Bayesian design of experiments for industrial and scientific applications via Gaussian processes

David C. Woods
Southampton Statistical Sciences Research Institute,
University of Southampton,
Southampton SO17 1BJ UK
D.Woods@southampton.ac.uk

Abstract

The design of an experiment can be considered to be at least implicitly Bayesian, with prior knowledge used informally to aid decisions such as the variables to be studied and the choice of a plausible relationship between the explanatory variables and measured responses. Bayesian methods allow uncertainty in these decisions to be incorporated into design selection through prior distributions that encapsulate information available from scientific knowledge or previous experimentation. Further, a design may be explicitly tailored to the aim of the experiment through a decision-theoretic approach using an appropriate loss function. We review the area of decision-theoretic Bayesian design, with particular emphasis on recent advances in computational methods.

For many problems arising in industry and science, Bayesian design is often seen as impractical, particularly for finding designs for nonlinear models that have intractable expected loss and for larger factorial experiments with many potential response models. We describe how Gaussian process emulation, commonly used in computer experiments, can play an important role in facilitating Bayesian design for realistic problems. A main focus is the combination of Gaussian process regression to approximate the expected loss with cyclic descent (coordinate exchange) optimisation algorithms to allow optimal designs to be found for previously infeasible problems. The methods are motivated and illustrated using applications through the pharmaceutical and biological sciences.

Keywords: Computer experiments; *D*-optimality; generalised linear models; high-dimensional design; nonlinear models; Smoothing.

1   Introduction

Design of experiments is an "a priori" activity, taking place before data has been collected, and hence the Bayesian paradigm is a particularly appropriate approach to take. Bayesian methods allow available prior information on the model to be incorporated into both the design of the experiment and the analysis of the resulting data, and produce posterior distributions that are interpretable by scientists. They also reduce reliance on unrealistic assumptions and asymptotic results that may be inappropriate for small- to medium-sized experiments. The Bayesian approach to design enables realistic and coherent accounting for the substantial model and parameter uncertainties that usually exist before an experiment is performed and is also a natural framework for sequential inference and design.

An important problem where Bayesian methods can have substantial impact is optimal design for nonlinear modelling, which relies on some prior information being available about the unknown values of the model parameters (see Atkinson et al., 2007, ch. 17). A Bayesian approach relaxes the assumption of specifying particular values of the parameters, as required by local optimality criteria. Fully Bayesian design, as outlined below, also relies less on the asymptotic assumptions that underpin most classical design for nonlinear models.

A decision-theoretic Bayesian optimal design is found through minimisation of the expectation of a loss function that is chosen to encapsulate the aims of the experiment. Suppose that we require a design for q

1

variables in n points, with the ith point defined as $\mathbf{x}_i = (x_{i1}, \ldots, x_{iq})^{\mathrm{T}} \in \mathcal{X} \subset \mathbb{R}^q$. Let $l(\xi, \mathbf{y}, \boldsymbol{\psi})$ be the loss function for a design $\xi = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\} \in \mathcal{X}^n$ producing data $\mathbf{y} = (y_1, \ldots, y_n)^{\mathrm{T}} \in \mathcal{Y}$. Assume a statistical model defined via likelihood $p(\mathbf{y}|\boldsymbol{\psi})$, with parameters $\boldsymbol{\psi} \in \Psi$ having prior density $p(\boldsymbol{\psi})$. Then an optimal design $\xi^\star$ is defined as

$$\xi^\star = \operatorname{argmin}_{\xi \in \Xi} \int_{\mathcal{Y}} \int_{\Psi} l(\xi, \mathbf{y}, \boldsymbol{\psi}) p(\boldsymbol{\psi}, \mathbf{y}|\xi) \, d\boldsymbol{\psi} \, d\mathbf{y} \,. \tag{1}$$

See the landmark review paper of Chaloner and Verdinelli (1995).

There are a number of difficulties in calculating the expected loss in (1):

(a) the evaluation of l itself is potentially non-trivial, as it may depend on the posterior distribution and only be available numerically;

(b) evaluation of the joint density $p(\boldsymbol{\psi}, \mathbf{y}|\xi) = p(\mathbf{y}|\boldsymbol{\psi}, \xi) p(\boldsymbol{\psi})$ may be complicated by the computational expense of calculating the likelihood for numerical models;

(c) the integrals in (1) may be very high dimensional, and are unlikely to be analytically tractable;

(d) the joint probability model must take account of any hierarchies in the experiment structure (e.g. through the application of linear or nonlinear mixed models).

Common choices for l include (i) the negative of the gain in Shannon information, or Kullback-Leibler (KL) divergence, between the posterior and prior densities, leading to a maximisation of the mutual information between $\mathbf{y}$ and $\boldsymbol{\psi}$, and (ii) the squared error loss between $\boldsymbol{\psi}$ and its posterior expectation. For some experiments, more bespoke losses may be required, for example, incorporating a cost for each run of the experiment. We demonstrate results using the negative Shannon information gain (NSIG), defined as

$$\Phi(\xi) = \int_{\mathcal{Y}} \int_{\Psi} \left[\log p(\boldsymbol{\psi}) - \log p(\boldsymbol{\psi}|\mathbf{y}, \xi) \right] p(\boldsymbol{\psi}, \mathbf{y}|\xi) \, d\boldsymbol{\psi} \, d\mathbf{y} \,.$$

We refer to a design minimising this expected loss as NSIG-optimal.

However, until very recently, optimal Bayesian design has not evolved far from the methods reviewed in the landmark paper of Chaloner and Verdinelli (1995). Development and application of methods for Bayesian design have lagged behind the progress made in inference and modelling due to the additional complexity introduced by the need to integrate over the (as yet) unobserved responses, in addition to unknown model parameters. Hence, methodology has been restricted to simple models and fully sequential, one-point-at-a-time, procedures (Ryan et al., 2015a).

In this short paper, we present two examples of practical importance where Bayesian design methods are a natural choice, overview commonly used approaches to Bayesian design, and present and demonstrate methodology for high-dimensional Bayesian design using recently developed methods of approximate coordinate exchange (Overstall and Woods, 2016). We finish with a short discussion, highlighting issues that can prevent the uptake of Bayesian design in practice, and propose some potential remedies.

2 Examples

We discuss two classes of example, both of which involve estimation of a nonlinear model. These examples are typical of the situations where Bayesian design may (arguably) be most effective.

2.1 Pharmacokinetic studies

Consider a compartmental model, as might be used in pharmacokinetic studies of the movement of a drug through the body, described by

$$y(t) \sim N\left(c(\boldsymbol{\theta}) \mu(\boldsymbol{\theta}; t), \sigma^2 \nu(\boldsymbol{\theta}; t) \right) \,, \quad t \in [0, 24] \, \text{hours} \,, \tag{2}$$
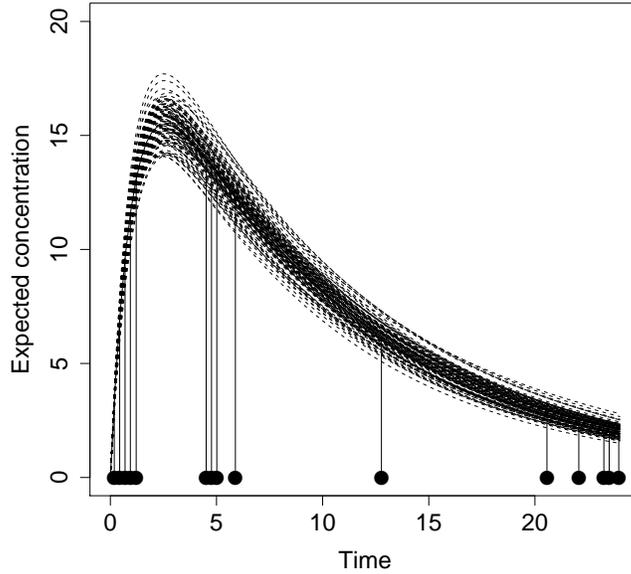
Figure 1: Realisations from compartmental model (2) with NSIG-optimal sampling times for $n = 15$ points subject to the constraint $|t_u - t_v| \geq 0.25$ $(u, v = 1, \ldots, n; u \neq v)$.

with

$$\mu(\boldsymbol{\theta}; t) = \exp(-\theta_1 t) - \exp(-\theta_2 t), \quad c(\boldsymbol{\theta}) = \frac{D}{\theta_3} \frac{\theta_2}{\theta_2 - \theta_1}, \quad \nu(\boldsymbol{\theta}; t) = 1 + \frac{\tau^2}{\sigma^2} c(\boldsymbol{\theta})^2 \mu(\boldsymbol{\theta}; t).$$

Following Ryan et al. (2014), we fix $D = 400$, $\tau^2 = 0.1$ and $\sigma^2 = 0.1$. Interest is in estimation of $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)^{\mathrm{T}}$, and normal prior distributions are assumed for $\log \boldsymbol{\theta}$, with $\mathrm{E}(\log \theta_1) = \log 0.1$, $\mathrm{E}(\log \theta_2) = 0$, $\mathrm{E}(\log \theta_3) = \log 20$ and $\mathrm{Var}(\log \theta_k) = 0.05$ for $k = 1, 2, 3$. A design is a choice of sampling times $t_1, \ldots, t_n$, here subject to the constraint $|t_u - t_v| \geq 0.25$ $(u, v = 1, \ldots, n; u \neq v)$. Figure 1 shows realisations from this model, obtained by sampling parameter values from the prior distributions for $\boldsymbol{\theta}$ and σ^2, and then sampling responses from (2). Also shown are the sampling times for an $n = 15$ NSIG-optimal design for estimating $\boldsymbol{\theta}$.

2.2 Experiments with discrete data

Generalised linear models (GLMs) are an important class of models for experiments in science and industry that measure responses that cannot be well described by a Normal linear model. Perhaps the most common cases involve discrete responses, for example, binary, binomial or count data. Woods et al. (2006) and Woods and van de Ven (2011) described examples from chemistry, food technology and engineering with binary (success/fail) responses. In particular, the crystallography experiment from Woods et al. (2006) involved measuring the formation, or not, of a pharmaceutical salt where a treatment was applied consisting of settings of four variables. A suitable GLM here might be a logistic regression. Let $y_i \sim \mathrm{Bernoulli}(\rho_i)$ be the response from the ith run of the experiment, with

$$\log \left(\frac{\rho_i}{1 - \rho_i} \right) = \beta_0 + \sum_{j=1}^{4} \beta_j x_{ji}, \tag{3}$$
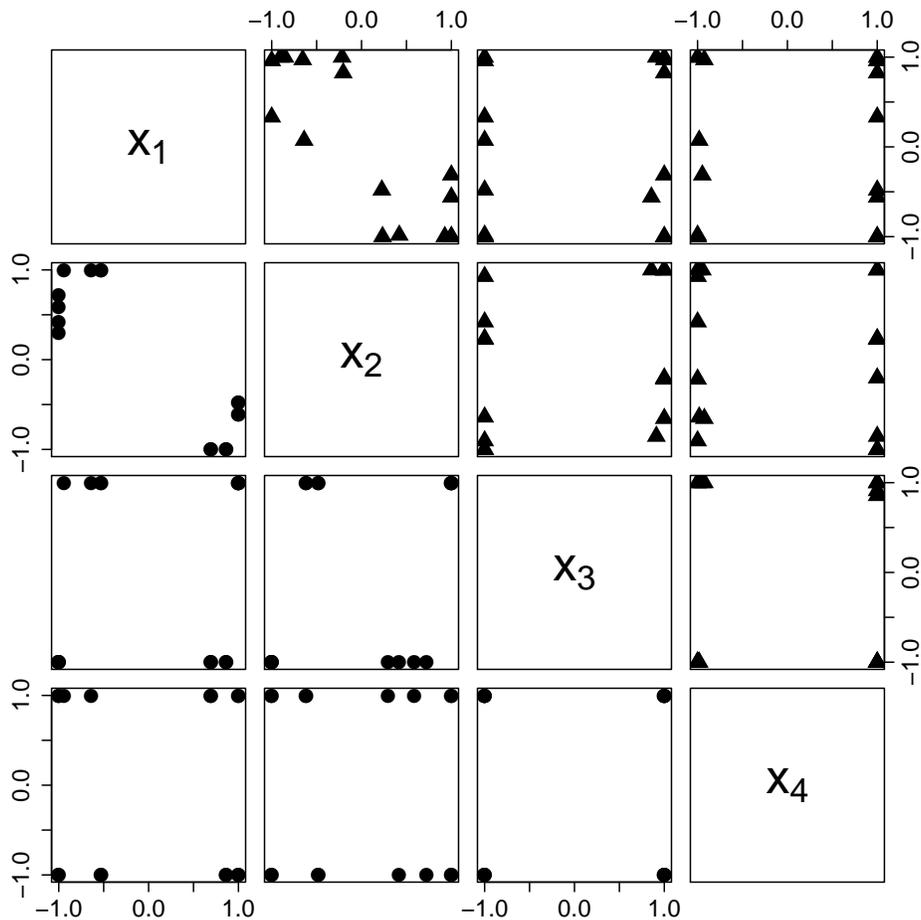
Figure 2: Two-dimensional projections of optimal designs with $n = 16$ points for model (3): NSIG-optimal design (lower panels, ●); pseudo-Bayesian D-optimal design (upper panels, ▲).

where $\beta_0, \ldots, \beta_4$ are unknown parameters to be estimated. For illustration, Woods et al. (2006) assumed the following prior distributions for the parameters:

$$\beta_0 \sim \mathrm{U}[-3, 3], \quad \beta_1 \sim \mathrm{U}[4, 10], \quad \beta_2 \sim \mathrm{U}[5, 11], \quad \beta_3 \sim \mathrm{U}[-6, 0], \quad \beta_4 \sim \mathrm{U}[-2.5, 3.5].$$

Figure 2 gives two-dimensional projections of the Bayesian NSIG-optimal design (lower panels) and a pseudo-Bayesian D-optimal design (upper panels; see Section 3.1), both with $n = 16$ points. We discuss how these designs were found and their relative performance in Section 4.3.

3 Approaches to Bayesian design

3.1 Asymptotic approximations

For experiments with large n, the inverse expected Fisher information matrix $M(\psi; \xi)$, typically evaluated at the posterior mode, is an asymptotic approximation to the posterior variance-covariance matrix of the

parameters $\boldsymbol{\psi}$. Use of this approximation leads to pseudo-Bayesian "alphabetic" optimality criteria. For example, under D-optimality, a design is selected to minimise

$$\Phi_D(\xi) = \int_\Psi -\log |M(\boldsymbol{\psi}; \xi)| \, \pi(\boldsymbol{\psi}) \, d\boldsymbol{\psi} \,. \tag{4}$$

The integral with respect to $\boldsymbol{\psi}$ is usually of low dimension and amenable to deterministic approximation. Such an approximation to the objective function can then be minimised using a conditional algorithm such as point or coordinate exchange; see, for, example, Gotwalt et al. (2009).

3.2 Simulation-based optimisation

In general, the expected loss can be approximated via Monte Carlo integration as

$$\tilde{\Phi}(\xi) = \frac{1}{B} \sum_{k=1}^B l\left(\xi, \mathbf{y}_k, \boldsymbol{\psi}_k\right) \,,$$

with $(\boldsymbol{\psi}_k, \mathbf{y}_k) \sim p(\boldsymbol{\psi}, \mathbf{y}|\xi)$, and the loss $l(\xi, \mathbf{y}_k, \boldsymbol{\psi}_k)$ often also requiring approximation (necessitating nested Monte Carlo simulation). Direct optimisation of this approximation requires large B to generate a suitably smooth objective function and/or expensive stochastic algorithms (e.g. simulated annealing or genetic algorithms), see for example, Hamada et al. (2001). Alternatively, the optimisation can be embedded within a Markov chain simulation scheme, and a modal design identified by sampling from the artificial joint distribution for the design, the model parameters and the data (Müller, 1999, Müller et al., 2004 and Amzal et al., 2006). Typically, an annealing step is employed to enable easier identification of the modal design. This approach is most effective for small experiments (both variables and runs). Recent extensions to this algorithm have allowed designs to be found for (i) models with intractable likelihoods using Approximate Bayesian Computation (Drovandi and Pettitt, 2013) and (ii) dynamic models with numerous sampling times using dimension-reduction (Ryan et al., 2014), importance sampling and Laplace approximations (Ryan et al., 2015b).

3.3 Sequential design

Most experiments are part of a sequence, where a Bayesian approach, with sequential updating from prior to posterior distributions, is natural. For point-sequential designs, approximation of the expected loss is greatly simplified by the reduction in the dimension of the integral. Recent methods have been suggested for estimation of, and discrimination between, nonlinear models, see Drovandi et al. (2013, 2014). A growing area is Bayesian optimisation of expensive black-box functions (e.g. in computer experiments), using Gaussian process surrogates to reduce the number of required function evaluations, following the seminal work of Jones et al. (1998). The computational efficiency of sequential design can be greatly aided through the use of sequential Monte Carlo for the necessary inference (Gramacy and Polson, 2011).

3.4 Smoothing-based optimisation

Smoothing-based design methods (Müller and Parmigiani, 1996) evaluate a computationally expensive, typically Monte Carlo, approximation to the expected loss (1) for a limited number of designs and then smooth these approximated losses to generate a surrogate function which can then be optimised in place of the expected loss. Recent research includes (i) extension of such methods to multi-variable experiments (Overstall and Woods, 2016) and (ii) use of surrogates to enable Bayesian D-optimal design for generalised linear mixed models (Waite and Woods, 2015).

We demonstrate the Müller and Parmigiani (1996) smoothing method by finding a simple design for compartmental model (2). We fix the first point in a design with $n = 2$ at $t_1 = 5$, and approximate the NSIG for 10 designs with different values of the second point t_2. Figure 3 shows the smoothed NSIG constructed
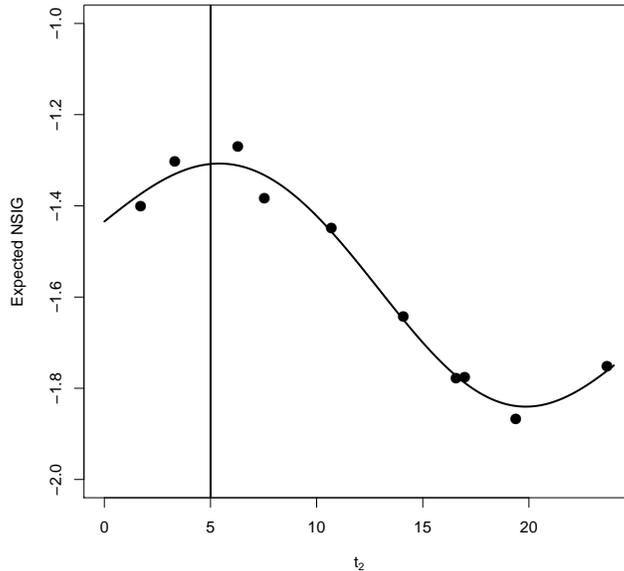
Figure 3: Compartmental example, $n = 2$ with the first design point fixed at time $t_1 = 5$ (vertical line) and smoothed approximate NSIG for 10 values of the second design point t_2.

from these 10 designs, as a function of t_2. Such a smoother may be used to find an optimal value of t_2, which is about $t_2 = 17$ in this example.

An extension to higher-dimensional examples, using conditional smoothing and optimisation, is described in the next section.

4 Bayesian optimal design via approximate coordinate exchange

In this section we describe and demonstrate multi-variable (factorial) Bayesian design using approximate coordinate exchange (ACE), as proposed by Overstall and Woods (2016).

4.1 The approximate coordinate exchange algorithm

ACE is a conditional optimisation algorithm that makes use of surrogates, or emulators, for the expected loss as a function of a single design coordinate (a value of a single variable in a single run). The algorithm steps through each coordinate of the design, and constructs a one-dimensional emulator, $\hat{\Phi}(x)$, for a Monte Carlo approximation, $\tilde{\Phi}(x_{ij}|\xi)$, to the expected loss for design ξ with ijth coordinate replaced by x_{ij} ($i = 1, \ldots, n$; $j = 1, \ldots, q$). We find the value of the coordinate that minimises this emulator, and perform an accept/reject step in order to decide whether to swap the current design coordinate with this minimum.

Algorithm 1 gives the basic steps of ACE, and the accept/reject step is described in Algorithm 2. This algorithm would typically be repeated multiple times (perhaps exploiting parallel computing) to avoid local optima. Overstall and Woods (2016) gave more details of the implementation and application of the algorithm, including its combination with a point-exchange algorithm to consolidate clusters of similar design points.

We employ a Gaussian process model (GP; see, for example, Rasmussen and Williams, 2006) in line 9 of

Input : Initial (randomly chosen) design ξ
Output: Φ-optimal design

1 begin
2 | **repeat**
3 | | **for** $i = 1 : n$ **do**
4 | | | **for** $j = 1 : q$ **do**
5 | | | | Generate a 1d space-filling design $\zeta_{ij} = \left\{ x_{ij}^1, \ldots, x_{ij}^Q \right\}$ in $\mathcal{X}_j \subset \mathbb{R}$;
6 | | | | **for** $k = 1 : Q$ **do**
7 | | | | | Evaluate $\tilde{\Phi}_k = \tilde{\Phi}(x_{ij}^k | \xi)$;
8 | | | | **end**
9 | | | | Construct a 1d emulator $\hat{\Phi}(x)$ as (5);
10 | | | | Set $x_{ij} = \operatorname{argmin}_{\mathcal{X}} \hat{\Phi}(x)$ with probability $p^\star$ obtained from Algorithm 2;
11 | | | **end**
12 | | **end**
13 | **until** convergence;
14 end

<div align="center">

Algorithm 1: The approximate coordinate exchange (ACE) algorithm.

</div>

Input : Current design ξ and proposed new coordinate x_{ij}
Output: Posterior probability $p^\star$ that $\tilde{\Phi}(x_{ij} | \xi) < \tilde{\Phi}(\xi)$

1 begin
2 | Let ξ_p be the design formed by replacing the ijth coordinate of ξ with x_{ij};
3 | **for** $k = 1 : B$ **do**
4 | | Sample $\tilde{\psi}$ from $\pi(\psi)$;
5 | | Sample $\mathbf{y}_1 \sim \pi(\mathbf{y} | \psi, \xi_p)$ and $\mathbf{y}_2 \sim \pi(\mathbf{y} | \psi, \xi)$;
6 | | Set $L_{1k} = l(\xi_p, \mathbf{y}_1, \tilde{\psi})$ and $L_{2k} = L_2 + l(\xi, \mathbf{y}_2, \tilde{\psi})$;
7 | **end**
8 | Assume $L_{1k} \sim N(b_1 + b_2, a)$ and $L_{2k} \sim N(b_1, a)$;
9 | Calculate the posterior probability, $p^\star$, that $b_2 < 0$ using "data" L_{1k} and L_{2k};
10 end

<div align="center">

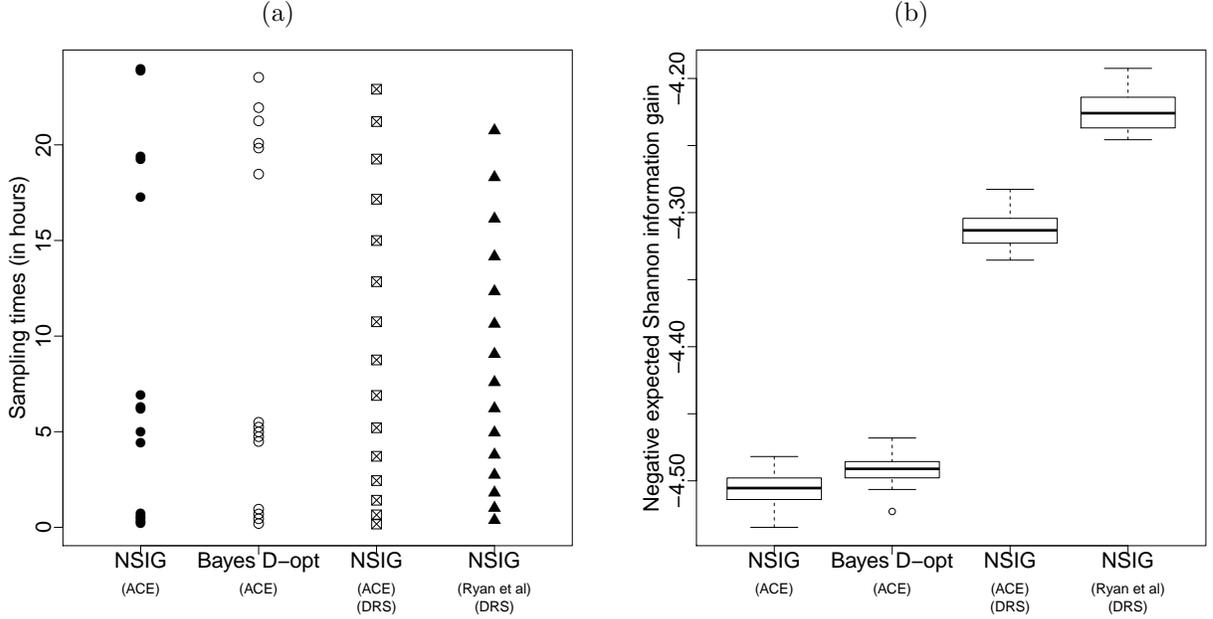**Algorithm 2:** Accept/reject step from line 10 of the ACE algorithm.

</div>

Figure 4: Bayesian optimal design for compartmental model (2): (a) Sampling times from the unrestricted NSIG-optimal, pseudo-Bayesian D-optimal and Beta DRS NSIG-optimal designs (all from ACE), together with the Ryan et al. (2014) Beta DRS NSIG-optimal design; (b) Boxplots for 20 evaluations of Monte Carlo approximations to NSIG for the four designs.

Algorithm 1. The emulator is given by the posterior mean function of the GP

$$\hat{\Phi}(x) = \hat{\mu}_{ij} + \hat{\sigma}_{ij}\boldsymbol{a}^{\mathrm{T}}(x, \zeta_{ij})A(\zeta_{ij})\boldsymbol{z}_{ij}, \tag{5}$$

with $\hat{\mu}_{ij} = \sum_{k=1}^{Q} \tilde{\Phi}(x_{ij}^k|\xi)/B$, $\hat{\sigma}_{ij}^2 = \sum_{k=1}^{Q} \left(\tilde{\Phi}(x_{ij}^k|\xi) - \hat{\mu}_{ij}\right)^2/(B-1)$, $\boldsymbol{z}_{ij}$ a Q-vector having kth entry $\left(\tilde{\Phi}(x_{ij}^k|\xi) - \hat{\mu}_{ij}\right)/\hat{\sigma}_{ij}$ and x_{ij}^k being points from a one-dimensional space-filling design (see Algorithm 1). Under the common assumption of a squared exponential correlation structure, the Q-vector $\boldsymbol{a}$ and $Q \times Q$ matrix A have entries

$$\boldsymbol{a}(x, \zeta_{ij})_u = \exp\left\{-\rho(x - x_{ij}^u)^2\right\}, \quad A(\xi)_{uv} = \exp\left\{-\rho(x_{ij}^u - x_{ij}^v)^2\right\} + \eta\,\mathrm{I}(u = v), \quad u, v = 1, \dots, Q,$$

with I the indicator function. The inclusion of a nugget η ensures the emulator will smooth, rather than interpolate, the $\tilde{\Phi}$ values. We estimate ρ and η via maximum likelihood.

The minimisation in line 10 is subject to both Monte Carlo error and emulator error. To remove the emulator error when making the decision to accept the exchange, we perform the steps in Algorithm 2 and use independent Monte Carlo samples to assess the improvement in the design. Algorithm 2 essentially describes a Bayesian t-test based on simulated data from the existing and proposed designs (c.f. Wang and Zhang, 2006). If the assumption of normality that underpins this test is invalid, a nonparametric procedure may be used instead.

4.2 Example: pharmacokinetic design

Overstall and Woods (2016) used ACE to find designs with $n = 15$ points for compartmental model (2) to compare to the design found by Ryan et al. (2014). We reproduce those results here, to demonstrate the effectiveness of ACE.
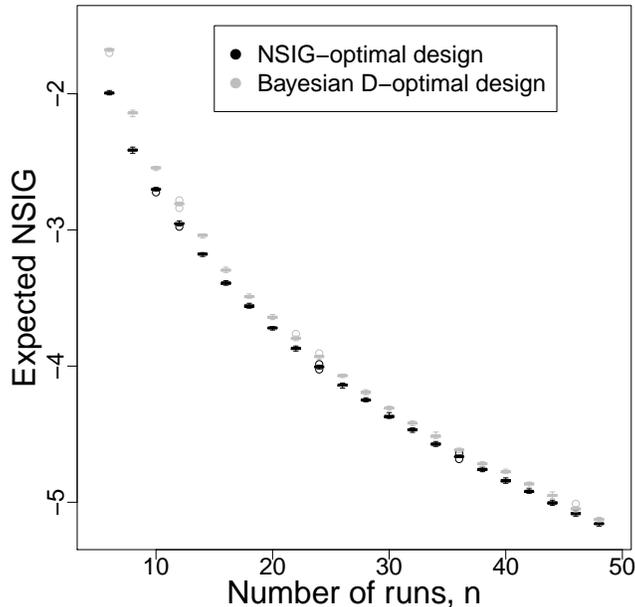
Figure 5: Boxplots of 20 evaluations of the NISG expected loss for NSIG-optimal and pseudo-Bayesian D-optimal designs for $n = 6, \ldots, 48$.

The design of Ryan et al. (2014) used a dimension reduction scheme (DRS) that restricted the search for NSIG-optimal designs to sampling times formed as scaled quantiles from a $\text{Beta}(\alpha_1, \alpha_2)$ distribution. Hence, the optimisation problem is reduced from finding $n = 15$ design points to finding values for the parameters α_1 and α_2. Ryan et al. (2014) performed this two dimensional optimisation using the simulation algorithm of Müller et al. (2004). These authors also imposed the constraint of a 15 minute interval between adjacent sampling times, hence ruling out any replicate runs in the design.

Figure 4 gives Bayesian optimal designs ($n = 15$) from four different strategies and evaluates their performance in terms of NSIG. Three strategies use ACE to find an optimal design: (i) NSIG-optimal design; (ii) pseudo-Bayesian D-optimal design, minimising (4); and (iii) NISG-optimal design found using the Beta DRS. The final design comes from the Ryan et al. (2014) strategy of using a Beta DRS with the Müller et al. (2004) simulation algorithm. Figure 4(a) shows that the DRS designs do not display the same clustering of design points as the two unrestricted designs. In Figure 4(b), we see that the DRS designs also have higher expected loss, up to 5% higher, than the unrestricted designs. The Bayesian D-optimal design is a reasonable surrogate for the NSIG-optimal design, and ACE is able to find a DRS design with lower expected loss than the simulation-based approach.

4.3   Example: binary data

The NISG-optimal and pseudo-Bayesian D-optimal designs in Figure 2 were both found using the ACE algorithm, followed by application of a point exchange algorithm to consolidate close clusters of design points. Both designs are quite different from a factorial design, having multiple values of x_1 and, especially, x_2 (the two variables that have parameters with prior support not including 0). The pseudo-Bayesian D-optimal design also includes some points in the interior of the design region. We compare the performance of these two designs, as well as designs for other values of n, by computing 20 independent Monte Carlo approximations to the NSIG, see Figure 5.

From these results, we can see that the difference in NISG between the two types of design decreases as n increases, with only substantial differences for small numbers of runs. This is not surprising, as the posterior distribution for the model parameters will be asymptotically well approximated by a normal distribution with variance-covariance matrix given by the inverse of the Fisher information (see, for example, Gelman et al., 2014, pp. 585-588). However, for small n, then can be substantial average differences in NSIG between the designs, of up to 20%. For $n = 16$, the average difference is just less than 2%. Without the use of an efficient computational algorithm, such as ACE, a comparative study to investigate the small-sample performance of pseudo-Bayesian D-optimal designs would not be possible.

5 Discussion

Optimal Bayesian design is challenging for high-dimensional problems with multi-variable models and/or many design points and there are few literature examples of such designs being used in practice. Reasons for this include the lack of scaleable algorithms for design selection, the complexity of available software for Bayesian design, an unwillingness in some areas to "bias" designs through the use of prior information and, in many application areas, a lack of appreciation that "DoE" can go beyond standard factorial designs. However, Bayesian design is a powerful tool for a variety of experiments. Here, we have focussed on using new computational methodology to find designs for nonlinear models, where some prior information is *necessary* to design informative experiments. Another important application area is in screening experiments, and other small n, big p ("fat data") problems. In fact, prior information is used implicitly whenever a fractional factorial experiment is employed (for example, by invoking the principle of effect hierarchy).

Methodology such as ACE removes some of the barriers to the implementation of Bayesian design, both by widening the scope of models and experiments that can be addressed, and by facilitating the provision of greater evidence for the effectiveness of the methods through rigorous scientific studies. More details of the methodology demonstrated in this paper can be found in Overstall and Woods (2016) and also in Overstall et al. (2015) who discussed optimal designs for uncertainty quantification of physical models, an application area of increasing importance. The ACE algorithm has been implemented in an R package, available on request.

Acknowledgements

References

Amzal, B., Bois, F. Y., Parent, E., and Robert, C. (2006), "Bayesian optimal design via interacting particle systems," *Journal of the American Statistical Association*, 101, 773–785.

Atkinson, A., Donev, A., and Tobias, R. (2007), *Optimum Experimental Design, with SAS*, Oxford: Oxford University Press.

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

Drovandi, C. C., McGree, J. M., and Pettitt, A. N. (2013), "Sequential Monte Carlo for Bayesian sequentially designed experiments for discrete data," *CSDA*, 57, 320–335.

— (2014), "A sequential Monte Carlo algorithm to incorporate model uncertainty in Bayesian sequential design," *JCGS*, 23, 3–24.

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

Gelman, A., Carlin, J., Stern, H., Dunson, D., Vehtari, A., and Rubin, D. (2014), *Bayesian Data Analysis*, Boca Raton: CRC, 3rd ed.

Gotwalt, C. M., Jones, B. A., and Steinberg, D. M. (2009), "Fast computation of designs robust to parameter uncertainty for nonlinear settings," *Technometrics*, 51, 88–95.

Gramacy, R. B. and Polson, N. G. (2011), "Particle learning of Gaussian process models for sequential design and optimization," *Journal of Computational and Graphical Statistics*, 20, 102–118.

Hamada, M., Martz, H. F., Reese, C. S., and Wilson, A. G. (2001), "Finding near-optimal Bayesian experimental designs via genetic algorithms," *The American Statistician*, 55, 175–181.

Jones, D., Schonlau, M., and Welch, W. (1998), "Efficient global optimization of expensive black-box functions," *Journal of Global Optimization*, 13, 455–492.

Müller, P. (1999), "Simulation-based optimal design," in *Bayesian Statistics 6*, eds. Bernardo, J. M., Bayarri, M. J., Berger, J. O., Dawid, A. P., Heckerman, D., and Smith, A. F. M., Oxford.

Müller, P. and Parmigiani, G. (1996), "Optimal design via curve fitting of Monte Carlo experiments," *Journal of the American Statistical Association*, 90, 1322–1330.

Müller, P., Sanso, B., and De Iorio, M. (2004), "Optimal Bayesian design by inhomogeneous Markov chain simulationMarkov Chain Simulation," *JASA*, 99, 788–798.

Overstall, A. M. and Woods, D. C. (2016), "Bayesian design of experiments using approximate coordinate exchange," *arXiv:1501.00264*.

Overstall, A. M., Woods, D. C., and Parker, B. M. (2015), "Bayesian optimal design for ordinary differential equation models," *arXiv:1509.04099*.

Rasmussen, C. E. and Williams, C. K. I. (2006), *Gaussian Processes for Machine Learning*, Cambridge, MA.: MIT Press.

Ryan, E. G., Drovandi, C. C., McGree, J. M., and Pettitt, A. N. (2015a), "A review of modern computational algorithms for Bayesian optimal design," *International Statistical Review*, in press.

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

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.

Waite, T. W. and Woods, D. C. (2015), "Designs for generalized linear models with random block effects via information matrix approximations," *Biometrika*, 102, 677–693.

Wang, L. and Zhang, L. (2006), "Stochastic optimization using simulated annealing with hypothesis test," *Applied Mathematics and Computation*, 174, 1329–1342.

Woods, D. C., Lewis, S., Eccleston, J. A., and Russell, K. G. (2006), "Designs for generalised linear models with several variables and model uncertainty," *Technometrics*, 48, 284–292.

Woods, D. C. and van de Ven, P. (2011), "Blocked designs for experiments with non-normal response," *Technometrics*, 53, 173–182.