Computational Statistics and Data Analysis 70 (2014) 45–60

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Towards Bayesian experimental design for nonlinear models that require a large number of sampling times

Elizabeth G. Ryan *, Christopher C. Drovandi, M. Helen Thompson, Anthony N. Pettitt

Mathematical Sciences, Queensland University of Technology, Brisbane, 4001, Australia

ARTICLE INFO

Article history:
Received 18 January 2013
Received in revised form 28 August 2013
Accepted 28 August 2013
Available online 11 September 2013

Keywords:
Bayesian optimal design
Sampling strategies
Robust design
Markov chain Monte Carlo
Stochastic optimisation

ABSTRACT

The use of Bayesian methodologies for solving optimal experimental design problems has increased. Many of these methods have been found to be computationally intensive for design problems that require a large number of design points. A simulation-based approach that can be used to solve optimal design problems in which one is interested in finding a large number of (near) optimal design points for a small number of design variables is presented. The approach involves the use of lower dimensional parameterisations that consist of a few design variables, which generate multiple design points. Using this approach, one simply has to search over a few design variables, rather than searching over a large number of optimal design points, thus providing substantial computational savings. The methodologies are demonstrated on four applications, including the selection of sampling times for pharmacokinetic and heat transfer studies, and involve nonlinear models. Several Bayesian design criteria are also compared and contrasted, as well as several different lower dimensional parameterisation schemes for generating the many design points.

1. Introduction

Optimal experimental design provides rules for the allocation of resources for studies which require data collection but where there is variability present, whether it is under the full control of the experimenter or not. Experimental designs are concerned with the incorporation of features into studies to control systematic error (bias), reduce random variations, and increase precision. Experimental design problems are commonly viewed as optimisation problems. Optimal experimental designs may be used to achieve the experimental goals more rapidly and hence reduce experimental costs.

Experimental design has been widely developed within the classical framework, in both theory and practice (e.g., Atkinson and Donev, 1992). In the classical framework, optimal experimental designs are commonly derived using optimality criteria that are based on the Fisher information matrix (e.g., Fedorov, 1972; Pukelsheim and Torsney, 1991; Atkinson and Donev, 1992).

Classical experimental design is well suited to linear or linearised models. However, for nonlinear models, designs are locally dependent on the values which are chosen for the model parameters. Since interest is often focused on the design of experiments that can provide accurate parameter estimates, this means that selection of the parameters from which to construct the design is of critical importance and the use of unsuitable parameter values may result in suboptimal designs.

* Correspondence to: Mathematical Sciences, Queensland University of Technology, P.O. Box 2434, Brisbane, Queensland 4001, Australia. Tel.: +61 0 7 3138 2308.

E-mail addresses: elizabeth.ryan@student.qut.edu.au, eg.ryan@qut.edu.au (E.G. Ryan), c.drovandi@qut.edu.au (C.C. Drovandi), helen.thompson@qut.edu.au (M.H. Thompson), a.pettitt@qut.edu.au (A.N. Pettitt).

0167-9473/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.csda.2013.08.017

In the classical framework, for nonlinear models, only locally optimal designs can be obtained. To overcome the dependence of the design on the initial estimates of the parameters, several studies have incorporated probability distributions on the parameters and average local design criteria over the distribution so that the designs obtained may be robust to the initial choice of the parameter values (e.g., D'Argenio, 1990; Duffull et al., 2005; Ogungbenro and Aarons, 2007; Duffull et al., 2012). However, these so-called "pseudo-Bayesian" design criteria are more computationally intensive than the classical design criteria.

Here we consider a Bayesian approach. The use of Bayesian methodologies for optimal experimental design has increased in the last few years (e.g., Müller, 1999; Stroud et al., 2001; Amzal et al., 2006; Müller et al., 2006; Cook et al., 2008). When constructing Bayesian optimal designs, the maximum expected utility principle is employed (see de Groot, 1970), in which the preferences of the decision maker are assumed to be encoded by a utility function, $U(\mathbf{d}, \boldsymbol{\theta}, \mathbf{y})$. The utility function describes the worth of choosing the design $\mathbf{d}$ from the design space $\mathbf{D}$, yielding data $\mathbf{y}$ from a sample space $\mathbf{Y}$, given model parameters (and latent variables) $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. The form of the utility function is specific to the application and should incorporate the experimental aims (see Lindley, 1972; Chaloner and Verdinelli, 1995). A probabilistic model, $p(\boldsymbol{\theta}, \mathbf{y}|\mathbf{d})$, is also required for all relevant random variables and future data. The probability model is decomposed into a prior distribution $p(\boldsymbol{\theta})$ and a sampling distribution $p(\mathbf{y}|\mathbf{d}, \boldsymbol{\theta})$.

Lindley (1972) suggests that the choice of a design should be regarded as a decision problem and that the design $\mathbf{d}$ which maximises the expected utility should be selected. Historical data may be incorporated into the model by conditioning the decision process on the available information. The Bayesian framework seeks to determine the optimal design, $\mathbf{d}^*$, that maximises the expected utility function $U(\mathbf{d})$ over the design space $\mathbf{D}$ with respect to the unknown future observations $\mathbf{y}$ and model parameters $\boldsymbol{\theta}$:

$$\begin{aligned}\mathbf{d}^* &= \arg\max_{\mathbf{d}\in\mathbf{D}} E[U(\mathbf{d}, \boldsymbol{\theta}, \mathbf{y})] \\ &= \arg\max_{\mathbf{d}\in\mathbf{D}} \int_{\mathbf{Y}} \int_{\boldsymbol{\Theta}} U(\mathbf{d}, \boldsymbol{\theta}, \mathbf{y}) p(\boldsymbol{\theta}, \mathbf{y}|\mathbf{d}) d\boldsymbol{\theta} d\mathbf{y}. \end{aligned} \tag{1}$$

In this article we assume that $\mathbf{d}^*$ exists.

There are several difficulties associated with maximising the expected utility surfaces. First, the decision space is often quite intricate in practical problems. Second, Eq. (1) does not usually have a closed form solution and requires multiple integrations and a maximisation (often) over a large decision space to obtain the optimum, $\mathbf{d}^*$. Even if the decision space is small and easily parameterised, the utility function may be difficult to integrate. Unless the likelihood and prior are specifically chosen to enable analytic evaluation of the integration problem, the maximisation and integration problem requires numerical approximation or stochastic solution methods. In this article we will focus on stochastic solution methods.

A range of stochastic algorithms have been proposed in the literature to approximate the maximisation and integration problem. These include: prior simulation (Müller, 1999); smoothing of Monte Carlo simulations (Müller, 1999); gridding methods which involve numerical quadrature or Laplace approximations (Brockwell and Kadane, 2003); Markov chain Monte Carlo (MCMC) simulation in an augmented probability model (Müller, 1999); and sequential Monte Carlo methods (Amzal et al., 2006). Most of these simulation methods are based on the assumption that the integral(s) (Eq. (1)) may be evaluated by Monte Carlo simulations with relative ease. In the majority of situations, $p(\boldsymbol{\theta}, \mathbf{y}|\mathbf{d})$ is available for efficient random variable generation and the utility function can be evaluated point-wise using the simulated $(\boldsymbol{\theta}_i, \mathbf{y}_i)$ for $i = 1, \ldots, M$. The integral may then be approximated by using:

$$\hat{U}(\mathbf{d}) = \frac{1}{M} \sum_{i=1}^{M} U(\mathbf{d}, \boldsymbol{\theta}_i, \mathbf{y}_i). \tag{2}$$

One can then use $\hat{U}(\mathbf{d})$ to find the optimal design, $\mathbf{d}^* = \arg\max \hat{U}(\mathbf{d})$, by using a suitable maximisation method (see Müller, 1999). However, straightforward Monte Carlo integration over $(\boldsymbol{\theta}, \mathbf{y})$ for each design $\mathbf{d}$ can be computationally intensive for high dimensional design problems since a large value of M is required to obtain reasonable accuracy of the estimate of $U(\mathbf{d})$.

Clyde et al. (1996), Bielza et al. (1999) and Müller (1999) instead treated the expected utility described in Eq. (1) as an unnormalised marginal probability density function. This was achieved by placing a joint distribution on $(\mathbf{d}, \boldsymbol{\theta}, \mathbf{y})$ to form an augmented probability model $h(\mathbf{d}, \boldsymbol{\theta}, \mathbf{y})$, which is given by:

$$h(\mathbf{d}, \boldsymbol{\theta}, \mathbf{y}) \propto U(\mathbf{d}, \boldsymbol{\theta}, \mathbf{y}) p(\boldsymbol{\theta}) p(\mathbf{y}|\mathbf{d}, \boldsymbol{\theta}),$$

assuming that $U(\mathbf{d}, \boldsymbol{\theta}, \mathbf{y})$ satisfies the appropriate conditions for $h(\cdot)$ to be positive and integrable over $(\mathbf{D}, \boldsymbol{\Theta}, \mathbf{Y})$. The resulting probability distribution $h(\cdot)$ is defined such that the marginal distribution of $\mathbf{d}$ is proportional to the expected utility, i.e.,

$$h(\mathbf{d}) \propto \iint U(\mathbf{d}, \boldsymbol{\theta}, \mathbf{y}) p(\boldsymbol{\theta}, \mathbf{y}|\mathbf{d}) d\boldsymbol{\theta} d\mathbf{y} = U(\mathbf{d}).$$

Usually it is assumed that the design space $\mathbf{D}$ is bounded and that the utility $U(\mathbf{d}, \boldsymbol{\theta}, \mathbf{y})$ is non-negative and bounded. One can then simulate from $h(\cdot)$ using, say, a Metropolis–Hastings MCMC scheme to solve the optimal design problem by selecting