 - \mathbf{e}^{-\theta_a(t_j^\gamma - t_{j-1}^\gamma)}\right)\right). \end{aligned} \tag{14}$$

This allows a more flexible variance and time scale than the standard normal approximation to the binomial distribution, which helps to illustrate our methodology including the possibility of redundant parameters. The auxiliary model is over parameterised but our approach accommodates this. Equation (14) is a poor approximation for $S(t_{j-1})$ close to 0, giving support to negative values of $S(t_j)$. Nevertheless, it is useful for illustrative purposes and performs well for the design problem. For this example, the auxiliary parameters are $\boldsymbol{\phi} = (\theta_a, \gamma, \sigma^2)$.

Optimal observation times

For the II approach to Bayesian experimental design, Algorithm 1 is used to estimate the map $g(\theta)$ resulting in precomputed values $\{\theta^i, \hat{\boldsymbol{\phi}}^i\}_{i=1}^n$ that can be used to estimate the utility. For this example, the prior predictive distribution of Figure 1(b) shows that the process is most likely complete by 10 days so we take the training design d_T as 20 equispaced single observations across 10 days. The choice of d_T is itself a design problem, which requires further investigation. The resulting values of $\hat{\sigma}^2$ and $\hat{\gamma}$ were centred close to 1, which was expected since the normal approximation to the binomial distribution for large $S(t_j)$ is (14), where $\sigma^2 = \gamma = 1$. This demonstrates that the auxiliary model can include redundant parameters. A strong relationship between $\hat{\theta}_a$ and θ can be seen in Figure 1(d), where the curve is a *lowess* smooth.

The Müller algorithm was carried out for $100,000$ iterations of each of one to four observation times. Designs d of interest are continuous observation times between $d_{min} = 0$ days and $d_{max} = 10$ days. The algorithm cycles through each design point in turn, given the current value of all other design points. Observation times are ordered such that $d_1 < d_2 < \cdots < d_k$, for $k = 1, \ldots, 4$. At iteration t, the proposed value of d_i^t is generated from a truncated normal random walk with variance 1, truncated at d_{i-1}^t below and d_{i+1}^{t-1} above for $i = 1, \ldots, k$ (where the notation assumes that $d_0^t = d_{min}$ and $d_{k+1}^{t-1} = d_{max}$). Figure 2 displays the resulting marginal densities of $d_1, \ldots, d_k$ for $k = 1, \ldots, 4$ observations (continuous lines). The truncated normal random walk mixes well for our examples with $k \leq 4$. Other proposals could be used if mixing was problematic, such as an independence sampler with uniform order statistics or block updates.

The likelihood of the death model is tractable, and the approach was also carried out using the exact likelihood in place of the auxiliary likelihood in the utility calculation of Algorithm 2. The resulting marginal densities for d are plotted in Figure 2 (dashed lines). The marginal densities inferred using II are similar, demonstrating the effectiveness of the proposed approach with the settings of $m = 3$, $n = 10,000$ in Algorithm 1 and $J = 10$ in the Müller algorithm.

We follow Drovandi and Pettitt (2013) and use a density estimation approach to determine the optimal design, that is, the mode estimated from the MCMC samples. For more than one observation time, it is necessary to find the multivariate mode since the designs are time ordered. For example, consider the case of two ordered observations at $0 < d_1 < d_2 < d_{max}$ drawn from the uniform density on the simplex, $0 < d_1 < d_2 < d_{max}$. The marginal modes of d_1, d_2 would appear at the endpoints 0 and d_{max}. Therefore, we use a Gaussian smoothing kernel to determine the multivariate mode, finding the design that has the most other designs in its vicinity. Further details of the mode finding approach can be found in Appendix A of Drovandi and Pettitt (2013).

The expected utility is calculated at the modal designs using simulations $\mathbf{y}_j$ from the generative model, where

$$u(d) \approx \sum_{j=1}^{10^6} u(d, \mathbf{y}_j), \tag{15}$$

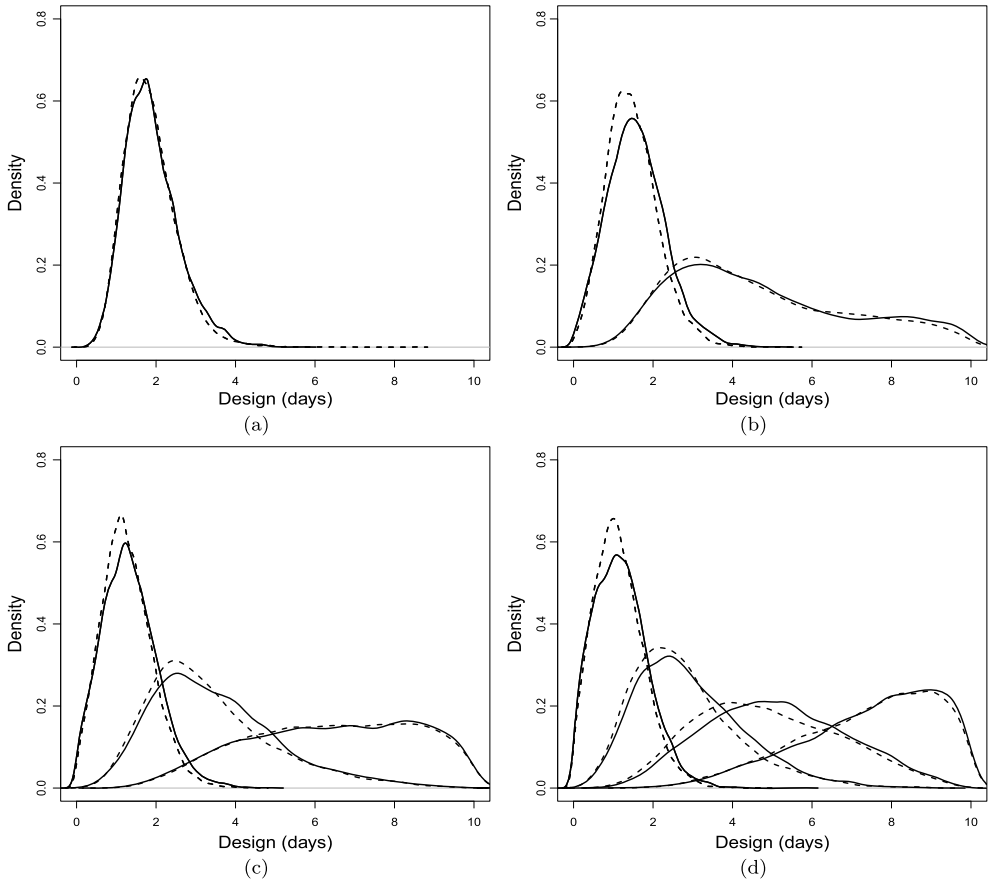


Figure 2: Marginal density estimates of the designs for 1 to 4 observations of the death model (plots (a) to (d), respectively) for 100, 000 iterations of the Müller algorithm using II. The dashed lines are based on 100, 000 iterations of the Müller algorithm using the exact binomial likelihood in place of the auxiliary likelihood in the utility calculation of Algorithm 2.

and where $u(d, \mathbf{y}_j)$ is calculated using the exact binomial likelihood in place of the auxiliary likelihood to calculate the importance weights in line 1, Algorithm 2. The resulting optimal designs are displayed in Table 1 and are similar to the optimal designs found by Cook et al. (2008) using the likelihood approximation approach and the optimal designs found using the exact likelihood. All designs for $k \geq 2$ are generally well spaced in the interval $(0.5, 5.3)$ rather than clustered about some time. This demonstrates that II design finds designs that are close to those that are found using the exact likelihood. Expected utility values in Table 1 are estimated with a standard error of less than 0.07. For $k = 3$ observations, the design found using the exact likelihood produces a lower estimate of the expected utility than the II approach and the approach of Cook

et al. (2008). This demonstrates the difficulty of estimating the mode of a multivariate MCMC estimated density.

k	II design	$u(d)$	Cook et al. (2008)	$u(d)$	Exact	$u(d)$
1	1.8	133.54	1.7	133.61	1.7	133.68
2	$(1.2, 3.3)$	142.30	$(0.9, 2.4)$	142.69	$(1.1, 2.9)$	142.70
3	$(0.8, 2.2, 4.2)$	146.06	$(0.7, 1.5, 2.9)$	146.34	$(1.0, 2.2, 4.4)$	145.93
4	$(0.6, 0.9, 2.0, 3.6)$	147.77	$(0.8, 1.7, 3.1, 5.3)$	148.01	$(0.5, 1.2, 2.6, 4.1)$	148.10

Table 1: Optimal designs for $k = 1, \ldots, 4$ observations of the death model using II design, the exact likelihood and the likelihood approximation (results from Cook et al. (2008)). Expected utilities (see (15)) are calculated by importance sampling at the optimal design inferred by each approach using the exact likelihood to compute importance weights. Expected utility values are estimated with a standard error of less than 0.07.

4.2 Macroparasite model example

The motivating example for this work is a study of the population evolution of L3 *Brugia pahangi* larvae in cats (Denham et al., 1972). The experiment involved the injection of approximately 100 or 200 larvae in 212 host cats. The cats are assumed independent and had never previously been exposed to the parasite. At various times between 24 and 1,193 days the cats were sacrificed (killed), and the number of live mature parasites in each cat was recorded. The data from the study (Denham et al., 1972) informs the prior distribution $p(\boldsymbol{\theta})$ for future experiments. A stochastic Markov model was developed (Michael et al., 1998; Riley et al., 2003) to explain the population evolution of *Brugia pahangi*. As with many real world problems, this model is of scientific interest but has an intractable likelihood and therefore an intractable utility for optimal design.

The aim of the experimental design in this example is to choose values of the controllable covariates (observation times and initial injections of larvae) for independent replications of the stochastic process, which is optimal for precise parameter estimation. That is, the set of optimal covariate values to use for each additional cat. Precise estimation of a subset of the model parameters is considered.

Generative model

The following stochastic model was developed to explain the within-host population dynamics of lymphatic filariasis (Michael et al., 1998; Riley et al., 2003). At time t, a host cat is described by the following random variables; the mature parasite count $M(t)$, the larvae count $L(t)$ and a discrete measure of the host's immunity $I(t)$. The host cats have never previously been exposed to the parasite and thus have no experience of infection, no immunity and no mature parasites. This gives initial conditions of $I(0) = 0$, $M(0) = 0$ and $L(0) = l_i$, since the initial number of larvae l_i for cat i is injected at time 0. The values of the states at time t are $M(t) = i$, $L(t) = j$ and $I(t) = k$. The larvae mature to adult parasites, die due to the natural death of the larvae, or die due to the immune response of the host. Note that the number of larvae is unobservable at

the time of sacrifice, as is the level of immunity. Only mature live parasites are counted. The mature parasites die at a rate μ_M per larva per day. Larvae mature at a rate γ per larva per day. Larvae die at a rate $\mu_L + \beta I(t)$ per day, where μ_L corresponds to the natural death rate of the larvae and β is the additional death rate due to the host's immune response. The host's immune response, $I(t)$, is assumed to increase at a rate ν per larva per day and decrease at a rate μ_I per unit of immunity. The immune response is assumed to affect larvae only and not the mature parasites. Thus for a small time interval Δt such that at most one event can occur, the transition probabilities at time $t + \Delta t$ are given by

$$\begin{aligned} p(i+1, j-1, k) &= \gamma j \Delta_t + o(\Delta_t), \\ p(i, j-1, k) &= (\mu_L + \beta k) j \Delta_t + o(\Delta_t), \\ p(i-1, j, k) &= \mu_M i \Delta_t + o(\Delta_t), \\ p(i, j, k+1) &= \nu j \Delta_t + o(\Delta_t), \\ p(i, j, k-1) &= \mu_I k \Delta_t + o(\Delta_t). \end{aligned} \tag{16}$$

The generative model parameters are $\boldsymbol{\theta} = (\nu, \mu_L)$, with $(\mu_I, \mu_M, \gamma, \beta)$ held fixed. Unfortunately, only the number of mature parasites can be counted at sacrifice time, which has been demonstrated to provide little information about the parameters β and μ_I (Drovandi and Pettitt, 2011). Thus we assume fixed values of $\beta = 1.1$ and $\mu_I = 0.31$ (Riley et al., 2003). Alternative experiments could be designed to provide information about these parameters and about μ_M and γ, which we assume known at estimates found in previous studies; $\gamma = 0.04$ (Suswillo, Denham, and McGreevy, 1982) and $\mu_M = 0.0015$ (Michael et al., 1998). The aim of this experimental design is the precise estimation of the parameters (ν, μ_L) for which the mature parasite count provides information. It would be possible to consider different fixed values for γ and μ_M to assess the robustness of the optimal design to these parameters. Further, to obtain a robust design, one could place prior distributions on these parameters, but not update them with the data. Either of these options will simply result in a different mapping function, which can be fed into the same design optimisation algorithm. Since our main aim here is to demonstrate the II design methodology, we do not pursue these options.

The prior distribution for (ν, μ_L) is displayed in Figure 3(a) and is the posterior taken from a previously published study (Drovandi et al., 2011), where $(\sqrt{\nu}, \sqrt{\mu_L})$ is approximated by a bivariate normal distribution with mean $(0.0361, 0.0854)$ and standard deviations $(0.0045, 0.0342)$ with a correlation of -0.6974. Figure 3(b) displays the prior predictive median (solid line) of $100,000$ data simulations from the model together with 95% prior prediction intervals (dashed lines).

Auxiliary model

An auxiliary beta-binomial distribution based on the observed sample of $N = 212$ data points has been demonstrated to fit the data well (Drovandi et al., 2011), where the probability function for the ith host is given by

$$p(M(t_i) = m_i | a_i, b_i, l_i) = \binom{l_i}{m_i} \frac{B(m_i + a_i, l_i - m_i + b_i)}{B(a_i, b_i)}, \tag{17}$$

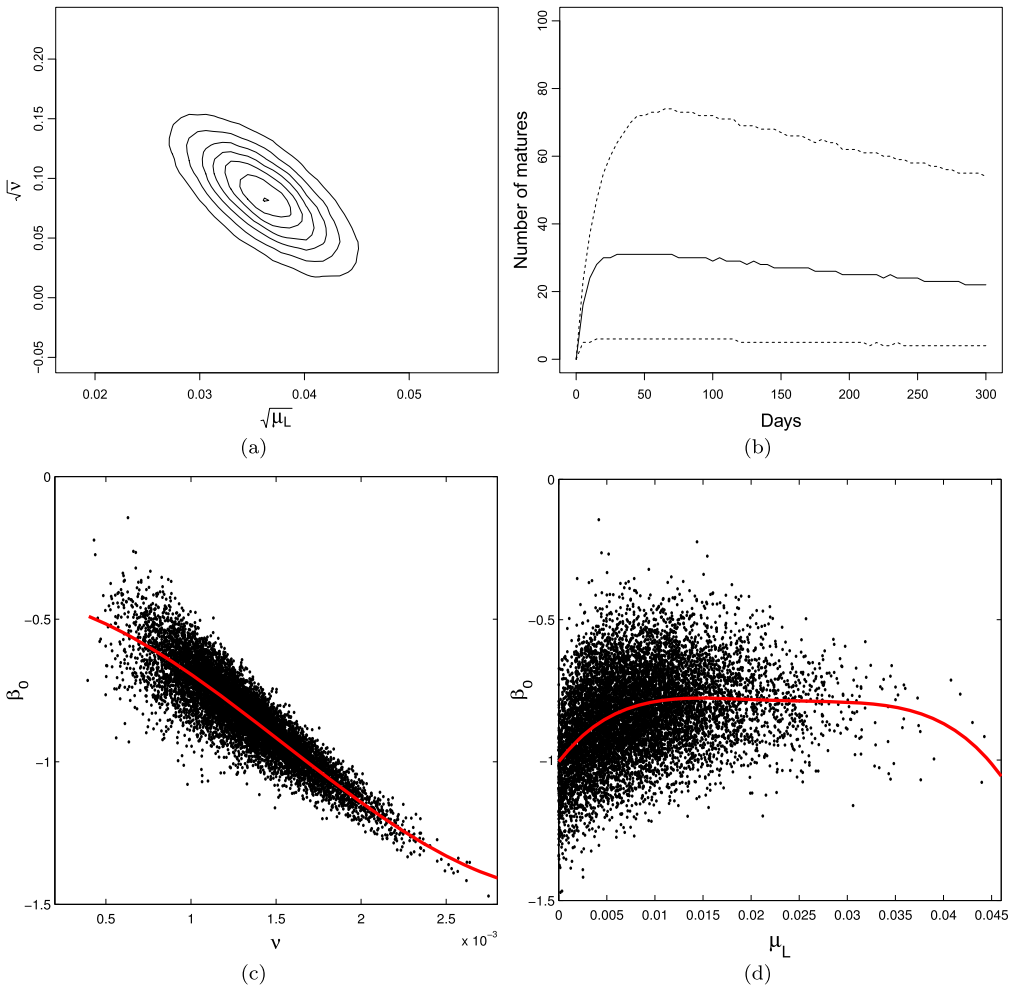


Figure 3: Plot (a) displays the bivariate normal prior distribution for $(\sqrt{\nu}, \sqrt{\mu_L})$. Plot (b) displays the prior predictive median (solid line) and 95% prior prediction intervals (dashed lines) of 100,000 simulated counts of mature parasites from the macroparasite model, where $l_i = 100$ initial larvae were injected in each host. Plots (c) and (d) display the marginal relationship between the auxiliary parameter estimate $\hat{\beta}_0$ and the generative model parameters $\boldsymbol{\theta} = (\nu, \mu_L)$, respectively, for the macroparasite example. The smooth curves in plots (c) and (d) are polynomial regression fits, which indicate a reasonably strong relationship between these generative and auxiliary parameters.

where the observed data is the initial larvae count l_i and the mature parasite count m_i at sacrifice time t_i for hosts $i = 1, \ldots, N$ and where $B(\cdot, \cdot)$ denotes the beta function. This is re-parameterised in terms of a proportion $p_i = a_i/(a_i + b_i)$ and overdispersion $\xi_i = 1/(a_i + b_i)$. Initially we attempt to determine the optimal sacrifice times when the

initial number of larvae is fixed at 100 for each host. Further, we only consider sacrifice times in the range (30, 300) days. For this purpose, we consider the following auxiliary model that is similar to the one used in Drovandi et al. (2011):

$$\begin{aligned} \text{logit}(p_i) &= f_p(t_i, l_i) = \beta_0 + \beta_1(\log(t_i) - \kappa) + \beta_2(\log(t_i) - \kappa)^2, \\ \log(\xi_i) &= \eta, \end{aligned}$$

where $\kappa = \log(165)$ is introduced to weaken the dependence between the auxiliary parameter estimates of $(\beta_0, \beta_1, \beta_2)$. Given the above re-parameterisation, the auxiliary parameters are $\boldsymbol{\phi} = (\beta_0, \beta_1, \beta_2, \eta)$.

Later, we consider optimising both the observation times (again between (30, 300) days) and the initial number of larvae (any integer between 100 and 200 inclusive). For this purpose we extend the auxiliary model:

$$\begin{aligned} \text{logit}(p_i) &= f_p(t_i, l_i) = \beta_0 + \beta_1(\log(t_i) - \kappa) + \beta_2(\log(t_i) - \kappa)^2, \\ \log(\xi_i) &= f_\xi(t_i, l_i) = \eta_0 + \eta_1 l_i, \end{aligned}$$

where the auxiliary parameters are now given by $\boldsymbol{\phi} = (\beta_0, \beta_1, \beta_2, \eta_0, \eta_1)$.

Optimal observation times

For the training design we consider 1000 randomly selected sacrifice times in (30, 300) days where the initial number of larvae is set to 100 for every host in the training design. The precomputation step, Algorithm 1, was carried out for $m = 1$ and $n = 10,000$. Since the number of observations induced by the training design is quite large, we use $m = 1$. Figures 3(c) and 3(d) display the marginal relationship between $\boldsymbol{\theta} = (\nu, \mu_L)$ and $\hat{\beta}_0$ as an example of the relationship between the generative model parameters $\boldsymbol{\theta}$ and the auxiliary parameters $\boldsymbol{\phi}$. There appears to be a reasonably strong relationship between these generative and auxiliary parameters. To assess that the beta–binomial auxiliary model provides a good description of data generated from the macroparasite population evolution model, a goodness-of-fit test of the auxiliary model for 10,000 simulations from the generative model was performed using a generalised Pearson test statistic. Details can be found in Appendix C of Drovandi et al. (2015).

The utility function for this problem was defined in (12) as the inverse of the determinant of the variance of the II posterior distributions for $\boldsymbol{\theta} = (\nu, \mu_L)$. Other choices of utility functions could be used such as the inverse of the trace of the posterior covariance matrix (Bayesian A-posterior precision) or the marginal posterior precision of μ_L or ν.

Results for 100,000 iterations of the Müller algorithm are displayed in Figure 4 for the macroparasite example. To make comparisons with the ABC approach of Drovandi and Pettitt (2013), designs are restricted to earlier than $d_{max} = 300$ days and later than $d_{min} = 30$ days, with the number of initial larvae fixed at 100 throughout the Müller algorithm. Here we attempt to update each observation time one-at-a-time and propose an observation time uniformly between adjacent observation times (and restricted to the interval mentioned earlier). Table 2 shows the optimal designs obtained from II

design and also those from ABC design in Drovandi and Pettitt (2013). As in the death model example, density estimation was used to find the optimal designs. The table also shows the estimated expected utility values when using the II utility and the ABC utility of Drovandi and Pettitt (2013). It is evident from the table that there seems to be some difference between the optimal designs obtained from both methods. However, from the estimated utility values, it appears there are a wide range of designs that are efficient. The designs obtained from the II method have generally higher utility than the ABC designs when using the II utility function, as expected. The II designs are also very competitive under the ABC utility, and perhaps slightly better for 3 and 4 design points. Since the approach of Drovandi and Pettitt (2013) does not use the adjusted utility in (6), it may have more difficulty identifying the mode.

method	design (days)	II utility $\times 10^{11}$ (std err $\times 10^9$)	ABC utility $\times 10^{11}$ (std err $\times 10^9$)
II design	77.4	5.00 (0.86)	5.95 (5.8)
ABC design	99	4.97 (0.84)	6.08 (6.2)
II design	(69.7, 97.5)	5.38 (1.2)	7.20 (6.6)
ABC design	(71, 127)	5.35 (1.2)	7.18 (6.7)
II design	(56.5, 82.8, 120.3)	5.73 (1.5)	7.39 (5.3)
ABC design	(95, 105, 231)	5.54 (1.4)	7.28 (5.3)
II design	(50.5, 75.2, 98.2, 143.8)	6.10 (1.8)	7.28 (4.3)
ABC design	(79, 121, 231, 273)	5.61 (1.5)	7.07 (4.1)

Table 2: Optimal designs for the macroparasite model for $k = 1, \ldots, 4$ observations, using the Müller algorithm with II design and using ABC (results from Drovandi and Pettitt (2013)). Both approaches use a Gaussian smoothing kernel to find the modal values. Expected utilities (see (15)), are calculated using II and ABC (Drovandi and Pettitt, 2013) at the optimal design inferred by each approach.

We ran II design also with $m = 10$. We found that the marginal distribution of the second and third auxiliary parameter estimates had higher precision, demonstrating an improved estimate of the mapping function. However, we obtained very similar marginal distributions of the design parameters as with $m = 1$. Thus it appears that $m = 1$ is suitable for this application.

Optimal observations times and initial larvae counts

We now use our II approach to perform optimal design across two dimensions; both the observation times and initial larvae counts for hosts. We consider one or two hosts. There are 1000 hosts in the training design, where the sacrifice times are sampled randomly in (30, 300) days and the initial larvae counts are randomly taken from the set of integers between 100 and 200 inclusive. The precomputation step, Algorithm 1, was carried out for $m = 1$ and $n = 10,000$.

In the Müller algorithm, we propose observation times using the same process as the previous section. Proposed initial larvae counts are drawn uniformly over the allowable values (100–200). We attempt to update each design parameter one-at-a-time. We

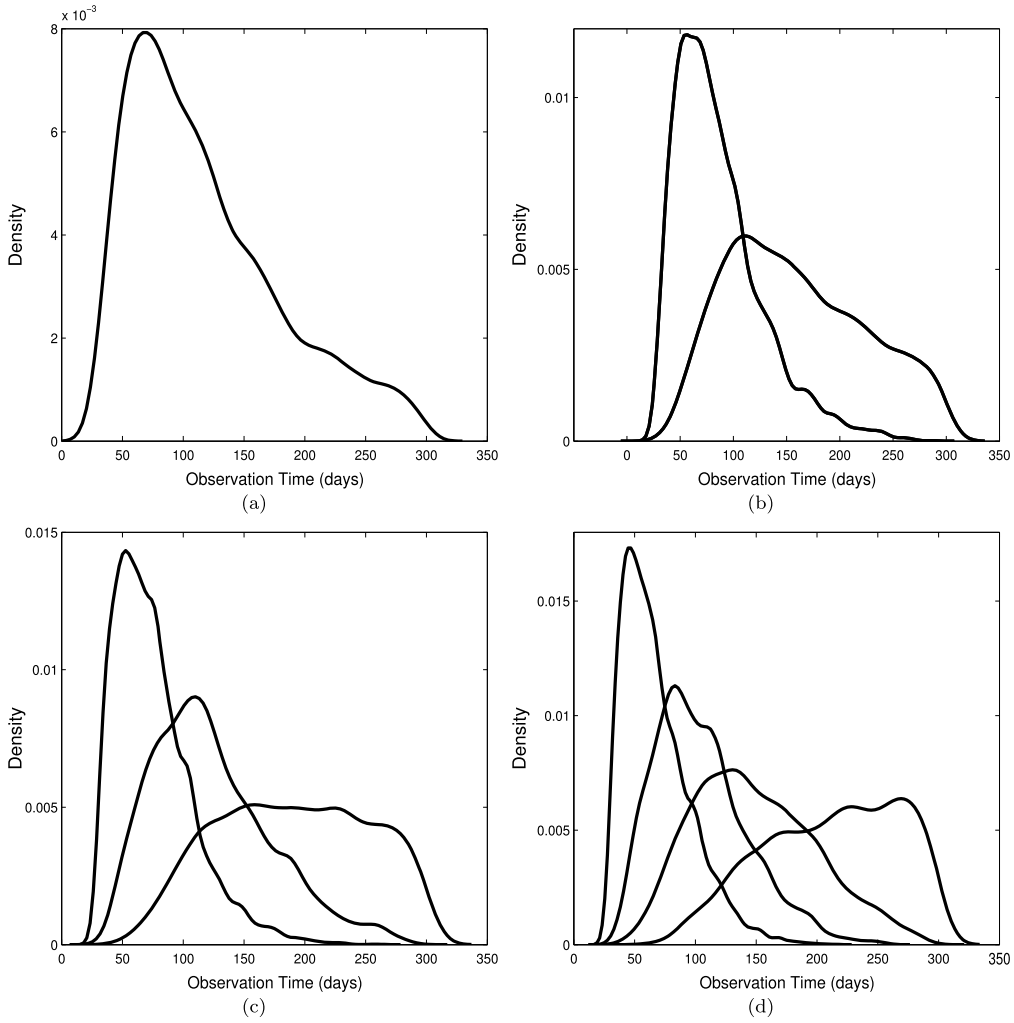
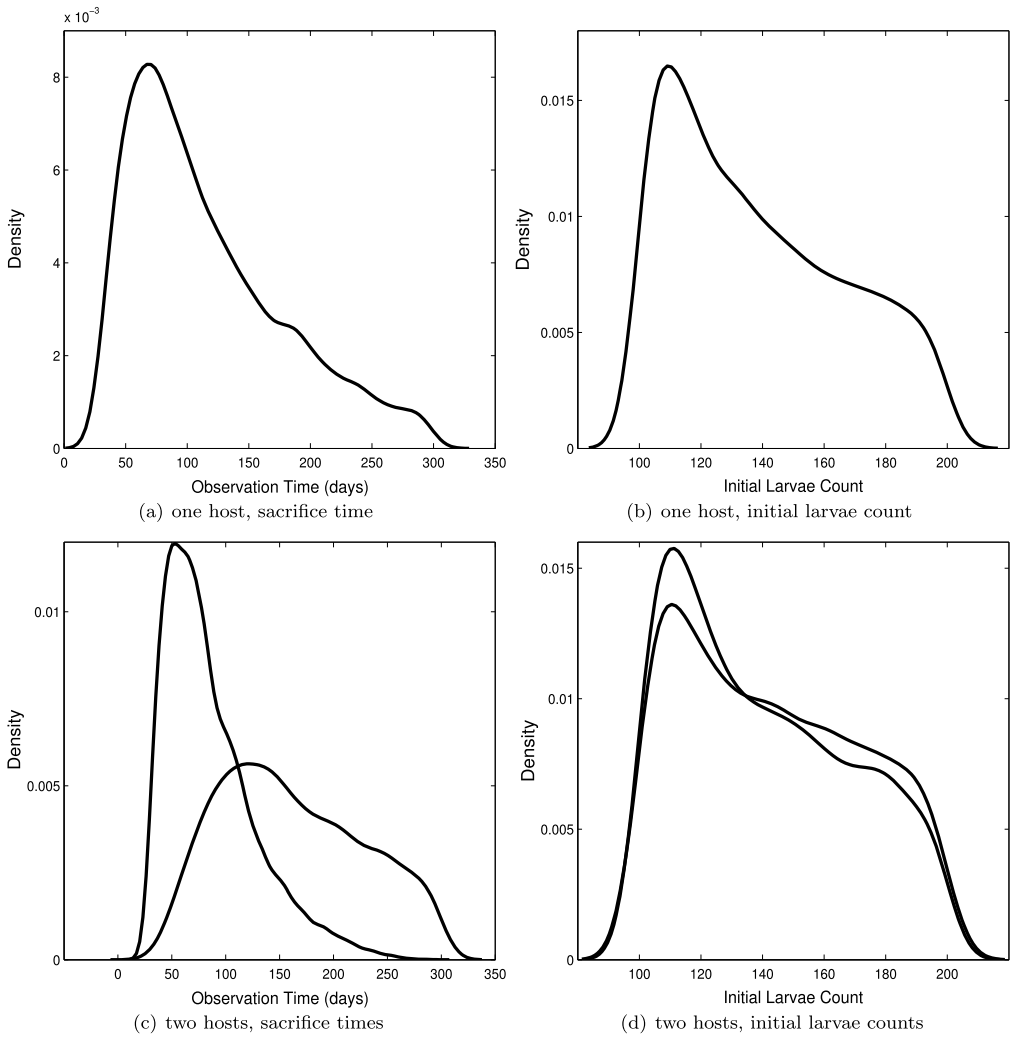


Figure 4: Marginal density estimates of the designs for 1 to 4 observations of the macroparasite model (plots (a) to (d), respectively) for 100, 000 iterations of the Müller algorithm using II design.

run the Müller algorithm for 100,000 iterations. Marginal density plots for the design parameters for 1 and 2 hosts are shown in Figure 5. The multivariate mode for each design scenario is estimated using the procedure described earlier. For one host, the optimal observation time is 69.7 days with an initial larvae count of 111. For two hosts, the first host has an optimal observation time of 60.9 days with an initial larvae count of 108 while the corresponding optimal design for the second host is 135.3 days and 105. The estimated expected utilities (with estimated standard errors in parentheses) at these designs are 4.74×10^{11} (8.1×10^{8}) and 5.13×10^{11} (1.1×10^{9}). The optimal



(a) one host, sacrifice time

(b) one host, initial larvae count

(c) two hosts, sacrifice times

(d) two hosts, initial larvae counts

Figure 5: Marginal density estimates of the observation times and initial larvae counts designs for 1 (top row) and 2 (bottom row) observations of the macroparasite model for $100,000$ iterations of the Müller algorithm using II design. The left column shows the marginal density estimates for the observation times and the right column shows the marginal density estimates for the initial larvae counts.

larvae counts might be 100 due to the difficulties associated with edge effects and kernel density estimation.

The fact that the optimal initial larvae counts are close to 100 might seem counter-intuitive. The expected information obtained from a binomial experiment increases with the number of trials. In an attempt to validate our optimal designs, we estimated the

expected utilities of designs for one host given by those with an optimal observation time of 70 days and different initial larvae counts (100, 150, 200) using the ABC rejection approach of Drovandi and Pettitt (2013). We perform enough simulations from the prior predictive distribution so that 200 draws are sampled from (almost) the true posteriors. These 200 samples are then used the estimate the utility. The process is repeated for many possible future datasets to estimate the expected utility. The expected utilities for 100, 150 and 200 initial larvae are roughly 6.05×10^{11} (6.2×10^9), 5.32×10^{11} (4.3×10^9) and 4.93×10^{11} (2.0×10^9). It is clear that taking 100 initial larvae is the preferred design.

5 Discussion

A novel approach to Bayesian experimental design has been presented for models with intractable likelihoods. The major advantages of our method are that a tractable analytic auxiliary likelihood is used to calculate the utility and optimal designs are found on a continuous design space. Furthermore, the precomputation needs to be carried out once only and its storage requirements are not necessarily dependent on the complexity of the designs. These advantages enable the methodology to be extended to complex designs such as designs with a large number of design points, high dimensional designs and designs with a large number of observations at each design point. A further advantage of our methodology is that our simple modification to the utility (see (6)) substantially sharpened the utility surface without requiring a large value of J. Computation times were substantially reduced compared with using our methodology with large tempering values J in the tempered target distribution in the Müller algorithm. This idea is not specific to the II design approach but is generally applicable to the Müller algorithm and perhaps other design optimisation approaches. The methodology was demonstrated using two examples and compared to published results from existing methods. The death model served to validate the approach and algorithm settings. Results obtained were similar to the optimal designs inferred using the model's exact tractable likelihood. The motivating example was a stochastic model for the population evolution of macroparasites, which has an intractable likelihood and therefore an intractable utility function.

In the macroparasite example, there are obvious costs with keeping a host animal alive longer that could be incorporated into the utility function if this real monetary cost could be calibrated in terms of the utility used here. Additionally, although not the case here, one could have costs associated with the initial conditions, here the number of larvae injected into the host. Again these costs could be incorporated into the utility.

It is important to note that our II approach for estimating the utility function can be used within any optimal design search algorithm. Another option for the low-dimensional designs considered in this paper is to estimate the expected utility by Monte Carlo integration at several design points, fit a smooth surface and optimise the surface directly (see, for example, Kuo et al. (1999)).

The II approach that we use to estimate the posterior distribution has similarities with the use of emulation (e.g. Gaussian processes) within the computer experiment lit-

erature (see Bayarri et al. (2007) for a review). However, the emulator within computer experiments is used to avoid expensive model simulations at new parameter values, whereas we use II to avoid expensive likelihood calculations for different datasets encountered during the design search algorithm.

5.1 Design challenges

There are a number of open design challenges for models with tractable and intractable likelihoods to which our methodology can contribute. We outline three such challenges below.

(a) Increase in the number of design points

One general difficulty with the Müller approach to Bayesian experimental design is that, to find an optimal design for $k \geq 2$ design points, one must find a multivariate mode. This becomes more difficult as the number of design points increases. This is particularly challenging in the problem of choosing optimal observation times considered in this paper since for more than one observation time of the stochastic process, the times are ordered. We used density estimation to determine the multivariate modes of the observation times. However, this approach becomes more difficult as the number of observations k increases. As an example, we can consider the macroparasite model analysed in Section 4.2. We design for the set of optimal times of sacrifice of up to 4 times to observe independent replications of the stochastic process. That is, the set of optimal times of sacrifice of 4 additional cats, for additional observations to the 212 already observed. However, it may be of interest to design for a larger set of observation times, for example, the set of 10 optimal times of sacrifice of each additional cat.

Further, as noted by Ryan et al. (2015) in the context of Bayesian static design, importance sampling for approximating the posterior becomes inefficient when there is an increase in the number of design points. Ryan et al. (2015) considered a Laplace approximation which helps to overcome the issues associated with importance sampling. For future research we plan on exploring the possibility of forming a Laplace approximation based on the II posterior.

(b) Increase in the number of design variables

In this paper we were able to accommodate two design variables through our II design approach. In order to handle the extra variable (initial larvae count), an additional parameter was required in the auxiliary model. Bayesian design with several design variables remains a challenging problem.

(c) Increase in the size of the data set observed at each design point

In our examples, the number of observations k was equal to the number of time points. However, many problems concern a much larger number of observations k, but with a

small number of times when to observe. One such design is a weighted design, where each factor level (time point in our case), is given a weight or proportion of units to be observed at that level. This design was considered in a recent article (Weaver et al., 2015) for an accelerated life test model, where the design variable was temperature and where the design was $k = 165$ observations weighted over 2 distinct temperatures. In designing for the macroparasite model, the size of the dataset observed at each time point could be increased to ≥ 5 say, 10 host animals at 4 times (40 cats in total). Or, as a combination of design challenges (2) and (3), a design could be optimised for 10 host animals at each of 10 time points (100 host animals in total).

5.2 Advantages of our indirect inference approach to Bayesian experimental design

Using II design, it is possible to address the three key challenges outlined above. This would be difficult for competing approaches as currently presented. General Bayesian design methodology based on ABC is presented in Drovandi and Pettitt (2013) and Hainy et al. (2013). Both use the Müller algorithm but Drovandi and Pettitt (2013) precompute simulated data at a number of fixed discrete design points while Hainy et al. (2013) propose computing an ABC approximate posterior at a randomly proposed continuous design point as part of the MCMC step of the Müller algorithm. The former method is applied to the same examples as considered here while the latter method is applied to a different problem. We see the former method as a competitor for the approach presented here. Both ABC approaches suffer from deficiencies of all ABC methods such as large storage requirements and computational times, and the necessity to choose a tolerance level for matching simulated data to observed data. The ABC approach of Drovandi and Pettitt (2013), is restricted to a discretised design space, unlike our II approach, which is carried out over a continuous design space. Searching over a continuous design space using the Müller algorithm is much more efficient than discretising the design space. This is relevant to the three design challenges above. As (i) the number of design points, (ii) the number of design variables or (iii) the size of the dataset increase, the ABC approach will become computationally infeasible as the design space will need to be discretised.

A principled approach to tackle (i), an increase in the number of design points, is to re-parameterise a large number of experimental observation times over a lower dimensional space as in Ryan et al. (2014), where the Müller algorithm is operated over the lower dimensional parameterisation using the quantiles of a beta distribution. This is a flexible approach requiring the optimisation of only two parameters of the beta distribution. Using this methodology, II design could be extended to optimise a large number of experimental observations, for example, 10 or more. This is due to the fact that the set of values $\{\boldsymbol{\theta}^i, \hat{\boldsymbol{\phi}}^i\}_{i=1}^n$ is precomputed once only and storage of simulated data in this precomputation step is not required. Moreover, the size of $\{\boldsymbol{\theta}^i, \hat{\boldsymbol{\phi}}^i\}_{i=1}^n$ does not necessarily grow substantially with an increase in the number of design points. For the same reasons, our methodology can be extended to (ii), an increase in the number of design variables since the size of the precomputation storage does not necessarily scale

poorly with the number of design variables. In our paper, only one additional auxiliary parameter was required to handle the initial larvae counts.

In contrast, it would be difficult to tackle problems (i) and (ii) using the ABC approach as currently presented. This is due to the discrete nature of the design search and the necessity in the ABC approach to store all the simulated data, the size of which grows exponentially with the number of design variables. Furthermore, as (i), the number of observations to design for grows, the ABC tolerance will increase. This can be mitigated by increasing the number of prior simulations but this again increases the storage requirements. This might be where a summary statistic approach could be used. This is an area of further research.

Unlike the ABC approach, II design can also address (iii), an increase in the size of the dataset observed at each factor level or time point. Since the auxiliary likelihood is analytic, the auxiliary posterior can be approximated reasonably efficiently for a large design dataset. Secondly, the same arguments hold as above since the estimation and size of the set of precomputed values $\{\boldsymbol{\theta}^i, \hat{\boldsymbol{\phi}}^i\}_{i=1}^n$ are not necessarily dependent on the size of the dataset observed at each factor level. A more complex auxiliary model may be required for more complex design problems. However, our II methodology will scale more readily than the competing ABC approach.

5.3 Other areas for further research

A critical assumption of II design is that an adequate auxiliary statistical model for the data is available (Assumption (iv), Section 3). However, it may not be possible or straightforward to find a suitable auxiliary model. One solution is to use a summary statistic approach, for example a Bayesian version of the synthetic likelihood (Wood, 2010). In this approach the auxiliary likelihood is a multivariate normal distribution fitted to a set of summary statistics rather than the full simulated data. This is a simple auxiliary model, where the mean and variance of the normal distribution are the auxiliary parameters. Our method could be extended for problems which violate Assumption (iv) using the synthetic likelihood approach. One issue is the choice of suitable summary statistics.

In the current setup, a training design d_T is used in line 3, Algorithm 1, to estimate the relationship between the generative model and the auxiliary model through a noisy mapping $\boldsymbol{\phi} = \hat{g}(\boldsymbol{\theta})$. The choice of design d_T is itself a design problem and requires further research. For simplicity, in the macroparasite example, we used a simple random training design but it may be possible to improve estimation by relaxing this restriction. For the death model, we found that 20 observations with $m = 3$ replicates in line 3, Algorithm 1, was sufficiently informative for estimation of $g(\boldsymbol{\theta})$. Alternatively, with the same computational budget, one could use the 60 observations with $m = 1$ replicate or 10 observations with $m = 6$ replicates. A general guideline is to choose the training design d_T that produces the least variable estimate of $g(\boldsymbol{\theta})$ within the available precomputation time.

Reducing the noise in the map estimation is another area of further research. Assuming that the auxiliary estimator is consistent, as $m \to \infty$, $\hat{g}(\boldsymbol{\theta}) \to g(\boldsymbol{\theta})$, reducing the

noise in the auxiliary parameter estimates. However, as the noise or error decreases at a rate $1/\sqrt{m}$, very large values of m would be required to have a substantial effect. This substantial increase in computational complexity may be unnecessary in the context of optimal experimental design, since high accuracy is not required. However, one could avoid the computational cost of increasing m and simultaneously (perhaps) improve the optimal design results, by using for example, a local multivariate smoother such as a spline, a kernel smoother or a Gaussian process, to recover a smoother estimate of $g(\boldsymbol{\theta})$, based on the noisy mapping $\hat{g}(\boldsymbol{\theta})$.

References

Bayarri, M. J., Berger, J. O., Paulo, R., Sacks, J., Cafeo, J. A., Cavendish, J., Lin, C.-H., and Tu, J. (2007). "A framework for validation of computer models." *Technometrics*, 49(2): 138–154. MR2380530. doi: http://dx.doi.org/10.1198/004017007000000092. 877

Bernardo, J. M. and Smith, A. F. (2000). *Bayesian theory*. John Wiley & Sons. MR1274699. doi: http://dx.doi.org/10.1002/9780470316870. 858

Chaloner, K. and Larntz, K. (1989). "Optimal Bayesian design applied to logistic regression experiments." *Journal of Statistical Planning and Inference*, 21(2): 191–208. MR0985457. doi: http://dx.doi.org/10.1016/0378-3758(89)90004-9. 858

Chaloner, K. and Verdinelli, I. (1995). "Bayesian experimental design: A review." *Statistical Science*, 10(3): 273–304. MR1390519. 858

Clyde, M. A. (2001). "Experimental design: A Bayesian perspective." *International Encyclopedia of the Social and Behavioral Sciences*, 8: 5075–5081. 858

Cook, A. R., Gibson, G. J., and Gilligan, C. A. (2008). "Optimal observation times in experimental epidemic processes." *Biometrics*, 64(3): 860–868. MR2526637. doi: http://dx.doi.org/10.1111/j.1541-0420.2007.00931.x. 858, 865, 866, 868, 869

Cox, D. R. (1961). "Tests of separate families of hypotheses." In: *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, 105–123. MR0131927. 863, 864

Denham, D. A., Ponnudurai, T., Nelson, G. S., Guy, F., and Rogers, R. (1972). "Studies with *Brugia pahangi*." – I. Parasitological observations on primary infections of cats *(Felis catus)*. *International Journal for Parasitology*, 2(2): 239–247. 869

Drovandi, C. C., McGree, J. M., and Pettitt, A. N. (2013). "Sequential Monte Carlo for Bayesian sequentially designed experiments for discrete data." *Computational Statistics & Data Analysis*, 57(1): 320–335. MR2981091. doi: http://dx.doi.org/10.1016/j.csda.2012.05.014. 860

Drovandi, C. C. and Pettitt, A. N. (2011). "Estimation of parameters for macroparasite population evolution using approximate Bayesian computation." *Biometrics*, 67(1): 225–233. MR2898834. doi: http://dx.doi.org/10.1111/j.1541-0420.2010.01410.x. 870

Drovandi, C. C. and Pettitt, A. N. (2013). "Bayesian experimental design for models with intractable likelihoods." *Biometrics*, 69(4): 937–948. MR3146789. doi: http://dx.doi.org/10.1111/biom.12081. 858, 867, 872, 873, 876, 878

Drovandi, C. C., Pettitt, A. N., and Faddy, M. J. (2011). "Approximate Bayesian computation using indirect inference." *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 60(3): 317–337. MR2767849. doi: http://dx.doi.org/10.1111/j.1467-9876.2010.00747.x. 870, 872

Drovandi, C. C., Pettitt, A. N., and Lee, A. (2015). "Bayesian indirect inference using a parametric auxiliary model." *Statistical Science*, 30(1): 72–95. MR3317755. doi: http://dx.doi.org/10.1214/14-STS498. 858, 861, 863, 865, 872

Gallant, A. R. and McCulloch, R. E. (2009). "On the determination of general scientific models with application to asset pricing." *Journal of the American Statistical Association*, 104(485): 117–131. MR2663037. doi: http://dx.doi.org/10.1198/jasa.2009.0008. 858, 861

Gallant, A. R. and Tauchen, G. (1996). "Which moments to match?" *Econometric Theory*, 12(04): 657–681. MR1422547. doi: http://dx.doi.org/10.1017/S0266466600006976. 858

Gillespie, D. T. (1977). "Exact stochastic simulation of coupled chemical reactions." *The Journal of Physical Chemistry*, 81(25): 2340–2361. 865

Gourieroux, C., Monfort, A., and Renault, E. (1993). "Indirect inference." *Journal of Applied Econometrics*, 8(S1): S85–S118. 858

Hainy, M., Müller, W. G., and Wagner, H. (2013). *Likelihood-free simulation-based optimal design*. Technical report, Johannes Kepler University of Linz. 878

Heggland, K. and Frigessi, A. (2004). "Estimating functions in indirect inference." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 66(2): 447–462. MR2062387. doi: http://dx.doi.org/10.1111/j.1369-7412.2003.05341.x. 858, 864

Kuo, L., Soyer, R., and Wang, F. (1999). "Optimal design for quantal bioassay via Monte Carlo methods." *Bayesian Statistics VI*, 795–802. 876

Lindley, D. V. (1972). *Bayesian Statistics, a Review*. Capital City Press, Montepelier, Vermont. 858

Michael, E., Grenfell, B. T., Isham, V. S., Denham, D. A., and Bundy, D. A. P. (1998). "Modelling variability in lymphatic filariasis: Macrofilarial dynamics in the *Brugia pahangi*–cat model." In: *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 265(1391): 155–165. 859, 865, 869, 870

Moores, M. T., Drovandi, C. C., Mengersen, K. L., and Robert, C. P. (2015). "Preprocessing for approximate Bayesian computation in image analysis." *Statistics and Computing*, 25(1): 23–33. MR3304900. doi: http://dx.doi.org/10.1007/s11222-014-9525-6. 863

Müller, P., Sansó, B., and De Iorio, M. (2004). "Optimal Bayesian design by inhomogeneous Markov chain simulation." *Journal of the American Statistical Association*, 99(467): 788–798. MR2090911. doi: http://dx.doi.org/10.1198/016214504000001123. 857, 858, 859, 860

Nelder, J. A. and Mead, R. (1965). "A simplex method for function minimization." *Computer Journal*, 7(4): 308–313. 865

Ottesen, E. A. (2006). "Lymphatic filariasis: Treatment, control and elimination." *Advances in Parasitology*, 61: 395–441. 859

Raiffa, H. and Schlaifer, R. (1961). *Applied Statistical Decision Theory*. Division of Research, Graduate School of Business Administration, Harvard University. 858

Reeves, R. W. and Pettitt, A. N. (2005). "A theoretical framework for approximate Bayesian computation." In: *Proceedings of the 20th International Workshop on Statistical Modelling, Sydney*, 393–396. 858, 861

Renshaw, E. (1991). *Modelling Biological Populations in Space and Time*. Cambridge University Press, Cambridge. MR1130616. doi: http://dx.doi.org/10.1017/CBO9780511624094. 859, 865

Riley, S., Donnelly, C. A., and Ferguson, N. M. (2003). "Robust parameter estimation techniques for stochastic within-host macroparasite models." *Journal of Theoretical Biology*, 225(4): 419–430. MR2079299. doi: http://dx.doi.org/10.1016/S0022-5193(03)00266-2. 858, 859, 865, 869, 870

Ryan, E. G., Drovandi, C. C., and Pettitt, A. N. (2015). "Fully Bayesian experimental design for pharmacokinetic studies." *Entropy*, 17(3): 1063–1089. 877

Ryan, E. G., Drovandi, C. C., Thompson, M. H., and Pettitt, A. N. (2014). "Towards Bayesian experimental design for nonlinear models that require a large number of sampling times." *Computational Statistics & Data Analysis*, 70: 45–60. MR3125477. doi: http://dx.doi.org/10.1016/j.csda.2013.08.017. 878

Smith, A. A. (1993). "Estimating nonlinear time-series models using simulated vector autoregressions." *Journal of Applied Econometrics*, 8(S1): S63–S84. 858

Suswillo, R. R., Denham, D. A., and McGreevy, P. B. (1982). "The number and distribution of *Brugia pahangi* in cats at different times after a primary infection." *Acta Tropica*, 39(2): 151–156. 870

Van Laarhoven, P. J. M. and Aarts, E. H. L. (1987). *Simulated annealing: Theory and applications*. D. Reidel Publishing Company, Dordrecht. MR0904050. doi: http://dx.doi.org/10.1007/978-94-015-7744-1. 860

Weaver, B. P., Williams, B. J., Anderson-Cook, C. M., and Higdon, D. M. (2015). "Computational enhancements to Bayesian design of experiments using Gaussian processes." *Bayesian Analysis*. doi: http://dx.doi.org/10.1214/15-BA945. 878

Wood, S. N. (2010). "Statistical inference for noisy nonlinear ecological dynamic systems." *Nature*, 466(7310): 1102–1104. 879

Acknowledgments

The work of Caitríona M. Ryan was supported by the Australian Research Council Discovery Project DP1210269 and that of Christopher C. Drovandi and Anthony N. Pettitt by an Australian Research Council Discovery Project DP110100159. Anthony N. Pettitt is a chief investigator in the ARC Centre of Excellence for Mathematical and Statistical Frontiers (ACEMS). The authors thank two referees and an associate editor for comments that led to improvements in this article.