



Sequential Monte Carlo for Bayesian sequentially designed experiments for discrete data

Christopher C. Drovandi*, James M. McGree, Anthony N. Pettitt

Mathematical Sciences, Queensland University of Technology, P.O. Box 2434, Brisbane, Queensland 4001, Australia

ARTICLE INFO

Article history:

Received 10 January 2012

Received in revised form 16 May 2012

Accepted 16 May 2012

Available online 30 May 2012

Keywords:

Clinical trials

Generalised linear model

Generalised non-linear model

Sequential design

Sequential Monte Carlo

Target stimulus

ABSTRACT

In this paper we present a sequential Monte Carlo algorithm for Bayesian sequential experimental design applied to generalised non-linear models for discrete data. The approach is computationally convenient in that the information of newly observed data can be incorporated through a simple re-weighting step. We also consider a flexible parametric model for the stimulus–response relationship together with a newly developed hybrid design utility that can produce more robust estimates of the target stimulus in the presence of substantial model and parameter uncertainty. The algorithm is applied to hypothetical clinical trial or bioassay scenarios. In the discussion, potential generalisations of the algorithm are suggested to possibly extend its applicability to a wide variety of scenarios.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

In sequential or adaptive design, one makes use of currently available data in order to make an informed decision about the selection of the next design point. This decision is based on optimising some utility function over a set of allowable design points, which takes into account that many different observables are plausible at each design point. In many instances the utility function involves computing the expected information gain about the parameters, or a function thereof, of a particular model that is assumed to generate the observed data. Sequential designs are useful as they are generally more efficient than static designs for non-linear models in the presence of parameter and model uncertainty (see, for example, [Dror and Steinberg, 2008](#)). Therefore such approaches are critical in applications where data are costly to collect.

The incorporation of parameter and model uncertainty is most rigorously achieved within a Bayesian framework. Consequently, many authors have adopted such an approach in optimal experimental design (see, for example, [Chaloner and Verdinelli, 1995](#); [Müller et al., 2007](#); [Amzal et al., 2006](#)). The sequential nature of how the data are observed lends itself naturally to the use of an on-line updating algorithm such as sequential Monte Carlo (SMC). To our knowledge there has been little application of this methodology to sequential design problems. [Gramacy and Polson \(2011\)](#) present a particle learning algorithm for sequential design, but have a specific focus on Gaussian process models. The aim of this article is to present an algorithm for sequential design problems that can be applied when there exists a finite set of design points and a discrete observation process. We believe that this approach has the potential to be generalised to other settings (see the discussion for more details).

The implementation of sequential design methodology is prevalent throughout the literature, for example, in computer experiments ([Loeppky et al., 2010](#)), neural networks ([Pavel and Miroslav, 2010](#)), item response theory ([Chang and Ying, 2009](#)), neurophysiology experiments ([Lewi et al., 2009](#)), clinical trials ([Liu et al., 2009](#)) and bioassays ([Tian and Wang, 2009](#)).

* Correspondence to: Mathematical Sciences, Queensland University of Technology, P.O. Box 2434, Brisbane, Queensland 4001, Australia.

E-mail addresses: c.drovandi@qut.edu.au (C.C. Drovandi), james.mcgree@qut.edu.au (J.M. McGree), a.pettitt@qut.edu.au (A.N. Pettitt).

In this paper we illustrate the algorithms and utility functions we develop on two motivating examples. We focus on utility functions where the objective is to learn about a particular value on the stimulus–response curve; however some of the utilities can be applied generally and alternative utilities can be incorporated straightforwardly.

In both motivating examples we use a generalised linear or non-linear model framework (see, for example, [Biedermann and Woods, 2011](#)) where the predictor is a function of the stimulus applied. The first example is binary; for example it could be whether or not a particular toxic event (or adverse reaction) occurs after dose administration in phase I clinical trials or an all-or-nothing response in quantal assay. Secondly, we consider a count response (for example, the number of seizures in epileptic studies such as in phase II clinical trials) modelled via a Poisson distribution (see [Russell et al., 2009](#)). We assume that each subject receives a single stimulus, and that each response is available before the next subject. In both cases we are interested in estimating the stimulus value based on a particular expected response.

Bayesian sequential design is suited to SMC methods as the sequence of targets is built through data annealing. SMC algorithms (also known as particle filters) sample from a sequence of target distributions that evolve smoothly (see [Del Moral et al., 2006](#)). Weighted samples (particles) are generated from a tractable importance distribution and are propagated through the sequence of targets via iteratively applying re-weighting (importance sampling), resampling and mutation steps. New data are incorporated via a simple re-weighting step rather than sampling the full posterior distribution via Markov chain Monte Carlo (MCMC) as in [McGree et al. \(2012\)](#) for example. Therefore the major motivation for an SMC approach is computational. We make use of an efficient Metropolis–Hastings algorithm within the SMC to diversify the particle population, called the mutation step, after resampling to avoid particle degeneration, although this does not occur after every re-weight. Moreover, this step is potentially parallelisable. The use of MCMC as a mutation kernel within SMC results in an algorithm complexity that is linear in the number of particles. Furthermore, we find that SMC provides a convenient way to estimate the Kullback–Leibler divergence ([Kullback and Leibler, 1951](#)), which is a general utility of interest in Bayesian design problems ([Bernardo and Smith, 1994](#)).

In the stimulus selection phase of the algorithm, a new posterior distribution must be approximated for each combination of new stimulus and response in order to evaluate the utility value for each design point. It is important to note, however, that these posteriors do not form part of our sequence of targets, and are solely used for the stimulus selection. In this part we use importance sampling (as per [McGree et al., 2012](#)) to obtain estimated utilities via the weighted samples. We use the algorithm to efficiently compare the performance of several fully-Bayesian utilities for adaptive designs where the interest is in estimating the target stimulus, which is a function of model parameters. We implement a new hybrid utility for this purpose that can be useful in the presence of little prior information.

In the parametric approach to adaptive studies an underlying model must be assumed for the observed data. In this article we investigate the impact of model mis-specification on the estimate of the target stimulus and propose a more general model that can lead to more robust estimates.

This paper is organised as follows. In Section 2 we detail the parametric models we consider as well as the priors we place on the parameters of such models. SMC is introduced in Section 3 together with the self-tuning SMC algorithm we develop for Bayesian sequential design. In Section 4 we present the various Bayesian utilities we consider to assist in the stimulus selection process. The results are given in Section 5 followed by a concluding discussion in Section 6. One part of the discussion involves suggestions for generalising the algorithm to continuous responses and non-finite design spaces.

2. Problem and Bayesian framework

2.1. Inferential problem

In the sequential experimental design set-up, the current data together with a utility function is used to make an informed decision about the values of the controllable variables to use for the next experimental unit. Using a Bayesian approach, the current information is encapsulated within the posterior distribution, $\pi_t(\theta|\mathbf{y}_{1:t}, \mathbf{D}_{1:t}) \propto \pi(\theta) \prod_{i=1}^t f(y_i|\theta, D_i)$, if the responses $\mathbf{y}_{1:t} = (y_1, \dots, y_t)$ are conditionally independent given $\mathbf{D}_{1:t}$. The value of the controllable variable for the t th subject is given by D_t . The prior, $\pi(\theta)$, contains the information about the model parameter, θ , before any experiments. The problem is then to determine an appropriate value for D_{t+1} (the so-called stimulus in this paper) to apply to the $t + 1$ th experimental unit in order to achieve some inferential goal as quickly as possible. The aim of the experiment is captured by the utility function which depends on the currently collected data, $U(d|\mathbf{y}_{1:t}, \mathbf{D}_{1:t})$. Maximising this utility over the design space, $\mathbf{D}_a$, we obtain

$$D_{t+1} = \arg \max_{d \in \mathbf{D}_a} U(d|\mathbf{y}_{1:t}, \mathbf{D}_{1:t}).$$

Such an approach is referred to as a one-at-a-time or myopic design. The utility value is given by the expectation of the user-specified utility function, $U(d, \theta, z|\mathbf{y}_{1:t}, \mathbf{D}_{1:t})$, over the posterior parameter space and response space. For discrete data we obtain

$$U(d|\mathbf{y}_{1:t}, \mathbf{D}_{1:t}) = \sum_{z \in \mathcal{S}} f(z|\mathbf{y}_{1:t}, \mathbf{D}_{1:t}, d) \int_{\theta} U(d, \theta, z|\mathbf{y}_{1:t}, \mathbf{D}_{1:t}) \pi(\theta|(\mathbf{y}_{1:t}, z), (\mathbf{D}_{1:t}, d)) d\theta,$$

Download English Version:

<https://daneshyari.com/en/article/10327505>

Download Persian Version:

<https://daneshyari.com/article/10327505>

[Daneshyari.com](https://daneshyari.com)