ARTICLE IN PRESS

Computational Statistics and Data Analysis xx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

Multilevel rejection sampling for approximate Bayesian computation

David J. Warne [a,*], Ruth E. Baker [b], Matthew J. Simpson [a]

[a] School of Mathematical Sciences, Queensland University of Technology, Brisbane, Australia
[b] Mathematical Institute, University of Oxford, Oxford, United Kingdom

ARTICLE INFO

Article history:
Received 1 June 2017
Received in revised form 28 February 2018
Accepted 28 February 2018
Available online xxxx

Keywords:
Bayesian inference
Approximate Bayesian computation
Multilevel Monte Carlo
Rejection sampling
Likelihood-free methods

ABSTRACT

Likelihood-free methods, such as approximate Bayesian computation, are powerful tools for practical inference problems with intractable likelihood functions. Markov chain Monte Carlo and sequential Monte Carlo variants of approximate Bayesian computation can be effective techniques for sampling posterior distributions in an approximate Bayesian computation setting. However, without careful consideration of convergence criteria and selection of proposal kernels, such methods can lead to very biased inference or computationally inefficient sampling. In contrast, rejection sampling for approximate Bayesian computation, despite being computationally intensive, results in independent, identically distributed samples from the approximated posterior. An alternative method is proposed for the acceleration of likelihood-free Bayesian inference that applies multilevel Monte Carlo variance reduction techniques directly to rejection sampling. The resulting method retains the accuracy advantages of rejection sampling while significantly improving the computational efficiency.

1. Introduction

Statistical inference is of fundamental importance to all areas of science. Inference enables the testing of theoretical models against observations, and provides a rational means of quantifying uncertainty in existing models. Modern approaches to statistical inference, based on Monte Carlo sampling techniques, provide insight into many complex phenomena (Beaumont et al., 2002; Pooley et al., 2015; Ross et al., 2017; Stumpf, 2014; Sunnåker et al., 2013; Tavaré et al., 1997; Thorne and Stumpf, 2012; Vo et al., 2015).

Suppose we have: a set of observations, $\mathcal{D}$; a method of determining the likelihood of these observations, $\mathcal{L}(\boldsymbol{\theta}; \mathcal{D})$, under the assumption of some model characterised by parameter vector $\boldsymbol{\theta} \in \boldsymbol{\Theta}$; and a prior probability density, $p(\boldsymbol{\theta})$. The posterior probability density, $p(\boldsymbol{\theta} \mid \mathcal{D})$, can be computed using Bayes' Theorem,

$$p(\boldsymbol{\theta} \mid \mathcal{D}) = \frac{\mathcal{L}(\boldsymbol{\theta}; \mathcal{D})p(\boldsymbol{\theta})}{\int_{\boldsymbol{\Theta}} \mathcal{L}(\boldsymbol{\theta}; \mathcal{D})p(\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta}}. \tag{1}$$

* Correspondence to: 2 George St, Brisbane, Queensland, 4000, Australia.
E-mail address: david.warne@qut.edu.au (D.J. Warne).

https://doi.org/10.1016/j.csda.2018.02.009
0167-9473/

Explicit expressions for likelihood functions are rarely available (Tavaré et al., 1997; Warne et al., 2017; Wilkinson, 2009); motivating the development of likelihood-free methods, such as approximate Bayesian computation (ABC) (Stumpf, 2014; Sunnåker et al., 2013). ABC methods approximate the likelihood through evaluating the discrepancy between data generated by a simulation of the model of interest and the observations, yielding an approximate posterior,

$$p(\boldsymbol{\theta} \mid d(\mathcal{D}, \mathcal{D}_s) < \epsilon) \propto \mathbb{P}(d(\mathcal{D}, \mathcal{D}_s) < \epsilon \mid \boldsymbol{\theta}) p(\boldsymbol{\theta}). \tag{2}$$

Here, $\mathcal{D}_s \sim f(\mathcal{D} \mid \boldsymbol{\theta})$ is data generated by the model simulation process, $f(\mathcal{D} \mid \boldsymbol{\theta})$, d is a discrepancy metric, and $\epsilon > 0$ is the acceptance threshold. Due to this approximation, Monte Carlo estimators based on Eq. (2) are biased (Barber et al., 2015). In spite of this bias, however, ABC methods have proven to be very powerful tools for practical inference applications in many scientific areas, including evolutionary biology (Beaumont et al., 2002; Tavaré et al., 1997; Thorne and Stumpf, 2012), ecology (Stumpf, 2014), cell biology (Ross et al., 2017; Johnston et al., 2014; Vo et al., 2015) and systems biology (Wilkinson, 2009).

1.1. Sampling algorithms for ABC

The most elementary implementation of ABC is ABC rejection sampling (Pritchard et al., 1999; Sunnåker et al., 2013), see Algorithm 1. This method generates N independent and identically distributed samples $\boldsymbol{\theta}^1, \ldots, \boldsymbol{\theta}^N$ from the posterior distribution by accepting proposals, $\boldsymbol{\theta}^* \sim p(\boldsymbol{\theta})$, when the data generated by the model simulation process $f(\mathcal{D} \mid \boldsymbol{\theta}^*)$ is within ϵ of the observed data, $\mathcal{D}$, under the discrepancy metric $d(\mathcal{D}, \cdot)$. ABC rejection sampling is simple to implement, and samples are independent and identically distributed. Therefore ABC rejection sampling is widely used in many applications (Browning et al., 2018; Grelaud et al., 2009; Navascués et al., 2017; Ross et al., 2017; Vo et al., 2015). However, ABC rejection sampling can be computationally prohibitive in practice (Barber et al., 2015; Fearnhead and Prangle, 2012). This is especially true when the prior density is highly diffuse compared with the target posterior density (Marin et al., 2012), as most proposals are rejected.

Algorithm 1 ABC rejection sampler

```
for i = 1, ..., N do
    repeat
        Sample prior, θ* ~ p(θ).
        Generate data, D_s ~ f(D | θ*).
    until d(D, D_s) ≤ ε
    Set θ^i ← θ*.
end for
```

To improve the efficiency of ABC rejection sampling, one can consider a likelihood-free modification of Markov chain Monte Carlo (MCMC) (Beaumont et al., 2002; Marjoram et al., 2003; Tanaka et al., 2006) in which a Markov chain is constructed with a stationary distribution identical to the desired posterior. Given the Markov chain is in state $\boldsymbol{\theta}^i$, a state transition is proposed via a proposal kernel, $K(\boldsymbol{\theta} \mid \boldsymbol{\theta}^i)$.

The Metropolis–Hastings (Hastings, 1970; Metropolis et al., 1953) state transition probability, h, can be modified within an ABC framework to yield

$$h = \begin{cases} \min\left(\dfrac{p(\boldsymbol{\theta}^*)K(\boldsymbol{\theta}^i \mid \boldsymbol{\theta}^*)}{p(\boldsymbol{\theta}^i)K(\boldsymbol{\theta}^* \mid \boldsymbol{\theta}^i)}, 1\right) & \text{if } d(\mathcal{D}, \mathcal{D}_s) \leq \epsilon, \\ 0 & \text{otherwise.} \end{cases}$$

The stationary distribution of such a Markov chain is the desired approximate posterior (Marjoram et al., 2003). Algorithm 2 provides a method for computing N_T iterations of this Markov chain.

While MCMC-ABC sampling can be highly efficient, the samples in the sequence, $\boldsymbol{\theta}^1, \ldots, \boldsymbol{\theta}^{N_T}$, are not independent. This can be problematic as it is possible for the Markov chain to take long excursions into regions of low posterior probability. This incurs additional bias that is potentially significant (Sisson et al., 2007). A poor choice of proposal kernel can also have considerable impact upon the efficiency of MCMC-ABC (Green et al., 2015). The question of how to choose the proposal kernel is non-trivial. Typically proposal kernels are determined heuristically. However, automatic and adaptive schemes are available to assist in obtaining near optimal proposals in some cases (Cabras et al., 2015; Roberts and Rosenthal, 2009). Another additional complication is that of determining when the Markov Chain has converged; this is a challenging problem to solve in practice (Roberts and Rosenthal, 2004).

Sequential Monte Carlo (SMC) sampling was introduced to address these potential inefficiencies (Del Moral et al., 2006) and later extended within an ABC context (Sisson et al., 2007; Drovandi and Pettitt, 2011; Toni et al., 2009). A set of samples, referred to as particles, is evolved through a sequence of ABC posteriors defined through a sequence of T acceptance thresholds, $\epsilon_1, \ldots, \epsilon_T$ (Sisson et al., 2007; Beaumont et al., 2009). At each step, $t \in [0, T]$, the current ABC posterior, $p(\boldsymbol{\theta} \mid d(\mathcal{D}, \mathcal{D}_s) < \epsilon_t)$, is approximated by a discrete distribution constructed from a set of N_P particles $\boldsymbol{\theta}_t^1, \ldots, \boldsymbol{\theta}_t^{N_P}$ with