Bayesian Analysis (2020) **15**, Number 1, pp. 103–131

Bayesian Design of Experiments for Intractable Likelihood Models Using Coupled Auxiliary Models and Multivariate Emulation

Antony Overstall* and James McGree†

Abstract. A Bayesian design is given by maximising an expected utility over a design space. The utility is chosen to represent the aim of the experiment and its expectation is taken with respect to all unknowns: responses, parameters and/or models. Although straightforward in principle, there are several challenges to finding Bayesian designs in practice. Firstly, the utility and expected utility are rarely available in closed form and require approximation. Secondly, the design space can be of high-dimensionality. In the case of intractable likelihood models, these problems are compounded by the fact that the likelihood function, whose evaluation is required to approximate the expected utility, is not available in closed form. A strategy is proposed to find Bayesian designs for intractable likelihood models. It relies on the development of an automatic, auxiliary modelling approach, using multivariate Gaussian process emulators, to approximate the likelihood function. This is then combined with a copula-based approach to approximate the marginal likelihood (a quantity commonly required to evaluate many utility functions). These approximations are demonstrated on examples of stochastic process models involving experimental aims of both parameter estimation and model comparison.

Keywords: approximate Bayesian computation, approximate coordinate exchange, indirect inference, Gaussian process, model comparison, parameter estimation.

1 Introduction

Often, the dynamics underpinning a complex physical phenomenon can be modelled by a stochastic process. It is commonly the situation that the stochastic process (or model) depends on unknown parameters, time and, potentially, other controllable variables. In this paper, we consider the case where an experiment is to be performed to learn about the phenomenon by estimating the unknown parameters. That is, the physical phenomenon of interest is observed at a series of time points, after the specification of any controllable variables, and the stochastic model is *fitted* to the observed responses. In particular, we focus on the optimal choice of time points and controllable variables (collectively referred to as design variables) to best learn about the unknown process.

*Southampton Statistical Sciences Research Institute (S3RI), University of Southampton, Southampton, United Kingdom, A.M.Overstall@soton.ac.uk

†Mathematical Sciences, Queensland University of Technology, Brisbane, Australia, james.mcgree@qut.edu.au

 https://doi.org/10.1214/19-BA1144

A feature of the stochastic models studied in this paper is that, although the dynamics behind each process can be relatively simple, the probability model linking parameters and design variables to responses is typically only defined implicitly. The development of new statistical methodology, so called *likelihood-free* methodology, to analyse observed responses under these intractable likelihood models has received much attention in recent years, e.g. approximate Bayesian computation (Tavaré et al. 1997); synthetic likelihood (Wood, 2010); variational Bayes (Tran et al., 2017) and auxiliary modelling (Gourieroux et al. 1993). The task of designing the experiment, i.e. specifying the design variables, has received significantly less attention. Under the frequentist approach to statistical inference, Pagendam and Pollett (2013) and Parker et al. (2015) used numerical approximations to the Fisher information to find D-optimal designs (e.g., Atkinson et al., 2007, Chapter 11) for stochastic epidemic and queueing models, respectively. In this paper, the Bayesian approach to statistical inference is used. Under such an approach, a utility function is specified representing the aim of the experiment. Then, a Bayesian design (Chaloner and Verdinelli, 1995) is found by maximising the expectation of the utility where expectation is with respect to all unknown quantities (i.e. parameters and unobserved responses) and the maximisation is over the space of all possible designs. Finding optimal designs under the Bayesian approach, even for tractable likelihood models, is a significant computational challenge; see recent reviews of the field by Ryan et al. (2016b) and Woods et al. (2017). Typically, neither the utility function nor its expectation are available in closed form, and the space of all possible designs can be high-dimensional. For intractable likelihood models, the problem is further exacerbated by there being no closed form expression for the likelihood. Approaches for finding Bayesian designs have been proposed that use approximate Bayesian computation (Drovandi and Pettitt 2013, Hainy et al. 2013, Price et al. 2016) and auxiliary modelling (also known as indirect inference; Ryan et al. 2016a) to approximate the likelihood. A common feature of these methodologies is that they have only been applied for examples of experiments with low-dimensional design spaces rendering them of limited practical relevance.

In this paper, we aim to overcome the shortcomings of existing approaches. The contribution is threefold. First, we apply the latest methods (Overstall and Woods, 2017) for maximising the approximate expected utility which are suitable for high-dimensional design spaces. Secondly, we develop an automatic, flexible non-parametric auxiliary modelling approach to approximate the likelihood that uses a multivariate emulator to provide an approximate link between parameters, design variables and the responses. Thirdly, we develop a novel copula-based approximation to the marginal likelihood which is a key quantity for many commonly-used utility functions and is rarely analytically tractable. The paper is organised as follows. In Section 2, we review the necessary background on Bayesian design and likelihood-free methodology. In Section 3 we describe the proposed automatic auxiliary modelling approach and the approximation to the marginal likelihood, before demonstrating the proposed methodology on illustrative yet challenging examples in Section 4. Lastly, we describe and demonstrate how the approach can be extended to the experimental aim of model comparison in Section 5.

2 Background

2.1 Intractable likelihood model

Suppose the experiment consists of n runs. For $k = 1, \ldots, n$, the kth run involves the specification of a $w \times 1$ vector of design variables $\mathbf{d}_k \in \mathcal{D}$. Let y_k be the corresponding response from the phenomenon for the kth run. It is assumed that, independently,

$$y_k \sim \mathcal{F}\left(\boldsymbol{\theta}, \mathbf{d}_k\right), \tag{1}$$

where $\mathcal{F}$ is a distribution depending on a $p \times 1$ vector of unknown parameters $\boldsymbol{\theta} \in \Theta$, with Θ the parameter space. Let $f(y|\boldsymbol{\theta}, \mathbf{d})$ denote the probability density function (pdf) or probability mass function (pmf) of the distribution $\mathcal{F}$. The likelihood is

$$\pi(\mathbf{y}|\boldsymbol{\theta}, \mathbf{D}) = \prod_{k=1}^{n} f(y_k|\boldsymbol{\theta}, \mathbf{d}_k),$$

where $\mathbf{y} = (y_1, \ldots, y_n)$ is the $n \times 1$ vector of responses, $\mathbf{D} = (\mathbf{d}_1, \ldots, \mathbf{d}_n) \in \Delta = \mathcal{D}^n$ is the vector giving the design, and Δ the q-dimensional design space with $q = nw$. For the models considered in this paper, $\mathcal{F}$ is only defined implicitly with the result that $f(y|\boldsymbol{\theta}, \mathbf{d})$ and $\pi(\mathbf{y}|\boldsymbol{\theta}, \mathbf{D})$ are not available in closed form. The marginal model is the distribution of $\mathbf{y}|\mathbf{D}$ having marginalised over the parameters $\boldsymbol{\theta}$. The pdf/pmf of this distribution is called the marginal likelihood (also known as evidence) and given by

$$\pi(\mathbf{y}|\mathbf{D}) = \int_{\Theta} \pi(\mathbf{y}|\boldsymbol{\theta}, \mathbf{D})\pi(\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta}, \tag{2}$$

where $\pi(\boldsymbol{\theta})$ is the pdf of the prior distribution for $\boldsymbol{\theta}$.

2.2 Bayesian optimal design of experiments

We initially describe the concept of Bayesian optimal design of experiments for the experimental aim of parameter estimation. We consider the extension to model comparison in Section 5. Bayesian optimal design of experiments begins with the specification of a utility function denoted by $u(\boldsymbol{\theta}, \mathbf{y}, \mathbf{D})$ which represents the *utility* of estimating $\boldsymbol{\theta}$ using observed responses $\mathbf{y}$ generated via design $\mathbf{D}$. A Bayesian optimal design is given by maximising (over Δ) the expected utility function given by

$$U(\mathbf{D}) = \int u(\boldsymbol{\theta}, \mathbf{y}, \mathbf{D})\mathrm{dP}_{\boldsymbol{\theta},\mathbf{y}|\mathbf{D}}, \tag{3}$$

where the expectation is with respect to the joint distribution of all unknown quantities; $\boldsymbol{\theta}$ and $\mathbf{y}$. In this paper we consider a class of utility functions which we term *likelihood-based*. This is where the utility function is a functional of the likelihood, $\pi(\mathbf{y}|\boldsymbol{\theta}, \mathbf{D})$, and the marginal likelihood, $\pi(\mathbf{y}|\mathbf{D})$.

This class of utility function includes many commonly-employed utilities. As an example, consider the Shannon information gain (SIG; Lindley 1956) given by

$$u_S(\boldsymbol{\theta}, \mathbf{y}, \mathbf{D}) = \log \pi(\mathbf{y}|\boldsymbol{\theta}, \mathbf{D}) - \log \pi(\mathbf{y}|\mathbf{D}). \tag{4}$$

The Bayesian design under the SIG utility is equivalently the design that maximises the expected (with respect to the marginal distribution of $\mathbf{y}$) Kullback Leibler divergence between the prior and posterior distributions of $\boldsymbol{\theta}$.

Although conceptually straightforward, there are at least two hurdles to finding Bayesian designs in practice (even for tractable likelihood models). Firstly, neither the utility function $u(\boldsymbol{\theta}, \mathbf{y}, \mathbf{D})$ nor its expectation $U(\mathbf{D})$ are usually analytically tractable and will require approximation. Secondly, the design space Δ can be of high dimensionality, i.e. q can be relatively large.

Methods proposed in the literature for approximately maximising the expected utility can be broadly classified into simulation- or smoothing-based. The simulation-based approach of Müller (1999) places an artificial joint distribution on $\boldsymbol{\theta}$, $\mathbf{y}$ and $\mathbf{D}$ such that the marginal pdf of $\mathbf{D}$ is proportional to $U(\mathbf{D})$. Simulation methods are used to generate a sample from this joint distribution which is then used to estimate the marginal mode of $\mathbf{D}$, i.e. the Bayesian design. This method has been further refined, for example, by Müller et al. (2004) and Amzal et al. (2006). However, difficulties in efficiently sampling over a high dimensional space mean that the typical limit of dimensionality for the design space under these methods is considered to be $q = 4$ (e.g. Ryan et al. 2016b).

Smoothing-based approaches are based on the following Monte Carlo approximation to the expected utility

$$\tilde{U}(\mathbf{D}) = \frac{1}{B} \sum_{i=1}^{B} u(\boldsymbol{\theta}_i, \mathbf{y}_i, \mathbf{D}), \tag{5}$$

where $\{\boldsymbol{\theta}_i, \mathbf{y}_i\}_{i=1}^{B}$ is a sample of size B generated from the joint distribution of $\boldsymbol{\theta}$ and $\mathbf{y}$ (given $\mathbf{D}$). Due to the stochastic nature of the Monte Carlo approximation, application of standard optimisation methods (e.g. Lange, 2013) is difficult. Instead, Müller and Parmigiani (1995) proposed a method whereby $\tilde{U}(\mathbf{D})$ is evaluated at a series of designs and a statistical model (a smoother or emulator) is fitted that is able to predict $\tilde{U}(\mathbf{D})$ (and therefore $U(\mathbf{D})$) at any $\mathbf{D} \in \Delta$. This predictor is then maximised over the design space, Δ. Müller and Parmigiani (1995) were able to consider design spaces with dimensionality of $q = 2$. This method has been further refined by Weaver et al. (2016) (with maximum $q = 3$) and Jones et al. (2016) (with maximum $q = 9$).

To aid in the applicability to design spaces of higher dimensionality, Overstall and Woods (2017) proposed the approximate coordinate exchange (ACE) algorithm. Here a cyclic ascent algorithm (usually referred to as coordinate exchange in the design of experiments literature; see Meyer and Nachtsheim 1995) is used to maximise the expected utility. At each of the q elements (coordinates) of the design, $\tilde{U}(\mathbf{D})$ is evaluated at a series of designs which only differ in that coordinate and a Gaussian process smoother is fitted to the resulting evaluations and used to predict $U(\mathbf{D})$ for any design. This prediction is then maximised over the one-dimensional design space of the coordinate under study. By using this methodology, Overstall and Woods (2017) were able to find approximately optimal designs for experiments in examples with design spaces of up to $q = 192$ dimensions, i.e. nearly two orders of magnitude greater than existing methods. A brief description of the ACE algorithm is given in Section 1 of the Supplementary Material (Overstall and McGree, 2019). Furthermore, the algorithm is implemented in

the `acebayes` (Overstall et al., 2018b) `R` package. The ACE algorithm is currently the state of the art in computing Bayesian designs for realistic-sized design spaces and, for this reason, we use it in all examples. However the methodology we propose is suitable to use with any optimisation method which only requires the evaluation of the Monte Carlo approximation to the expected utility given by (5).

To apply any optimisation method relying on evaluation of the Monte Carlo approximation to the expected utility given by (5), it is a requirement to be able to evaluate the utility function $u(\boldsymbol{\theta}, \mathbf{y}, \mathbf{D})$. However, it is usually the case that the utility function itself is analytically intractable. Specifically, likelihood-based utilities depend on the marginal likelihood, e.g. the SIG utility given by (4), which is typically not available in closed form. The obvious approach is to use a further (inner) Monte Carlo approximation resulting in a *nested Monte Carlo* approximation to the expected utility (Ryan, 2003; Huan and Marzouk, 2013; Overstall and Woods, 2017). For example, to approximate the SIG utility, we generate a further sample, $\{\tilde{\boldsymbol{\theta}}_j\}_{j=1}^{C}$, of size C from the prior distribution of $\boldsymbol{\theta}$. The SIG utility is then approximated by

$$\tilde{u}_S(\boldsymbol{\theta}, \mathbf{y}, \mathbf{D}) = \log \pi(\mathbf{y}|\boldsymbol{\theta}, \mathbf{D}) - \log \tilde{\pi}(\mathbf{y}|\mathbf{D}), \tag{6}$$

where the inner Monte Carlo approximation to the marginal likelihood is

$$\tilde{\pi}(\mathbf{y}|\mathbf{D}) = \frac{1}{C}\sum_{j=1}^{C} \pi(\mathbf{y}|\tilde{\boldsymbol{\theta}}_j, \mathbf{D}). \tag{7}$$

2.3 Bayesian design for intractable likelihood models

Finding Bayesian designs becomes impossible under an intractable likelihood model using the methods described in Section 2.2 which rely on a large number of evaluations of the likelihood $\pi(\mathbf{y}|\boldsymbol{\theta}, \mathbf{D})$ to approximate the expected utility. In the Monte Carlo approximation to the expected utility given by (5), the utility function is evaluated B times where each evaluation of the utility needs at least C evaluations of the likelihood for the inner Monte Carlo approximation to the marginal likelihood given by (7).

We assume at this point that, although we are unable to evaluate the likelihood, it is possible to generate samples from the intractable likelihood model. All models considered in this paper are examples of Markov process models where samples can be straightforwardly generated using the Gillespie method (Gillespie, 1977). In recent years there has been an explosion of novel methodology to evaluate the posterior distribution under an intractable likelihood depending only on the ability to generate from the model. The most popular of these methods is approximate Bayesian computation (ABC; Tavaré et al. 1997). Here, the likelihood is approximated by the ABC likelihood

$$\pi_{ABC}(\mathbf{y}|\boldsymbol{\theta}, \mathbf{D}) = \int I(\delta(\tilde{\mathbf{y}}, \mathbf{y}) \leq \epsilon) \mathrm{dP}_{\tilde{\mathbf{y}}|\boldsymbol{\theta}, \mathbf{D}},$$

where $I(A)$ is the indicator function for event A, $\delta(\tilde{\mathbf{y}}, \mathbf{y}) \geq 0$ is a discrepancy function (with $\delta(\tilde{\mathbf{y}}, \mathbf{y}) = 0$ if and only if $\tilde{\mathbf{y}} = \mathbf{y}$) and $\epsilon \geq 0$ is a specified tolerance. If $\epsilon = 0$,

the ABC likelihood is equal to the likelihood. The ABC likelihood is approximated via Monte Carlo. There is typically a trade-off between choosing ϵ to be sufficiently small to ensure accurate inference and large enough for computational efficiency. A similar approximation exists for the marginal likelihood given by (2).

Some authors (e.g., Drovandi and Pettitt, 2013; Hainy et al., 2013; Price et al., 2016) have used ABC to approximate the utility function when finding Bayesian designs. However, the ABC methodology needs the successful simultaneous specification of a discrepancy function, δ, and tolerance, ϵ. In inferential settings with fixed observations $\mathbf{y}$ and design, $\mathbf{D}$, it is possible to tailor these choices. However for Bayesian design, $\mathbf{y}$ is unknown meaning the choice of discrepancy function and tolerance need to be suitable for all observations under the marginal model and for any design $\mathbf{D} \in \Delta$. This means applying ABC techniques to find Bayesian designs for anything other than small n is difficult (e.g Dehideniya et al., 2018) and therefore Bayesian design for intractable likelihood models using ABC has been limited to design spaces of small dimensionality.

As mentioned in Section 1, an alternative methodology for inference under an intractable likelihood is auxiliary modelling (also known as indirect inference). This is a well established methodology for both frequentist (Gourieroux et al., 1993; Heggland and Frigessi, 2004) and Bayesian (Drovandi et al., 2011, 2015) inference. A *conditional auxiliary model* $\mathcal{F}_X(\boldsymbol{\theta}, \mathbf{d})$ is used to approximate the distribution, $\mathcal{F}(\boldsymbol{\theta}, \mathbf{d})$, of y given in (1). We use the term *conditional* since it is conditional on parameters $\boldsymbol{\theta}$ and to distinguish it from the marginal auxiliary model which we introduce in Section 3. Suppose the pdf/pmf of $\mathcal{F}_X$ is denoted by $f_X(y|\boldsymbol{\theta}, \mathbf{d})$, then the auxiliary likelihood is

$$\pi_X(\mathbf{y}|\boldsymbol{\theta}, \mathbf{D}) = \prod_{k=1}^{n} f_X(y_k|\boldsymbol{\theta}, \mathbf{d}_k), \tag{8}$$

which is used to approximate the likelihood $\pi(\mathbf{y}|\boldsymbol{\theta}, \mathbf{D})$ in the utility function.

The conditional auxiliary model $\mathcal{F}_X(\boldsymbol{\theta}, \mathbf{d})$ is specified by assuming that $\mathcal{F}(\boldsymbol{\theta}, \mathbf{d}) = \mathcal{H}_X(\boldsymbol{\phi}_f(\boldsymbol{\theta}, \mathbf{d}))$, where $\mathcal{H}_X(\boldsymbol{\phi}_f)$ is a probability distribution depending on v *auxiliary parameters*, $\boldsymbol{\phi}_f$, which are a function of the parameters $\boldsymbol{\theta}$ and design variables $\mathbf{d}$. The function $\boldsymbol{\phi}_f(\boldsymbol{\theta}, \mathbf{d})$ is estimated by generating samples from the model $\mathcal{F}(\boldsymbol{\theta}, \mathbf{d})$ under different parameters and design variables. Then $\mathcal{F}_X(\boldsymbol{\theta}, \mathbf{d}) = \mathcal{H}_X(\hat{\boldsymbol{\phi}}_f(\boldsymbol{\theta}, \mathbf{d}))$ where $\hat{\boldsymbol{\phi}}_f$ is the estimate of $\boldsymbol{\phi}_f$.

Ryan et al. (2016a) have previously used the auxiliary modelling approach to find Bayesian designs under intractable likelihood models. However they estimated $\boldsymbol{\phi}_f$ by imposing a parametric form which may lack flexibility. They also found designs using sampling-based approaches (see Section 2.2) and so were restricted to design spaces of small dimensionality. In Section 3.1, we consider a more flexible non-parametric form for the function $\boldsymbol{\phi}_f$.

An additional barrier to overcome is that to consider likelihood-based utilities we also need to approximate the marginal likelihood given by (2). The standard nested Monte Carlo approach (see Section 2.2) would be to replace evaluation of the likelihood $\pi(\mathbf{y}|\boldsymbol{\theta}, \mathbf{D})$ by evaluation of the auxiliary likelihood $\pi_X(\mathbf{y}|\boldsymbol{\theta}, \mathbf{D})$ in the inner Monte Carlo

approximation to the marginal likelihood given by (7). However there exists a subtle disadvantage relating to the complexity of the conditional auxiliary model. For all pairs of $i = 1, \ldots, B$ and $j = 1, \ldots, C$, in the inner Monte Carlo approximation to the marginal likelihood we need to evaluate the auxiliary likelihood

$$\pi_X(\mathbf{y}_i|\tilde{\boldsymbol{\theta}}_j, \mathbf{D}) = \exp\left(\sum_{k=1}^{n} \log f_X(y_{ik}|\tilde{\boldsymbol{\theta}}_j, \mathbf{d}_k)\right), \tag{9}$$

where y_{ik} is the kth element of $\mathbf{y}_i$. In general, the exponent on the right hand side of (9) can be decomposed as

$$\sum_{k=1}^{n} \log f_X(y_{ik}|\tilde{\boldsymbol{\theta}}_j, \mathbf{d}_k) = \sum_{k=1}^{n} \alpha(y_{ik}, \mathbf{d}_k) + \sum_{k=1}^{n} \beta(\tilde{\boldsymbol{\theta}}_j, \mathbf{d}_k) + \sum_{k=1}^{n} \gamma(y_{ik}, \tilde{\boldsymbol{\theta}}_j, \mathbf{d}_k), \tag{10}$$

for functions α, β and γ whose form depend on the exact form of $f_X(y|\boldsymbol{\theta}, \mathbf{D})$. Therefore, to evaluate the nested Monte Carlo approximation to the expected utility, α and β are evaluated $B \times n$ and $C \times n$ times each, respectively. However, γ is evaluated $B \times C \times n$ times, which can result in a high computational burden. In cases where the nested Monte Carlo approximation has been applied previously for tractable likelihood models (e.g Huan and Marzouk, 2013; Overstall and Woods, 2017), the actual (not conditional auxiliary) model is from the exponential family of distributions and γ can be decomposed as follows

$$\gamma(y, \boldsymbol{\theta}, \mathbf{d}) = \gamma_y(y, \mathbf{d})\gamma_\theta(\boldsymbol{\theta}, \mathbf{d}) \tag{11}$$

which significantly reduces the computational burden of evaluation. It transpires that γ_y and γ_θ need only be evaluated $B \times n$ and $C \times n$ times each, respectively. However, as we demonstrate in Section 4, the conditional auxiliary model typically needs to possess characteristics which are not found in exponential family distributions. For example, for count models, we have found that the negative binomial distribution provides a far more adequate conditional auxiliary model (see Section 4) than the Poisson, the latter being an exponential family distribution. In these cases, the decomposition given by (11) will typically not hold. This apparently simple complication significantly increases the computational burden of evaluating the nested Monte Carlo approximation. As an alternative to nested Monte Carlo, in Section 3.2, we propose an auxiliary modelling approximation to the marginal likelihood. In Section 4, we compare designs found under nested and auxiliary Monte Carlo both in terms of accuracy and computational time.

3 Methodology

3.1 Non-parametric estimation of ϕ_f

To estimate the function $\boldsymbol{\phi}_f$, we propose an automatic approach originating from the field of computer experiments (see, e.g., Dean et al. 2015, Section V). In this area, the goal is to approximate an unknown function (which is usually computationally expensive). To do this, the function is evaluated a "small" number of times at a specified meta-design of arguments, and a statistical model (known as an emulator) fitted to the

output. The emulator provides a prediction of the unknown function for any argument. We use the multivariate Gaussian process (MGP; Conti and O'Hagan 2010) model as the emulator. This is a multivariate generalisation of the Gaussian process model which is a commonly employed emulator in computer experiments.

We begin by generating a training sample $\{\mathbf{d}^{(i)}\}_{i=1}^{M}$ of size M from $\mathcal{D}$. We employ the usual design used for computer experiments, i.e. a space-filling Latin hypercube design (e.g. Santner et al., 2003, Chapter 5). We then generate a sample $\{\boldsymbol{\theta}^{(i)}\}_{i=1}^{M}$ of size M from the prior distribution of $\boldsymbol{\theta}$. Finally, for $i = 1, \ldots, M$ we generate an independent sample $\mathbf{y}_f^{(i)} = (y_f^{(i1)}, \ldots, y_f^{(iN)})$ of size N from $\mathcal{F}(\boldsymbol{\theta}^{(i)}, \mathbf{d}^{(i)})$.

For each of these M samples we compute the maximum likelihood estimate (MLE) of $\boldsymbol{\phi}_f$ under $\mathcal{H}_X(\boldsymbol{\phi}_f)$, i.e. let

$$\hat{\boldsymbol{\phi}}_f^{(i)} = \arg\max_{\boldsymbol{\phi}_f} \prod_{j=1}^{N} h_X(y_f^{(ij)}|\boldsymbol{\phi}_f),$$

for $i = 1, \ldots, M$, where $h_X(y|\boldsymbol{\phi}_f)$ is the pdf/pmf of $\mathcal{H}_X(\boldsymbol{\phi}_f)$. Typically, the MLE is not available in closed form so numerical methods are used.

We now have $\{\hat{\boldsymbol{\phi}}_f^{(i)}, \mathbf{d}^{(i)}, \boldsymbol{\theta}^{(i)}\}_{i=1}^{M}$ and we learn the relationship between $\boldsymbol{\phi}_f$ and $(\mathbf{d}, \boldsymbol{\theta})$ using a MGP as follows.

Let $\mathbf{Z}_f$ be the $v \times M$ matrix where the ith column (for $i = 1, \ldots, M$) is given by $\mathbf{z}_f^{(i)} = \lambda(\hat{\boldsymbol{\phi}}_f^{(i)})$ where λ is a monotonic and differentiable link function applied element-wise to $\hat{\boldsymbol{\phi}}_f^{(i)}$. The link function is applied so that the elements of $\mathbf{Z}_f$ are in $\mathbb{R}$, e.g. a log link if the auxiliary parameters are positive. Under the MGP, we assume that

$$\mathbf{Z}_f|\boldsymbol{\beta}_f, \boldsymbol{\Sigma}_f, \mathbf{A}_f \sim \mathrm{MN}\left(\boldsymbol{\beta}_f \mathbf{1}_M, \mathbf{A}_f, \boldsymbol{\Sigma}_f\right), \tag{12}$$

where $\mathrm{MN}(\boldsymbol{\beta}\mathbf{1}_M, \mathbf{A}_f, \boldsymbol{\Sigma}_f)$ denotes the matrix-normal distribution with $v \times M$ mean matrix $\boldsymbol{\beta}_f \mathbf{1}_M$, $v \times v$ unstructured row covariance matrix $\boldsymbol{\Sigma}_f$ and $M \times M$ column correlation matrix $\mathbf{A}_f$. In (12), $\mathbf{1}_M$ is an $1 \times M$ matrix of ones, and $\boldsymbol{\beta}_f$ is an $v \times 1$ matrix of coefficients. The ijth element of $\mathbf{A}_f$ is given by

$$A_{fij} = \kappa\left(\mathbf{x}_i, \mathbf{x}_j; \boldsymbol{\rho}_f\right) + \eta_f I(i = j), \tag{13}$$

where $\mathbf{x}_i = (\boldsymbol{\theta}^{(i)}, \mathbf{d}^{(i)})$, $\kappa(\cdot, \cdot; \boldsymbol{\rho}_f)$ is a valid correlation function depending on parameters $\boldsymbol{\rho}_f$, and $\eta_f > 0$ is referred to as a nugget. We will assume that $\boldsymbol{\phi}_f$ is a smooth function meaning that a suitable correlation function is the following squared exponential

$$\kappa_{SE}\left(\mathbf{x}_i, \mathbf{x}_j; \boldsymbol{\rho}_f\right) = \exp\left(-\sum_{l=1}^{s} \rho_{fl} \left(x_{il} - x_{jl}\right)^2\right), \tag{14}$$

where x_{il} and ρ_{fl} are the lth elements of $\mathbf{x}_i$ and $\boldsymbol{\rho}_f$, respectively. Note that $\boldsymbol{\rho}_f$ is an $s \times 1$ vector where $s = p + w$.

Suppose we wish to predict the value of $\mathbf{z}_f = \lambda(\boldsymbol{\phi}_f(\mathbf{x}))$ for any value of $\mathbf{x} = (\boldsymbol{\theta}, \mathbf{d})$. Under (12), the predictive distribution for $\mathbf{z}_f$ is given by

$$\mathbf{z}_f|\boldsymbol{\beta}_f, \boldsymbol{\Sigma}_f, \boldsymbol{\rho}_f, \eta_f \sim \mathrm{N}\left(\boldsymbol{\beta}_f + \left(\mathbf{Z}_f - \boldsymbol{\beta}_f \mathbf{1}_M\right) \mathbf{A}_f^{-1}\mathbf{a}_f, \left(1 + \eta_f - \mathbf{a}_f^{\mathrm{T}}\mathbf{A}_f^{-1}\mathbf{a}_f\right) \boldsymbol{\Sigma}_f\right), \quad (15)$$

where $\mathbf{a}_f$ is an $M \times 1$ vector with ith element $a_{fi} = \kappa\left(\mathbf{x}_i, \mathbf{x}; \boldsymbol{\rho}_f\right)$.

For simplicity, we specify that the function $\hat{\boldsymbol{\phi}}_f(\mathbf{x})$ is given by the inverse link of the predictive mean, i.e. the mean of (15). This depends on the parameters $\boldsymbol{\beta}_f$, $\boldsymbol{\rho}_f$ and η_f. We replace these parameters by their MLEs where the likelihood is given by (12). Thus

$$\hat{\boldsymbol{\phi}}_f(\mathbf{x}) = \lambda^{-1}\left(\hat{\boldsymbol{\beta}}_f + \left(\mathbf{Z}_f - \hat{\boldsymbol{\beta}}_f \mathbf{1}_M\right) \hat{\mathbf{A}}_f^{-1}\hat{\mathbf{a}}_f\right), \quad (16)$$

where $\hat{\mathbf{a}}_f$ and $\hat{\mathbf{A}}_f$ are $\mathbf{a}_f$ and $\mathbf{A}_f$, respectively, with $\boldsymbol{\rho}_f$ and η_f replaced by their MLEs, $\hat{\boldsymbol{\rho}}_f$ and $\hat{\eta}_f$, respectively.

Finding the conditional auxiliary model using the above approach does carry high computational burden. However it can be entirely completed *off-line*, i.e. prior to starting any algorithm for maximising the approximate expected utility. Although we have prescribed an automatic approach to estimating the function $\boldsymbol{\phi}_f$, we still need to specify the distribution $\mathcal{H}_X$. In Section 3.4, we propose methods for assessing the adequacy of auxiliary models. We advocate an iterative approach whereby an auxiliary model is fitted, assessed for adequacy and, in light of this assessment, potentially updated, i.e. the same approach one would use for statistical modelling of physical data.

3.2 Approximating the marginal likelihood

For $k = 1, \ldots, n$, let $\mathcal{G}(\mathbf{d}_k)$ be the marginal distribution of y_k having marginalised over the parameters $\boldsymbol{\theta}$, i.e. the pdf/pmf of $\mathcal{G}(\mathbf{d}_k)$ is given by

$$g(y_k|\mathbf{d}_k) = \int_\Theta f(y_k|\boldsymbol{\theta}, \mathbf{d}_k)\pi(\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta}.$$

If the elements of $\mathbf{y}$ are continuous then, by Sklar's theorem (e.g. Nelson, 1998, Section 2.3), the marginal likelihood is uniquely given by

$$\pi(\mathbf{y}|\mathbf{D}) = c\left(G(y_1|\mathbf{d}_1), \ldots, G(y_n|\mathbf{d}_n)|\mathbf{D}\right) \times \prod_{k=1}^{n} g(y_k|\mathbf{d}_k), \quad (17)$$

where c is the pdf of the copula $\mathcal{C}$ of the marginal model, $\mathbf{y}|\mathbf{D}$, and $G(y_k|\mathbf{d}_k)$ is the cumulative distribution function (cdf) of $\mathcal{G}(\mathbf{d}_k)$, for $k = 1, \ldots, n$. Suppose that $\mathbf{u} = (G(y_1|\mathbf{d}_1), \ldots, G(y_n|\mathbf{d}_n))$, then the marginal distribution of each element of $\mathbf{u}$ is $\mathrm{U}[0, 1]$ and the copula is the joint distribution of $\mathbf{u}$. Essentially a continuous multivariate probability distribution can be decomposed into the marginal distributions of each element and the copula which controls the dependency structure.

Suppose we find a suitable auxiliary model (termed the *marginal auxiliary model*), denoted by $\mathcal{G}_X(\mathbf{d})$, for $\mathcal{G}(\mathbf{d})$, with pdf $g_X(y|\mathbf{d})$ and cdf $G_X(y|\mathbf{d})$, and an *auxiliary copula*, $\mathcal{C}_X$, with pdf $c_X(\mathbf{u}|\mathbf{D})$, then the marginal likelihood can be approximated using

$$\pi_X(\mathbf{y}|\mathbf{D}) = c_X\left(G_X(y_1|\mathbf{d}_1), \ldots, G_X(y_n|\mathbf{d}_n)|\mathbf{D}\right) \times \prod_{k=1}^{n} g_X(y_k|\mathbf{d}_k). \tag{18}$$

The decomposition given by (17) is only unique for continuous $\mathbf{y}$. However, this does not preclude its use for discrete $\mathbf{y}$ (see, e.g., Panagiotelis et al., 2012).

The marginal auxiliary model is constructed in an analogous way to the construction of the conditional auxiliary model, i.e. we assume that $\mathcal{G}(\mathbf{d}) = \mathcal{H}_X(\boldsymbol{\phi}_g(\mathbf{d}))$ and we set $\mathcal{G}_X(\mathbf{d}) = \mathcal{H}_X(\hat{\boldsymbol{\phi}}_g(\mathbf{d}))$ where $\hat{\boldsymbol{\phi}}_g$ is the estimate of $\boldsymbol{\phi}_g$. The estimate $\hat{\boldsymbol{\phi}}_g$ is found using a MGP in a similar way to how we estimate $\boldsymbol{\phi}_f$ in Section 3.1. Full details are given in Section 2 of the Supplementary Material. Note that similar to the construction of $\hat{\boldsymbol{\phi}}_f$, we complete this off-line, prior to starting any algorithm for maximising the approximate expected utility.

3.3 Constructing the auxiliary copula

We now consider specifying the auxiliary copula. Similar to forming the conditional and marginal auxiliary models, we choose a family for the copula depending on an $r \times 1$ vector of *copula parameters*, $\boldsymbol{\zeta}$. In contrast to the conditional and marginal auxiliary models, the choice of copula family will be less intuitive. However, for all examples in this paper, the t-copula (e.g. Demarta and McNeil, 2005) sufficed to produce an adequate coupled auxiliary model. We compared to the simpler normal copula (e.g. Joe, 1997) and found negligible difference in terms of approximate expected utility of the designs found. However we favour the more complex t-copula for its flexibility in accounting for extreme values. Unlike the conditional and marginal auxiliary models, we propose that the specification of the auxiliary copula be made *on-line*, i.e. during evaluation of $\tilde{U}(\mathbf{D})$ within the chosen algorithm for the maximisation of the approximate expected utility. The reasoning for this difference is as follows. In the case of the conditional and marginal auxiliary models, the dimensionality of the arguments of the functions $\boldsymbol{\phi}_f$ and $\boldsymbol{\phi}_g$ are $p+w$ and w, respectively, i.e. relatively small. However, if we were to allow $\boldsymbol{\zeta}$ to be a function of $\mathbf{D}$, the dimensionality of this argument is $q = nw$, i.e. relatively large, for which it may not be possible to estimate $\boldsymbol{\zeta}$ reliably for all $\mathbf{D} \in \Delta$. At each evaluation of $\tilde{U}(\mathbf{D})$, since $\mathbf{D}$ is fixed, $\boldsymbol{\zeta}$ is independent of $\mathbf{D}$ and its value estimated using a copula training sample generated from the model. Therefore, we write the copula as $\mathcal{C}_X(\boldsymbol{\zeta})$ with pdf $c_X(\mathbf{u}|\boldsymbol{\zeta})$, i.e. independent of $\mathbf{D}$. The pdf for the t-copula with δ degrees of freedom at $\mathbf{u} = (u_1, \ldots, u_n)$ is

$$c_X(\mathbf{u}|\boldsymbol{\zeta}) = |\mathbf{R}(\boldsymbol{\gamma})|^{-\frac{1}{2}} \left[\frac{\delta + \mathbf{v}^T\mathbf{v}}{\delta + \mathbf{v}^T\mathbf{R}(\boldsymbol{\gamma})^{-1}\mathbf{v}}\right]^{\frac{\delta+n}{2}}. \tag{19}$$

In (19), $\mathbf{v}$ is an $n \times 1$ vector with kth element $v_k = T_\delta^{-1}(u_k)$, T_δ is the distribution function of the standard univariate t-distribution with δ degrees of freedom, and $\mathbf{R}(\boldsymbol{\gamma})$

is an $n \times n$ correlation matrix with $\frac{1}{2}n(n-1)$ unique elements given by the elements of $\boldsymbol{\gamma}$. The $r = \frac{1}{2}n(n-1)+1$ copula parameters, given by $\boldsymbol{\zeta} = (\boldsymbol{\gamma}, \delta)$, are estimated via a two-stage process where $\boldsymbol{\gamma}$ is estimated via method of moments and δ by maximum likelihood (e.g. Demarta and McNeil, 2005).

3.4 Assessing adequacy of auxiliary models

Before applying the auxiliary models described in the previous section to approximate the expected utility, their adequacy should be assessed for plausibility, i.e. do they provide a reasonable approximation to the assumed model. The approach proposed is based on posterior predictive assessments (see, for example, Gelman et al. 2014, Chapter 6). Here M_0 test samples are generated from the assumed and auxiliary models, and sample statistics from each compared. We propose to separately assess a) the conditional and marginal auxiliary models and; b) the coupled auxiliary model.

Assessing the conditional and marginal auxiliary models

We generate M_0 samples of size N from the assumed, $\{\bar{\mathbf{y}}_f^{(i)}\}_{i=1}^{M_0}$, and auxiliary conditional, $\{\bar{\mathbf{y}}_{fX}^{(i)}\}_{i=1}^{M_0}$, models using the following two steps.

1. Generate test samples, $\{\bar{\boldsymbol{\theta}}^{(i)}\}_{i=1}^{M_0}$ and $\{\bar{\mathbf{d}}^{(i)}\}_{i=1}^{M_0}$, of size M_0 from the prior distribution of $\boldsymbol{\theta}$ and uniformly over $\mathcal{D}$, respectively.
2. For $i = 1, \ldots, M_0$ and $j = 1, \ldots, N$ generate

$$\bar{y}_f^{(ij)} \sim \mathcal{F}\left(\bar{\boldsymbol{\theta}}^{(i)}, \bar{\mathbf{d}}^{(i)}\right) \qquad \bar{y}_{fX}^{(ij)} \sim \mathcal{H}_X\left(\hat{\boldsymbol{\phi}}_f\left(\bar{\boldsymbol{\theta}}^{(i)}, \bar{\mathbf{d}}^{(i)}\right)\right).$$

Let

$$\bar{\mathbf{y}}_f^{(i)} = \left(\bar{y}_f^{(i1)}, \ldots, \bar{y}_f^{(iN)}\right) \qquad \bar{\mathbf{y}}_{fX}^{(i)} = \left(\bar{y}_{fX}^{(i1)}, \ldots, \bar{y}_{fX}^{(iN)}\right).$$

We propose two diagnostics to compare these samples. First, plot sample statistics of the $\bar{\mathbf{y}}_f^{(i)}$'s against $\bar{\mathbf{y}}_{fX}^{(i)}$'s, where suggested sample statistics are mean, variance, median, etc. If the conditional auxiliary model is adequate then the points should approximately lie on a straight line through the origin with unit slope. Second, a single number summary of conditional auxiliary model adequacy is given by the Bayesian posterior predictive p-value (Gelman et al., 2014, page 146) given by

$$\text{p-value}_f = \frac{1}{M_0}\sum_{i=1}^{M_0} I\left(\sum_{j=1}^{N} \log h_X\left(\bar{y}_f^{(ij)} | \hat{\boldsymbol{\phi}}_f(\bar{\boldsymbol{\theta}}^{(i)}, \bar{\mathbf{d}}^{(i)})\right) < \sum_{j=1}^{N} \log h_X\left(\bar{y}_{fX}^{(ij)} | \hat{\boldsymbol{\phi}}_f(\bar{\boldsymbol{\theta}}^{(i)}, \bar{\mathbf{d}}^{(i)})\right)\right).$$

A posterior predictive p-value close to zero or one indicate that the auxiliary model is inadequate. Similar diagnostics can be obtained for the marginal auxiliary model.

Assessing the coupled auxiliary model

To assess the adequacy of the coupled auxiliary model, we generate M_0 samples of size n from both the marginal model, $\{\breve{\mathbf{y}}^{(i)}\}_{i=1}^{M_0}$, and the coupled auxiliary model, $\{\breve{\mathbf{y}}_X^{(i)}\}_{i=1}^{M_0}$. The steps required to generate these samples is given in Section 3 of the Supplementary Material. To compare these samples, we use a posterior predictive p-value given by

$$\text{p-value} = \frac{1}{M_0}\sum_{i=1}^{M_0} I\left(\log \pi_X\left(\breve{\mathbf{y}}^{(i)}|\breve{\mathbf{D}}^{(i)}\right) > \log \pi_X\left(\breve{\mathbf{y}}_X^{(i)}|\breve{\mathbf{D}}^{(i)}\right)\right).$$

Similar to assessing the conditional and marginal auxiliary models, a posterior predictive p-value close to zero or one suggests an inadequate coupled auxiliary model.

3.5 The auxiliary Monte Carlo approximation to the expected likelihood-based utility

We now summarise the steps required to approximate the expected utility given a design $\mathbf{D} = (\mathbf{d}_1, \ldots, \mathbf{d}_n)$, a Monte Carlo sample size B and a copula training sample size L. Note that these steps rely on the *off-line* construction of both the conditional auxiliary model $\mathcal{F}_X(\boldsymbol{\theta}, \mathbf{d})$ (with pdf/pmf $f_X(y|\boldsymbol{\theta}, \mathbf{d})$) and the marginal auxiliary model $\mathcal{G}_X(\mathbf{d})$ (with pdf/pmf $g_X(y|\mathbf{d})$ and distribution function $G_X(y|\mathbf{d})$).

1. Generate sample, $\{\boldsymbol{\theta}_i\}_{i=1}^B$ from the prior distribution of $\boldsymbol{\theta}$. For $i = 1, \ldots, B$ and $k = 1, \ldots, n$, generate
$$y_{ik} \sim \mathcal{F}\left(\boldsymbol{\theta}_i, \mathbf{d}_k\right),$$
and let $\mathbf{y}_i = (y_{i1}, \ldots, y_{in})$. Now $\{\mathbf{y}_i, \boldsymbol{\theta}_i\}_{i=1}^B$ is the Monte Carlo sample from the joint distribution of $\mathbf{y}$ and $\boldsymbol{\theta}$ given $\mathbf{D}$.

2. Generate sample, $\{\bar{\boldsymbol{\theta}}_i\}_{i=1}^L$ from the prior distribution of $\boldsymbol{\theta}$. For $l = 1, \ldots, L$ and $k = 1, \ldots, n$, generate
$$\bar{y}_{lk} \sim \mathcal{F}\left(\bar{\boldsymbol{\theta}}_l, \mathbf{d}_k\right),$$
and let $\bar{\mathbf{y}}_l = (\bar{y}_{l1}, \ldots, \bar{y}_{ln})$. Now $\{\bar{\mathbf{y}}\}_{l=1}^L$ is the copula training sample from the marginal distribution of $\mathbf{y}$ given $\mathbf{D}$.

3. Calculate the maximum likelihood estimates, $\hat{\boldsymbol{\zeta}}$, of $\boldsymbol{\zeta}$ where
$$\hat{\boldsymbol{\zeta}} = \arg\max_{\boldsymbol{\zeta}} \prod_{l=1}^{L} c_X\left(G_X(\bar{y}_{l1}|\mathbf{d}_1), \ldots, G_X(\bar{y}_{ln}|\mathbf{d}_n)|\boldsymbol{\zeta}\right).$$
This maximisation will need to be computed numerically since closed form maximum likelihood estimates typically do not exist for copula parameters.

4. For $i = 1, \ldots, B$, calculate the following approximations to the likelihood and marginal likelihood
$$\pi_X(\mathbf{y}_i|\boldsymbol{\theta}_i, \mathbf{D}) \quad = \quad \prod_{k=1}^{n} f_X(y_{ik}|\boldsymbol{\theta}_i, \mathbf{d}_k),$$

$$\pi_X(\mathbf{y}_i|\mathbf{D}) \quad = \quad c_X\left(G_X(y_{i1}|\mathbf{d}_1),\ldots,G_X(y_{in}|\mathbf{d}_n)|\hat{\boldsymbol{\zeta}}\right)\prod_{k=1}^{n} g_X(y_{ik}|\mathbf{d}_k).$$

5. For $i = 1,\ldots,B$, approximate the likelihood-based utility, $u(\boldsymbol{\theta}_i, \mathbf{y}_i, \mathbf{D})$, by $u_X(\boldsymbol{\theta}_i, \mathbf{y}_i, \mathbf{D})$, wherein the likelihood and marginal likelihood are replaced by $\pi_X(\mathbf{y}_i|\boldsymbol{\theta}_i, \mathbf{D})$ and $\pi_X(\mathbf{y}_i|\mathbf{D})$, respectively. The resulting approximation to the expected utility, given by

$$\tilde{U}(\mathbf{D}) = \frac{1}{B}\sum_{i=1}^{B} u_X(\mathbf{y}_i, \boldsymbol{\theta}_i, \mathbf{D}),$$

is termed the *auxiliary Monte Carlo* approximation to the expected utility.

4 Examples

We apply the proposed methodology on a series of examples. To demonstrate the methodology and assess its efficacy, in Section 4.1, we consider an illustrative example of the compartmental non-linear model where the likelihood is available in closed form. We then apply the methodology to an aphid population growth model (Section 4.2) and a parasite model (Section 4.3), both of which have been used in the literature to demonstrate Bayesian design under intractable likelihood.

First we describe some implementation details common to all examples. For the training samples, we set $M = 500$ (number of marginal and conditional auxiliary model training samples), $N = 10000$ (size of training sample size) and $L = 500$ (number of copula training samples). These were found to be sufficient in all examples to provide adequate auxiliary models. To assess adequacy, we used $M_0 = 100$ test samples. To assess the coupled auxiliary model, we compute the posterior predictive p-value for all values of n considered for each example.

4.1 Compartmental model

In this section we apply the proposed methods to find a Bayesian design under the SIG utility for a compartmental model. For this model, the likelihood is available in closed form so the aim of this example is to assess the efficacy of the approach. Compartmental models simulate how materials flow through an organism. The design problem is to specify the n sampling times $\mathbf{D} = (t_1,\ldots,t_n)$ (in hours) at which to measure the concentration of a drug in an individual, following the administration of the drug at time $t = 0$. The concentration at time t_k is denoted by y_k where it is assumed that

$$y_k \sim \mathrm{N}\left(\mu(\boldsymbol{\theta}; t_k), \nu(\boldsymbol{\theta}; t_k)\right),$$

with $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)$ being the unknown parameters,

$$\mu(\boldsymbol{\theta}; t_k) = \frac{400\theta_2}{\theta_3\left(\theta_2 - \theta_1\right)}\left(\exp\left(-\theta_1 t_k\right) - \exp\left(-\theta_2 t_k\right)\right), \qquad \nu(\boldsymbol{\theta}; t_k) = 0.1 + 0.01\mu(\boldsymbol{\theta}; t_k)^2,$$

and $n = 15$. Following Ryan et al. (2014), independent prior distributions are assumed for the elements of $\boldsymbol{\theta}$, where, on the log scale, the common variance is 0.05 and the

Approach	Mean (standard error)
Nested Monte Carlo (exact likelihood)	4.51 (0.003)
Auxiliary Monte Carlo	4.28 (0.003)
Auxiliary Monte Carlo (exact likelihood)	4.48 (0.003)
Equally-spaced design	3.70 (0.003)

Table 1: Mean (standard error) nested Monte Carlo approximation (under the exact likelihood) to the expected SIG utility for the compartmental model under the designs found under the four different approaches.

expectations are log(0.1), log(1) and log(20), respectively. Additionally, a constraint is imposed on the design whereby sampling times must be at least 15 minutes apart. Overstall and Woods (2017) describe how such constraints can be easily incorporated into the ACE algorithm.

For the distribution, $\mathcal{H}_X$, we use the normal distribution dependent on $v = 2$ auxiliary parameters, $\boldsymbol{\phi} = (\phi_1, \phi_2)$ controlling the mean and variance, respectively. The variance parameter, ϕ_2, is positive so the λ link function is chosen to be the log function for this element. After fitting the auxiliary models, the posterior predictive p-values associated with the conditional and marginal auxiliary models are p-value$_f = 0.62$ and p-value$_g = 0.42$. The posterior predictive p-value associated with the coupled auxiliary model is 0.47. Figure 1 in the Supplementary Material shows plots of sample statistics (mean and variance) of the $\mathbf{y}_f^{(i)}$'s (the $\mathbf{y}_g^{(i)}$'s) against the $\mathbf{y}_{fX}^{(i)}$'s (the $\mathbf{y}_{gX}^{(i)}$'s). These plots and the posterior predictive p-values show that the auxiliary models appear adequate.

We find Bayesian designs under the SIG utility using ACE under two different approaches. In the first, we use the auxiliary Monte Carlo approximation to the expected utility, as described in Section 3, where we approximate both the likelihood and marginal likelihood using the auxiliary models. In the second approach, we only approximate the marginal likelihood, using the coupled auxiliary model, since the likelihood for the compartmental model is available in closed form. We compare the resulting two designs to a) the SIG design found by using nested Monte Carlo (under the exact likelihood) to approximate the expected utility; and b) the design given by equally-spaced sampling times. The former was found by Overstall and Woods (2017) using the ACE algorithm. Table 1 shows the mean and standard error of twenty nested Monte Carlo approximations (under the exact likelihood) to the expected utility under each of the four designs. The design found under the methodology proposed in this paper for intractable likelihood models (i.e. second row of Table 1) performs reasonably. Obviously being able to evaluate the exact likelihood (third row of Table 1) improves this design to a point that it has performance close to the design found under nested Monte Carlo with the exact likelihood. We conclude that the methodology is competitive.

4.2 Aphid population growth model

Now consider an experiment to learn about aphid infestation in cotton plants. In the experiment, the number of aphids, y_k, in a plot of cotton plants is recorded at sampling

time t_k (in days), for $k = 1, \ldots, n$. Let $N(t)$ and $C(t)$ denote the number of current and cumulative aphid population sizes at time t. Matis et al. (2007) proposed a Markov model for the aphid population where the dynamics are given by the following equations:

$$\begin{aligned} P(N(t+\delta t) = N(t)+1, C(t+\delta t) = C(t)+1) &= \theta_1 N(t)\delta t + o(\delta t), \\ P(N(t+\delta t) = N(t)-1, C(t+\delta t) = C(t)) &= \theta_2 N(t)C(t)\delta t + o(\delta t), \end{aligned}$$

where $\boldsymbol{\theta} = (\theta_1, \theta_2)$ are the unknown parameters. Therefore the population of aphids experiences a birth rate of $\theta_1 N(t)$ and a death rate of $\theta_2 N(t)C(t)$. Note that $y_k = N(t_k)$ and the design is $\mathbf{D} = (t_1, \ldots, t_n)$. Finding designs for this experiment was considered by Gillespie and Boys (2019) under a non-likelihood-based utility given by the determinant of the posterior precision matrix for $\boldsymbol{\theta}$. They used a moment closure approach to approximate the stochastic model (Gillespie and Golightly, 2010) but only considered experiments with a low-dimensional design space. Following Gillespie and Boys (2019), we assume *a-priori* that

$$\boldsymbol{\theta} \sim \mathrm{N}\left(\begin{pmatrix} 2.46 \times 10^{-1} \\ 1.34 \times 10^{-4} \end{pmatrix}, \begin{pmatrix} 6.24 \times 10^{-5} & 5.80 \times 10^{-8} \\ 5.80 \times 10^{-8} & 4.00 \times 10^{-10} \end{pmatrix}\right),$$

let $N(0) = C(0) = 28$ and $\Delta = [0, 49]$ days. Finally, we consider a range of different experiment sizes, i.e. $n = 5, 10, \ldots, 50$.

Since the response is a count, the natural choice for $\mathcal{H}_X$ is a Poisson distribution. This has $v = 1$ auxiliary parameter giving both the mean and variance of the auxiliary model. Since this parameter is positive we employ the log link. We fit the conditional auxiliary model and assess its adequacy. The first column of Figure 1 shows plots of the sample mean (a) and sample variance (d) of the $\mathbf{y}_{fX}^{(i)}$'s against the $\mathbf{y}_f^{(i)}$'s. Whereas there appears to be good agreement between the means of the two models, the conditional auxiliary model appears to be severely underestimating the variance. Instead we use a negative binomial distribution which is a common alternative to the Poisson distribution in the presence of over-dispersion. The negative binomial distribution has $v = 2$ positive auxiliary parameters so again the log link is used. The second column of Figure 1 shows plots of the sample mean (b) and sample variance (e) of the $\mathbf{y}_{fX}^{(i)}$'s against the $\mathbf{y}_f^{(i)}$'s. The third column shows the corresponding plots for the $\mathbf{y}_{gX}^{(i)}$'s against the $\mathbf{y}_g^{(i)}$'s. Clearly there is now agreement between both the sample means and the sample variances. The posterior predictive p-values for the conditional and marginal auxiliary models are p-value$_f = 0.43$ and p-value$_g = 0.37$, respectively. For the range of different values of n, the posterior predictive p-values for the coupled auxiliary model correspondingly ranged from 0.24 to 0.47. We conclude that the auxiliary models are adequate.

We consider finding Bayesian designs under the SIG utility and an alternative likelihood-based utility that we term *likelihood ratio* (LR), given by

$$u_{LR}(\boldsymbol{\theta}, \mathbf{y}, \mathbf{d}) = 1 - \pi(\mathbf{y}|\mathbf{d})^{\frac{1}{2}} \pi(\mathbf{y}|\boldsymbol{\theta}, \mathbf{d})^{-\frac{1}{2}}.$$

The design that maximises the expected LR utility, equivalently maximises the expected Hellinger distance between the prior and posterior distributions.

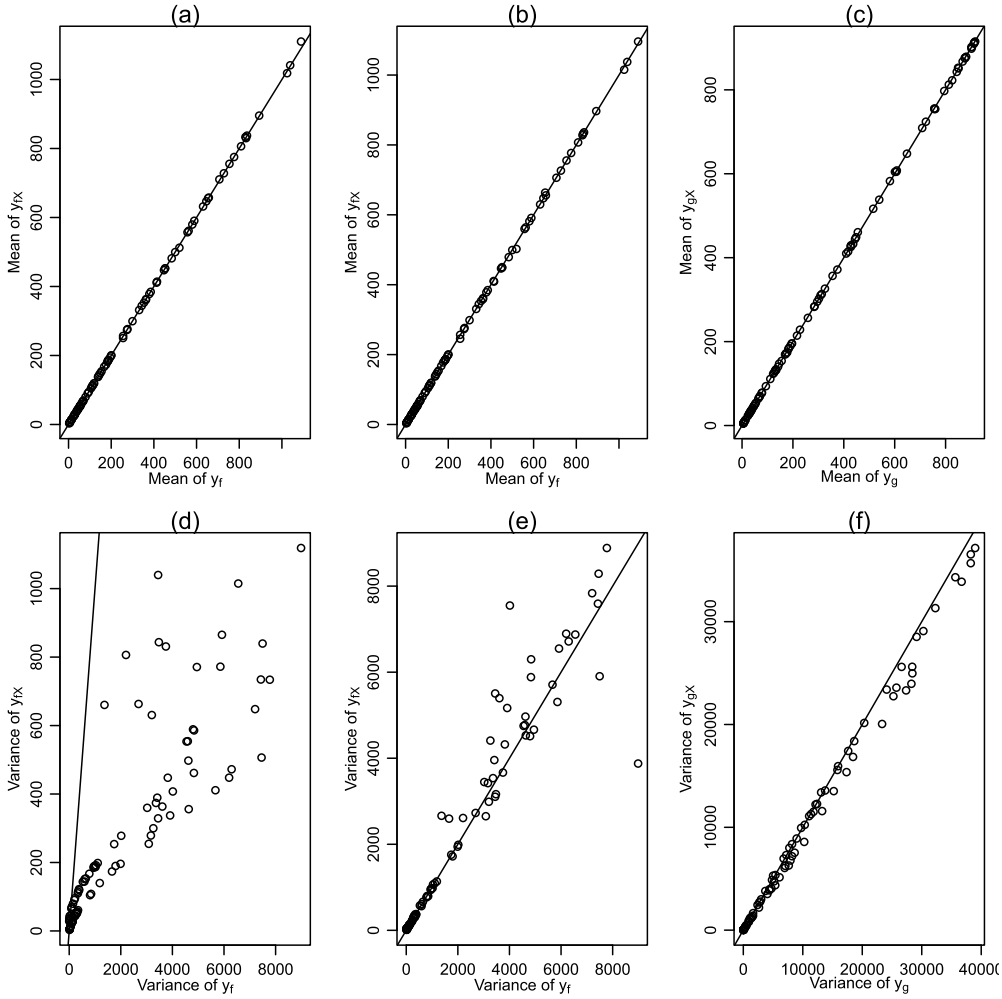

Figure 1: Graphics for assessing the adequacy of the auxiliary model for the aphid model. In the first row, (a) shows a plot of the sample mean of the $\mathbf{y}_{fX}^{(i)}$'s against the sample mean of the $\mathbf{y}_f^{(i)}$'s for the Poisson conditional auxiliary model. Corresponding plots for the negative binomial conditional (b) and marginal (c) auxiliary models. The second row shows the corresponding plots for the sample variance.

We find designs using ACE under auxiliary Monte Carlo for each value of n and each utility function. We compare these designs against the design formed from n equally spaced sampling times. Finally, designs are also found under each utility for $n = 5$ using nested Monte Carlo (using the auxiliary likelihood). Only $n = 5$ was considered due to the computational expense of finding designs using nested Monte Carlo for $n > 5$.

Number of runs, n	5	10	15	20	25	30	35	40	45	50
Auxiliary Monte Carlo	0.5	1.7	3.7	6.2	9.6	13.6	18.2	23.8	29.7	36.1
Nested Monte Carlo	8.5	-	-	-	-	-	-	-	-	-

Table 2: Average computing time (in hours) for designs found under auxiliary and nested Monte Carlo for the SIG utility and the aphid model.

The first row of Figure 2 shows the mean of twenty nested Monte Carlo approximations (under auxiliary likelihood) to the expected (a) SIG and (b) LR utilities against n for the three different types of design. As expected, in both cases, as n increases the expected gain in utility increases. The SIG and LR designs found under auxiliary Monte Carlo are superior to the equally-spaced designs, and, for $n = 5$, there is negligible difference between the designs found under nested and auxiliary Monte Carlo.

Table 2 shows the average computing times required to find the designs under the SIG utility for auxiliary and nested Monte Carlo for each value of n. The times for the LR utility are similar. For $n = 5$, it can be seen that finding the nested Monte Carlo design requires over 15 times as much computing time relative to the auxiliary Monte Carlo design. The additional computational expense of the nested Monte Carlo approximation can be explained by considering the decomposition given in (10). Under the negative binomial auxiliary model, $\alpha(y, \mathbf{d}) = -y!$, $\beta(\boldsymbol{\theta}, \mathbf{d}) = \hat{\phi}_{f2} \log \hat{\phi}_{f2} - \log \Gamma(\log \hat{\phi}_{f2})$ and

$$\gamma(y, \boldsymbol{\theta}, \mathbf{d}) = \log \Gamma(\hat{\phi}_{f2} + y) + y\hat{\phi}_{f1} - (y + \hat{\phi}_{f2}) \log(\hat{\phi}_{f1} + \hat{\phi}_{f2}),$$

where $(\hat{\phi}_{f1}, \hat{\phi}_{f2}) = \hat{\boldsymbol{\phi}}_f(\boldsymbol{\theta}, \mathbf{d})$ are the $v = 2$ estimated auxiliary parameters under the conditional auxiliary model. The function $\gamma(y, \boldsymbol{\theta}, \mathbf{d})$ cannot be written in the form of (11) due to the Gamma function and the same is true even after applying the Stirling approximation to the Gamma function (e.g. Abramowitz and Stegun, 2002, page 257).

Figure 2 (c) shows a plot of 200 samples of the aphid population size generated from the aphid model plotted against time. Figures 2 (d) and 2 (e) shows the SIG and LR designs, respectively, found under auxiliary Monte Carlo for each value of n. It can be seen in both cases, that for small values of n, the designs have sampling times concentrated in the middle of the sampling window corresponding to the "peak" in the aphid population. This qualitatively agrees with the designs found by Gillespie and Boys (2019) under their non-likelihood-based utility function. However, as n increases, we find the designs also include sampling times at the extremes of the sampling window.

4.3 Parasite model

We now consider a parasite model example modified from Drovandi and Pettitt (2013) and Ryan et al. (2016a). In the experiment, the kth host cat is injected with $d_{k1} \in [100, 200]$ *Brugia pahangi* larvae at time $t = 0$, for $k = 1, \ldots, n$. After time $d_{k2} \in (30, 300)$ (in days), the kth host cat is sacrificed and the number of mature parasites, y_k, are counted at autopsy. Riley et al. (2003) proposed a Markov process to simulate the population of parasites within the host cat. At time t, let $J(t)$ and $M(t)$ denote the number of juvenile and mature parasites, respectively. Furthermore, let $I(t)$ be a

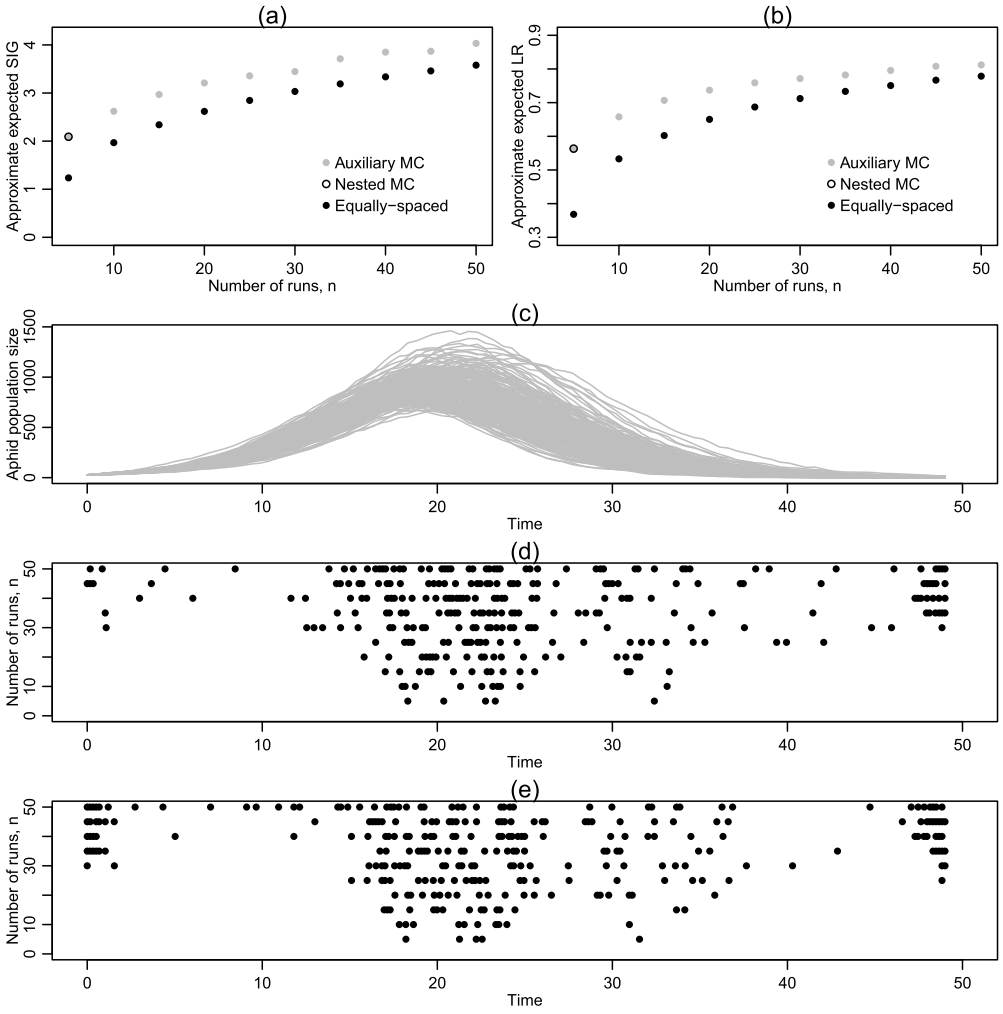

Figure 2: Plots summarising results from the aphid model. Plots (a) and (b) show the mean nested Monte Carlo approximation (under auxiliary likelihood) to the expected SIG and LR utilities, respectively, against n for designs found under the three different approaches. Plot (c) shows 200 samples of the aphid population size generated from the aphid model plotted against time. Plots (d) and (e) show the SIG and LR designs, respectively, found under auxiliary Monte Carlo for each value of n.

discrete representation of the host immunity. The dynamics of the model are as follows

$$\begin{aligned}
&\mathrm{P}\left(J(t+\delta t)=J(t)-1, M(t+\delta t)=M(t)+1, I(t+\delta t)=I(t)\right)=\theta_1 J(t)\delta t+o(\delta t),\\
&\mathrm{P}\left(J(t+\delta t)=J(t)-1, M(t+\delta t)=M(t), I(t+\delta t)=I(t)\right)\\
&\qquad\qquad =\left(\theta_4+\theta_5 I(t)\right)J(t)\delta t+o(\delta t),
\end{aligned}$$

$$\begin{aligned}
&\mathrm{P}\left(J(t+\delta t)=J(t), M(t+\delta t)=M(t)-1, I(t+\delta t)=I(t)\right)=\theta_2 M(t)\delta t+o(\delta t),\\
&\mathrm{P}\left(J(t+\delta t)=J(t), M(t+\delta t)=M(t), I(t+\delta t)=I(t)+1\right)=\theta_3 J(t)\delta t+o(\delta t),\\
&\mathrm{P}\left(J(t+\delta t)=J(t), M(t+\delta t)=M(t), I(t+\delta t)=I(t)-1\right)=\theta_6 I(t)\delta t+o(\delta t),
\end{aligned}$$

where $\boldsymbol{\theta}=(\theta_1,\ldots,\theta_6)$ are unknown parameters. Juvenile and mature parasites die with rates $\left(\theta_4+\theta_5 I(t)\right)J(t)$ and $\theta_2 M(t)$, respectively. Juvenile parasites mature with rate $\theta_1 J(t)$. The discrete measure of cat immunity increases or decreases by one unit with rates $\theta_3 J(t)$ or $\theta_6 I(t)$, respectively.

Note that for the kth run, $J(0)=d_{k1}$, $M(0)=0$, $I(0)=0$ and $M(d_{k2})=y_k$. Both Drovandi and Pettitt (2013) and Ryan et al. (2016a) fixed all parameter values except θ_3 and θ_4 and considered a design space with a maximum dimensionality of four. This was either by setting $n=4$ and fixing $d_{k1}=200$ or setting $n=2$. We consider all elements of $\boldsymbol{\theta}$ to be unknown and consider number of runs $n=2,4,6,8,10,20,30,40$, thus considering a design space with a maximum dimensionality of 80. The prior distributions for $\boldsymbol{\theta}$ follow from the analysis of responses from a previous experiment with $n=212$ (Denham et al., 1977). Following Drovandi and Pettitt (2013), the prior distribution for θ_3 and θ_4 is given by

$$\begin{pmatrix}\sqrt{\theta_3}\\ \sqrt{\theta_4}\end{pmatrix}\sim \mathrm{N}\left(\begin{pmatrix}0.0361\\ 0.0854\end{pmatrix},\begin{pmatrix}2.03\times10^{-5} & -1.07\times10^{-4}\\ -1.07\times10^{-4} & 1.17\times10^{-3}\end{pmatrix}\right).$$

The remaining parameters are given Gamma prior distributions with

$$\begin{array}{rclrcl}
\mathrm{E}(\theta_1) &=& 0.04 & \mathrm{var}(\theta_1) &=& 4.00\times10^{-4}\\
\mathrm{E}(\theta_2) &=& 0.00147 & \mathrm{var}(\theta_2) &=& 2.56\times10^{-7}\\
\mathrm{E}(\theta_5) &=& 1.10 & \mathrm{var}(\theta_5) &=& 0.21\\
\mathrm{E}(\theta_6) &=& 0.31 & \mathrm{var}(\theta_6) &=& 0.18.
\end{array} \tag{20}$$

The prior means in (20) are given by the assumed fixed values of Drovandi and Pettitt (2013), with prior standard deviations given by the corresponding standard errors found by Riley et al. (2003), inferred from 95% confidence intervals.

Since a mature parasite can only materialise from a juvenile, it means $y_k\in\{0,\ldots,d_{k1}\}$. The obvious choice is to use a binomial distribution but similar to the aphid model in Section 4.2, this was under-dispersed compared to the parasite model. Instead, we choose $\mathcal{H}_X$ to be the beta-binomial distribution. This distribution was also used by Ryan et al. (2016a). The posterior predictive p-values for the conditional and marginal auxiliary models were p-value$_f=0.20$ and p-value$_g=0.11$, respectively. The posterior predictive p-values for the coupled auxiliary models ranged from 0.39 to 0.68 (over the different values of n considered). Figure 2 in the Supplementary Material shows plots of sample statistics (mean and log variance) of the $\mathbf{y}_f^{(i)}$'s (the $\mathbf{y}_g^{(i)}$'s) against the $\mathbf{y}_{fX}^{(i)}$'s (the $\mathbf{y}_{gX}^{(i)}$'s). These plots and the posterior predictive p-values indicate that the auxiliary models are adequate.

We consider finding Bayesian designs under the SIG and LR utilities using ACE, under auxiliary Monte Carlo for each value of n. For $n=2$ we also use ACE to find

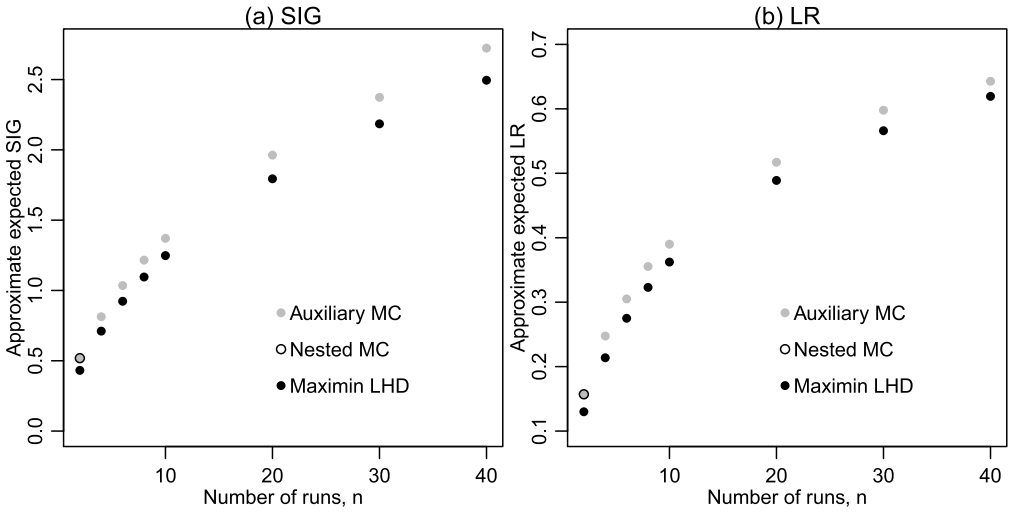

Figure 3: Plots summarising results from the parasite model. Plots (a) and (b) show the mean nested Monte Carlo approximation (under auxiliary likelihood) to the expected SIG and LR utilities, respectively, against n for designs found under the three different approaches.

a design using the nested Monte Carlo approximation (under the auxiliary likelihood). Similar to Section 4.2, we only consider $n = 2$ due to the computational expense of finding designs under nested Monte Carlo for $n > 2$. As a further comparison we also find maximin Latin hypercube designs (LHD) for each value of n.

Figure 3 shows the mean of twenty nested Monte Carlo approximations (under auxiliary likelihood) to the expected (a) SIG and (b) LR utilities against n for the three different types of design. For both utilities, the auxiliary Monte Carlo designs are superior to the maximin Latin hypercube designs and, for $n = 2$, there is negligible difference between the designs found under nested and auxiliary Monte Carlo. Table 1 in the Supplementary Material shows the average computing times required to find the designs under the SIG utility for auxiliary and nested Monte Carlo for each value of n. The computing time required to find a maximin Latin hypercube design is essentially negligible and not shown. Similar to Section 4.2, finding a design under nested Monte Carlo requires significantly more computing time than for auxiliary Monte Carlo.

5 Model comparison

5.1 Bayesian design of experiments for model comparison

Often interest lies in comparing a set $\mathcal{M}$ of competing stochastic models. An experimental aim of model comparison can be encapsulated by a utility function now denoted by $u(m, \mathbf{y}, \mathbf{D})$ where $m \in \mathcal{M}$ denotes the unknown model. Fully Bayesian inference in

this case centres on the posterior model probability of each model given by

$$\pi(m|\mathbf{y},\mathbf{d}) = \frac{\pi(\mathbf{y}|\mathbf{d},m)\pi(m)}{\sum_{m\in\mathcal{M}}\pi(\mathbf{y}|\mathbf{d},m)\pi(m)},$$

where

$$\pi(\mathbf{y}|\mathbf{d},m) = \int_{\Theta_m} \pi(\mathbf{y}|\boldsymbol{\theta}_m,\mathbf{d},m)\pi(\boldsymbol{\theta}_m|m)\mathrm{d}\boldsymbol{\theta}_m \tag{21}$$

is the marginal likelihood and $\pi(m)$ the prior model probability, respectively, for model m. In (21), $\pi(\mathbf{y}|\boldsymbol{\theta}_m,\mathbf{d},m)$ is the likelihood for model m with parameters $\boldsymbol{\theta}_m$ having prior distribution with pdf $\pi(\boldsymbol{\theta}_m|m)$. Two likelihood-based utility functions suitable for model comparison aims are

(a) the SIG utility for models given by $u_{SM}(m,\mathbf{y},\mathbf{D}) = \log\pi(\mathbf{y}|m,\mathbf{D}) - \log\pi(\mathbf{y}|\mathbf{D})$, where $\pi(\mathbf{y}|\mathbf{D}) = \sum_{m\in\mathcal{M}}\pi(\mathbf{y}|m,\mathbf{D})\pi(m)$; and

(b) the 0-1 utility given by $u_{01}(m,\mathbf{y},\mathbf{D}) = I(m=\tilde{m})$ where $\tilde{m} = \arg\max_{m\in\mathcal{M}}\pi(m|\mathbf{y},\mathbf{D})$ is the posterior modal model.

5.2 Marginal auxiliary model

Both of the utility functions for model comparison given in Section 5.1 depend on evaluation of the marginal likelihood, $\pi(\mathbf{y}|m,\mathbf{D})$, for each model $m\in\mathcal{M}$. Analogous to Section 3.2, for $k=1,\ldots,n$, let $\mathcal{G}(m,\mathbf{d})$ be the marginal distribution of y_k for model m having marginalised over the parameters $\boldsymbol{\theta}_m$. The marginal likelihood for model $m\in\mathcal{M}$ is then

$$\pi(\mathbf{y}|m,\mathbf{D}) = c\left(G(y_1|m,\mathbf{d}_1),\ldots,G(y_n|m,\mathbf{d}_n)|m,\mathbf{D}\right)\times\prod_{k=1}^{n} g(y_k|m,\mathbf{d}_k),$$

where $g(y_k|m,\mathbf{d}_k)$ and $G(y_k|m,\mathbf{d}_k)$ are the pdf/pmf and cdf of $\mathcal{G}(m,\mathbf{d})$, respectively, for $k=1,\ldots,n$, and $c(\cdot|m,\mathbf{D})$ is the copula for the marginal model, $\mathbf{y}|m,\mathbf{D}$.

A natural approach would be to find a separate coupled auxiliary model for each model, as described in Sections 3.2 and 3.3. However, one or more coupled auxiliary models may fit more adequately than the others, thus inflating the marginal likelihoods of these models. To mitigate this risk, we propose to find a separate copula for each model (as in Section 3), but form a marginal auxiliary model which is dependent on m. Specifically, we assume $\mathcal{G}(m,\mathbf{d}) = \mathcal{H}_X(\boldsymbol{\phi}_g(m,\mathbf{d}))$ and set the marginal auxiliary model to be $\mathcal{G}_X(m,\mathbf{d}) = \mathcal{H}_X(\hat{\boldsymbol{\phi}}_g(m,\mathbf{d}))$ where $\hat{\boldsymbol{\phi}}_g(m,\mathbf{d})$ is an estimate of $\boldsymbol{\phi}_g(m,\mathbf{d})$ formed using an MGP. Full details are given in Section 5 in the Supplementary Material. The most important point is that the squared exponential correlation function given in (14) is only suitable for quantitative arguments, i.e. $\mathbf{d}$ or $\boldsymbol{\theta}$, but not m. Qian et al. (2008) considered computer experiments where the arguments can be a mixture of quantitative

and categorical. We adopt their exchangeable correlation function, i.e.

$$\kappa_{SEE}\Big(\big(m^{(i)},\mathbf{d}^{(i)}\big),\big(m^{(j)},\mathbf{d}^{(j)}\big);\boldsymbol{\rho}\Big)=\exp\left(-\sum_{l=1}^{w}\rho_l(d_{il}-d_{jl})^2-\rho_{w+1}I\big(m^{(i)}\neq m^{(j)}\big)\right). \tag{22}$$

5.3 Model comparison in epidemiological dynamics

We now consider a modified version of the model comparison example considered by Dehideniya et al. (2018). The set $\mathcal{M}$ refers to a set of different epidemiological models for the spread of a disease in a given population of known size $K = 200$. The experiment involves observing y_k the number of infected individuals in the population at time t_k, for $k = 1, \ldots, n$. Thus the design is $\mathbf{D} = (t_1, \ldots, t_n)$. The population also includes exposed and susceptible individuals. Exposed individuals are those who have been exposed to the disease but are not yet infected. Susceptible individuals are those who are at risk of becoming exposed. Let $S(t)$, $E(t)$, and $I(t)$ be the number of susceptible, exposed and infected individuals, respectively, at time t, constrained such that $S(t)+E(t)+I(t) = K$. Assume that at time $t = 0$, $S(0) = K$ and $I(0) = E(0) = 0$ and note that $y_k = I(t_k)$. Dehideniya et al. (2018) considered the following four competing models.

1. **Death model** $(m = 1)$

   In the death model, individuals transition from susceptible to infected directly, i.e. they do not become exposed as an intermediate step. The rate of transition is proportional to the number of susceptible individuals left in the population. These dynamics are given by

   $$\mathrm{P}\left(S(t+\delta t)=S(t)-1, I(t+\delta t)=I(t)+1\right)=\theta_{11}S(t)\delta t+o(\delta t),$$

   where $\boldsymbol{\theta}_1 = (\theta_{11})$ is an unknown parameter.

2. **Susceptible-Infected (SI) model** $(m = 2)$

   The SI model modifies the death model so that the rate of transition from susceptible to infected is proportional to the rate at which susceptible and infected individuals meet. These dynamics are given by

   $$\mathrm{P}\left(S(t+\delta t)=S(t)-1, I(t+\delta t)=I(t)+1\right)=\left(\theta_{21}+\theta_{22}I(t)\right)S(t)\delta t+o(\delta t),$$

   where $\boldsymbol{\theta}_2 = (\theta_{21}, \theta_{22})$ are unknown parameters.

3. **Susceptible-Exposed-Infected (SEI) model** $(m = 3)$

   In the SEI model, the individuals can transition from susceptible to exposed to infected. The rate of these two transitions are proportional to the number of susceptible and exposed individuals, respectively. These dynamics are given by

   $$\begin{aligned}\mathrm{P}\left(S(t+\delta t)=S(t)-1, E(t+\delta t)=E(t)+1, I(t+\delta t)=I(t)\right)\\ =\theta_{31}S(t)\delta t+o(\delta t),\end{aligned}$$

$$\mathrm{P}\left(S(t+\delta t)=S(t), E(t+\delta t)=E(t)-1, I(t+\delta t)=I(t)+1\right) = \frac{E(t)}{\theta_{32}}\delta t + o(\delta t),$$

where $\boldsymbol{\theta}_3 = (\theta_{31}, \theta_{32})$ are unknown parameters.

4. **Susceptible-Exposed-Infected-II (SEI-II) model** ($m = 4$)

   The SEI-II model is a modification of the SEI model such that the rate of transition from susceptible to exposed is proportional to the rate at which susceptible and infected individuals meet. These dynamics are given by

$$\mathrm{P}\left(S(t+\delta t)=S(t)-1, E(t+\delta t)=E(t)+1, I(t+\delta t)=I(t)\right) = \left(\theta_{41}+\theta_{42}I(t)\right)S(t)\delta t + o(\delta t),$$

$$\mathrm{P}\left(S(t+\delta t)=S(t), E(t+\delta t)=E(t)-1, I(t+\delta t)=I(t)+1\right) = \frac{E(t)}{\theta_{43}}\delta t + o(\delta t),$$

   where $\boldsymbol{\theta}_4 = (\theta_{41}, \theta_{42}, \theta_{43})$ are unknown parameters.

We consider finding designs under the two likelihood-based utilities described in Section 5.1 for $n = 5, 10, \ldots, 50$. Dehideniya et al. (2018) also considered finding designs for the 0-1 utility function but used ABC to approximate the marginal likelihood and, therefore, only considered low-dimensional designs.

We set the prior model probabilities to be equal, i.e. $\pi(m) = 0.25$ therefore specifying that the models are *a-priori* equally likely. For the prior distribution of $\boldsymbol{\theta}_m$ under each model m, we let each element of $\boldsymbol{\theta}_m$ have the following uniform distributions

$$\begin{array}{lllllllll} \theta_{11} & \sim & \mathrm{U}[0,0.5], & & & & & & \\ \theta_{21} & \sim & \mathrm{U}[0,0.5], & \theta_{22} & \sim & \mathrm{U}[0,0.005], & & & \\ \theta_{31} & \sim & \mathrm{U}[0,0.5], & \theta_{32} & \sim & \mathrm{U}[0,10], & & & \\ \theta_{41} & \sim & \mathrm{U}[0,0.5], & \theta_{42} & \sim & \mathrm{U}[0,0.005], & \theta_{43} & \sim & \mathrm{U}[0,10]. \end{array} \tag{23}$$

These were chosen so that the distribution of $\mathbf{y}|m, \mathbf{D}$ is approximately the same for all m. To see this, Figure 3 in the Supplementary Material shows samples of $\mathbf{y}$ plotted against time under each of the models.

Since the number of infected individuals is bounded from above by K, it means that $y_k \in \{0, \ldots, K\}$. Therefore, similar to the parasite model in Section 4.3, we use the beta-binomial model for $\mathcal{H}_X$. The posterior predictive p-value for the marginal auxiliary model (for all m) is p-value$_g = 0.36$. The posterior predictive p-values for the coupled auxiliary models range from 0.20 to 0.41 over the different n considered. Figure 4 in the Supplementary Material shows plots of sample statistics (mean and variance) of the $\mathbf{y}_g^{(i)}$'s against the $\mathbf{y}_{gX}^{(i)}$'s. These plots appear to show some slight differences in the predictive variance between the four epidemiological dynamics models. However, in conjunction with the posterior predictive p-values, we consider the auxiliary models to be adequate.

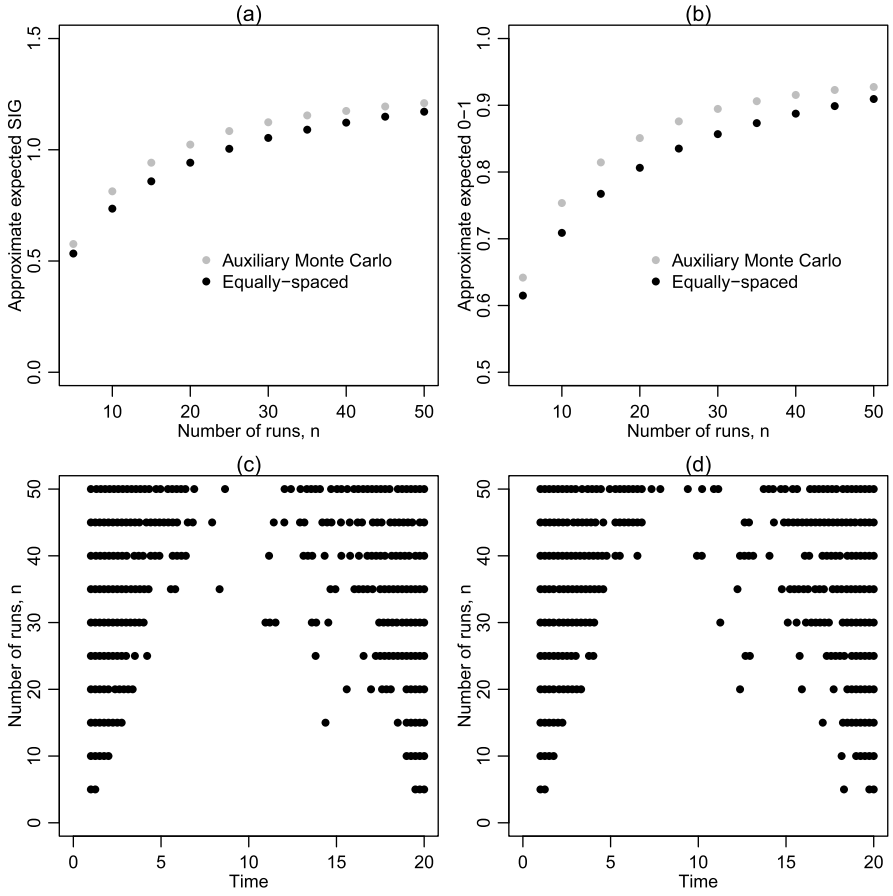

Figure 4: Plots summarising results from the epidemiological dynamics example. Plots (a) and (b) show the mean nested Monte Carlo approximation (under auxiliary likelihood) to the expected SIG and 0-1 utilities, respectively, against n for the design found under auxiliary Monte Carlo and the equally-spaced design. Plots (c) and (d) show the SIG and 0-1 designs, respectively, found under auxiliary Monte Carlo for each value of n.

The first row of Figure 4 shows the mean of twenty nested Monte Carlo approximations (under auxiliary likelihood) to the expected (a) SIG and (b) LR utilities against n for the design found under auxiliary Monte Carlo and the design given by equally-spaced sampling times. For both utilities, the auxiliary Monte Carlo designs are superior to the equally-spaced designs. The second row of Figure 4 shows the (c) SIG and (d) 0-1 designs found under auxiliary Monte Carlo for each value of n. It can be seen in both cases, that the designs appear to have sampling times at the beginning and end of the sampling window.

6 Discussion

In this paper we have introduced a general-purpose approach for finding Bayesian designs under intractable likelihood models. It is applicable for all likelihood-based utility functions, for realistic-sized experiments and for experimental aims of parameter estimation and model comparison.

The proposed methodology is not applicable for non-likelihood-based utility functions. Examples of such utilities are the negative trace of the posterior variance matrix (e.g. Overstall and Woods, 2017) or the determinant of the posterior precision matrix (e.g Gillespie and Boys, 2019). As stated by Gillespie and Boys (2019), these utilities are only suitable for non-skewed unimodal posterior distributions. Likelihood-based utilities, on the other hand, use the likelihood function to characterise information coming from the experiment, as opposed to a single summary, such as the posterior mean or variance, and therefore make no restrictions on the posterior distribution.

Ryan et al. (2016a) suggested the combination of auxiliary modelling and using normal-based approximations to posterior quantities (e.g. Long et al., 2013; Overstall et al., 2018a). However, as for our recommendation of likelihood-based utilities, we believe that the type of posterior distribution encountered in intractable likelihood models may not be well approximated by a normal distribution. Instead, other deterministic approximations, such as expectation propagation (e.g. Gelman et al., 2014, pages 338–343) could be more suitable.

Supplementary Material

Supplementary Material for "Bayesian design of experiments for intractable likelihood models using coupled auxiliary models and multivariate emulation"
(DOI: 10.1214/19-BA1144SUPP; .pdf).

References

Abramowitz, M. and Stegun, I. A. (2002). "Handbook of Mathematical Functions." Available at URL www.math.sfu.ca/~cbm/aands/toc.htm. MR0208797. 119

Amzal, B., Bois, F. Y., Parent, E., and Robert, C. P. (2006). "Bayesian-optimal design via interacting particle systems." *Journal of the American Statistical Association*, 101: 773–785. MR2281248. doi: https://doi.org/10.1198/016214505000001159. 106

Atkinson, A. C., Donev, A. N., and Tobias, R. D. (2007). *Optimum experimental designs, with SAS*. Oxford University Press Inc., New York. MR2323647. 104

Chaloner, K. and Verdinelli, I. (1995). "Bayesian Experimental Design: A Review." *Statistical Science*, 10: 273–304. MR1390519. 104

Conti, S. and O'Hagan, A. (2010). "Bayesian Emulation of Complex Multi-Output

and Dynamic Computer Models." *Journal of Statistical Planning and Inference*, 140: 640–651. MR2558393. doi: https://doi.org/10.1016/j.jspi.2009.08.006. 110

Dean, A., Morris, M., Stufken, J., and Bingham, D. (eds.) (2015). *Handbook of Design and Analysis of Experiments*. Boca Raton: CRC Press. MR3642972. 109

Dehideniya, M., Drovandi, C. C., and McGree, J. M. (2018). "Optimal Bayesian Design for Discriminating between Models with Intractable Likelihoods in Epidemiology." *Computational Statistics and Data Analysis*, 124: 277–297. MR3787626. doi: https://doi.org/10.1016/j.csda.2018.03.004. 108, 124, 125

Demarta, S. and McNeil, A. (2005). "The t Copula and Related Copulas." *International Statistical Review*, 73(1): 111–129. 112, 113

Denham, D., Ponnudurai, T., Nelson, G., Guy, F., and Rogers, R. (1977). "Studies with Brugia pahangi. I. Parasitological observations on primary infections of cats (Felis catus)." *Journal for Parasitology*, 2: 239–247. 121

Drovandi, C. C. and Pettitt, A. N. (2013). "Bayesian experimental design for models with intractable likelihoods." *Biometrics*, 69(4): 937–948. MR3146789. doi: https://doi.org/10.1111/biom.12081. 104, 108, 119, 121

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): 503–524. MR2767849. doi: https://doi.org/10.1111/j.1467-9876.2010.00747.x. 108

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: https://doi.org/10.1214/14-STS498. 108

Gelman, A., Carlin, J., Stern, H., Dunson, D., Vehtari, A., and Rubin, D. (2014). *Bayesian Data Analysis*. Chapman and Hall, 3rd edition. MR3235677. 113, 127

Gillespie, C. and Boys, R. (2019). "Efficient construction of Bayes optimal designs for stochastic process models." *Statistics and Computing*, In press. 117, 119, 127

Gillespie, C. and Golightly, A. (2010). "Bayesian inference for generalized stochastic population growth models with application to aphids." *Journal of Royal Statistical Society, Series C*, 59(2): 341–357. MR2744477. doi: https://doi.org/10.1111/j.1467-9876.2009.00696.x. 117

Gillespie, D. (1977). "Exact stochastic inference of coupled chemical reactions." *Journal of Physical Chemistry*, 81: 2340–2361. 107

Gourieroux, C., Monfort, A., and Renault, E. (1993). "Indirect Inference." *Journal of Applied Econometrics*, 8: 85–118. 104, 108

Hainy, M., Müller, W., and Wagner, H. (2013). "Likelihood-Free Simulation-Based Optimal Design: An Introduction." In Melas, V., Mignani, S., Monari, P., and Salmaso, L. (eds.), *Topics in Statistical Simulation*, 271–278. 104, 108

Heggland, K. and Frigessi, A. (2004). "Estimating functions in indirect infer-

ence." *Journal of the Royal Statistical Society Series B*, 104: 117–131. MR2062387. doi: https://doi.org/10.1111/j.1369-7412.2003.05341.x. 108

Huan, X. and Marzouk, Y. M. (2013). "Simulation-Based Optimal Bayesian Experimental Design for Nonlinear Systems." *Journal of Computational Physics*, 232(1): 288–317. MR2994301. doi: https://doi.org/10.1016/j.jcp.2012.08.013. 107, 109

Joe, H. (1997). *Multivariate models and dependence concepts*. Chapman & Hall. MR1462613. doi: https://doi.org/10.1201/b13150. 112

Jones, M., Goldstein, M., Jonathan, P., and Randell, D. (2016). "Bayes linear analysis for Bayesian optimal experimental design." *Journal of Statistical Planning and Inference*, 171: 115–129. MR3458072. doi: https://doi.org/10.1016/j.jspi.2015.10.011. 106

Lange, K. (2013). *Optimization*. Springer, 2nd edition. MR3052733. doi: https://doi.org/10.1007/978-1-4614-5838-8. 106

Lindley, D. (1956). "On a measure of the information provided by an experiment." *Annals of Mathematical Statistics*, 27: 986–1005. MR0083936. doi: https://doi.org/10.1214/aoms/1177728069. 105

Long, Q., Scavino, M., Tempone, R., and Wang, S. (2013). "Fast estimation of expected information gains for Bayesian experimental designs based on Laplace approximations." *Computer Methods in Applied Mechanics and Engineering* , 259: 24–39. MR3063259. doi: https://doi.org/10.1016/j.cma.2013.02.017. 127

Matis, J., Kiffe, T., Matis, T., and Stevenson, D. (2007). "Stochastic modeling of aphid population growth with nonlinear, power-law dynamics." *Mathematical Biosciences*, 208(2): 469–494. MR2348126. doi: https://doi.org/10.1016/j.mbs.2006.11.004. 117

Meyer, R. and Nachtsheim, C. (1995). "The coordinate-exchange algorithm for constructing exact optimal experimental designs." *Technometerics*, 37(1): 60–69. MR1322047. doi: https://doi.org/10.2307/1269153. 106

Müller, P. (1999). "Simulation-Based Optimal Design." *Bayesian Statistics*, 6: 459–474. MR1723509. 106

Müller, P. and Parmigiani, G. (1995). "Optimal Design via Curve Fitting of Monte Carlo Experiments." *Journal of the American Statistical Association*, 90(432): 1322–1330. MR1379474. 106

Müller, P., Sanso, B., and De Iorio, M. (2004). "Optimal Bayesian design by inhomogeneous Markov chain simulation." *Journal of the American Statistical Association*, 99: 788–798. MR2090911. doi: https://doi.org/10.1198/016214504000001123. 106

Nelson, R. (1998). *An Introduction to Copulas*. Springer Verlag New York. MR1653203. doi: https://doi.org/10.1007/978-1-4757-3076-0. 111

Overstall, A. and McGree, J. (2019). "Supplementary Material for "Bayesian Design of Experiments for Intractable Likelihood Models Using Coupled Auxiliary

Models and Multivariate Emulation".” *Bayesian Analysis*. doi: https://doi.org/10.1214/19-BA1144. 106

Overstall, A. M., McGree, J. M., and Drovandi, C. C. (2018a). “An approach for finding fully Bayesian optimal designs using normal-based approximations to loss functions.” *Statistics and Computing*, 28(2): 343–358. MR3747567. doi: https://doi.org/10.1007/s11222-017-9734-x. 127

Overstall, A. M. and Woods, D. C. (2017). “Bayesian Design of Experiments using Approximate Coordinate Exchange.” *Technometrics*, 59(4): 458–470. MR3740963. doi: https://doi.org/10.1080/00401706.2016.1251495. 104, 106, 107, 109, 116, 127

Overstall, A. M., Woods, D. C., and Adamou, M. (2018b). *acebayes: Optimal Bayesian Experimental Design using the ACE Algorithm*. R package version 1.6.0. URL https://cran.r-project.org/web/packages/acebayes/index.html 107

Pagendam, D. and Pollett, P. (2013). “Optimal design of experimental epidemics.” *Journal of Statistical Planning and Inference*, 143(3): 563–572. MR2995115. doi: https://doi.org/10.1016/j.jspi.2012.09.011. 104

Panagiotelis, A., Czado, C., and Joe, H. (2012). “Pair Copula Constructions for Multivariate Discrete Data.” *Journal of American Statistical Association*, 107: 1063–1072. MR3010894. doi: https://doi.org/10.1080/01621459.2012.682850. 112

Parker, B., Gilmour, S., Schormans, J., and Maruri-Aguilar, H. (2015). “Optimal design of measurements on queueing systems.” *Queueing Systems*, 79: 365–390. MR3316245. doi: https://doi.org/10.1007/s11134-014-9421-y. 104

Price, D., Bean, N., Ross, J., and Tuke, J. (2016). “On the efficient determination of optimal Bayesian experimental designs using ABC: A case study in optimal observation of epidemics.” *Journal of Statistical Planning and Inference*, 172: 1–15. MR3461907. doi: https://doi.org/10.1016/j.jspi.2015.12.008. 104, 108

Qian, P., Wu, H., and Wu, C. (2008). “Gaussian Process Models for Computer Experiments with Qualitative and Quantitative Factors.” *Technometrics*, 50: 383–396. MR2457574. doi: https://doi.org/10.1198/004017008000000262. 123

Riley, S., Donnelly, C., and Ferguson, N. (2003). “Robust parameter estimation techniques for stochastic within-host macroparasite models.” *Journal of Theoretical Biology*, 419–430. MR2079299. doi: https://doi.org/10.1016/S0022-5193(03)00266-2. 119, 121

Ryan, C., Drovandi, C., and Pettitt, A. N. (2016a). “Optimal Bayesian Experimental Design for Models with Intractable Likelihoods Using Indirect Inference Applied to Biological Process Models.” *Bayesian Analysis*, 11(3): 857–883. MR3543911. doi: https://doi.org/10.1214/15-BA977. 104, 108, 119, 121, 127

Ryan, E. G., Drovandi, C. C., McGree, J. M., and Pettitt, A. N. (2016b). “A Review of Modern Computational Algorithms for Bayesian Optimal Design.” *International Statistical Review*, 84: 128–154. MR3491282. doi: https://doi.org/10.1111/insr.12107. 104, 106

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 and Data Analysis*, 70: 45–60. MR3125477. doi: https://doi.org/10.1016/j.csda.2013.08.017. 115

Ryan, K. (2003). "Estimating expected information gains for experimental designs with application to the random fatigue-limit model." *Journal of Computational and Graphical Statistics*, 12: 585–603. MR2002637. doi: https://doi.org/10.1198/1061860032012. 107

Santner, T., Williams, B., and Notz, W. (2003). *The Design and Analysis of Computer Experiments*. Springer, New York. MR2160708. doi: https://doi.org/10.1007/978-1-4757-3799-8. 110

Tavaré, S., Balding, D. J., Griffiths, R., and Donnelly, P. (1997). "Inferring Coalescence Times from DNA Sequence Data." *Genetics*, 145(2): 505–518. 104, 107

Tran, M., Nott, D. J., and Kohn, R. (2017). "Variational Bayes with Intractable Likelihood." *Journal of Computational and Graphical Statistics*, To appear. MR3765351. doi: https://doi.org/10.1080/10618600.2017.1330205. 104

Weaver, B., Williams, B., Anderson-Cook, C., and Higdon, D. (2016). "Computational Enhancements to Bayesian Design of Experiments Using Gaussian Processes." *Bayesian Analysis*, 11(1): 191–213. MR3447096. doi: https://doi.org/10.1214/15-BA945. 106

Wood, S. (2010). "Statistical inference for noisy nonlinear ecological dynamic systems." *Nature*, 466: 1102–1104. 104

Woods, D., Overstall, A., Adamou, M., and Waite, T. (2017). "Bayesian design of experiments for generalised linear models and dimensional analysis with industrial and scientific application." *Quality Engineering*, 29(1): 91–103. 104

Acknowledgments

We would like to thank the editor, an associate editor and a referee for taking the time to read the paper. They have providing us with comments which have significantly improved the quality of the paper.