Approximate Bayesian Computation

Christopher C. Drovandi
School of Mathematical Sciences, Queensland University of Technology, Australia
email: c.drovandi@qut.edu.au

February 7, 2017

Abstract

Bayesian statistics provides a principled framework for performing statistical inference for an unknown parameter of a stochastic model assumed to be responsible for generating some observed data. However, standard Bayesian algorithms to sample from the posterior require that the likelihood function, the probability density of the data given the parameter represented as a function of the parameter for fixed observed data, is computationally tractable. However, there are an increasing number of models across Science and Technology where the likelihood function is difficult or impossible to compute. When simulation from the model is comparatively cheaper, a class of likelihood-free methods called approximate Bayesian computation (ABC) can be used. However, ABC introduces an approximation to the posterior. This paper gives an introduction to ABC, describes the approximation behaviour of ABC and provides advice on the successful implementation of ABC. Some current challenges facing ABC methods are also discussed.

Keywords: approximate Bayesian computation, pseudo-marginal methods, Markov chain Monte Carlo, sequential Monte Carlo, synthetic likelihood
1 Introduction

Statistical inferences in a Bayesian framework stem from the posterior distribution of unknowns

\[ p(\theta|y) \propto p(y|\theta)p(\theta), \]

where \( \theta \in \Theta \subseteq \mathbb{R}^p \) is a \( p \)-dimensional unknown parameter with \textit{a priori} uncertainty quantified by the prior distribution \( p(\theta) \) and updated through the likelihood function \( p(y|\theta) \) that is implicitly defined through the choice of a statistical model for observed data \( y \in \mathcal{Y} \), which is subsequently assumed fixed. In a model selection or averaging scenario, the set of unknowns may also include a random variable for the model indicator, but we do not consider this problem here. Many Bayesian approaches require the calculation of the likelihood \( p(y|\theta) \), which may be computationally expensive or impossible to obtain for complex models that practitioners may wish to consider. A possible approach for handling such models is approximate Bayesian computation (ABC), which is the focus of this paper.

The rest of the introduction highlights situations when ABC may be required and outlines the purpose of the paper.

In all but the simplest of applications, the posterior distribution \( p(\theta|y) \) does not have a recognisable distribution for which analytical properties are known. However, if it were possible to generate a sample \( \theta_1, \ldots, \theta_N \sim p(\theta|y) \) of size \( N \) from the posterior an estimate of the posterior expectation of a parameter functional \( f(\theta) \) can be obtained via Monte Carlo methods

\[
E_{\theta|y}[f(\theta)] \approx \frac{1}{N} \sum_{i=1}^{N} f(\theta_i). \tag{1}
\]

Unfortunately, it is generally not possible to sample directly from the posterior. Instead, indirect sampling methods may be useful. One such method is importance sampling (IS), where independent and identically distributed (iid) samples are drawn \( \theta_1, \ldots, \theta_N \sim g(\theta) \) where the importance distribution \( g(\theta) \) is chosen to be easy to sample from. The samples are then weighted to reflect the target posterior distribution \( p(\theta|y) \) where the weights are
given by

\[ W_i \propto \frac{p(y|\theta_i)p(\theta_i)}{g(\theta_i)}, \]

such that \( \sum_{i=1}^N W_i = 1 \). The average in (1) can be replaced by a weighted average. More sophisticated IS-based methods such as sequential Monte Carlo (Del Moral et al., 2006) also exist. Such methods involve similar weight calculations.

Markov chain Monte Carlo (MCMC) methods (Brooks et al., 2011) have also been developed as a powerful method for sampling from the posterior distribution. The idea of MCMC is to construct an ergodic Markov chain whose limiting distribution is \( p(\theta|y) \).

Assume that the current value of the chain is \( \theta \). The Metropolis-Hastings method, a very general MCMC algorithm, proceeds by proposing \( \theta^* \sim q(\cdot|\theta) \) from a user-specified proposal distribution \( q \) and accepting it as the new value of the chain with probability

\[
\min \left\{ 1, \frac{p(y|\theta^*)p(\theta^*)q(\theta^*|\theta)}{p(y|\theta)p(\theta)q(\theta^*|\theta)} \right\},
\]

otherwise the new value of the chain is the current \( \theta \). Following convergence, running the chain for \( N \) iterations produces \( N \) correlated samples from the posterior.

The key observation to motivate this paper is that indirect approaches such as IS and MCMC inspired methods require the ability to evaluate \( p(y|\theta) \) point-wise as a function of \( \theta \). If the likelihood calculation is even moderately expensive, the overall computation required to achieve a posterior sample with a suitable effective sample size may be too heavy. With a growing interest in developing more realistic models across a diverse set of disciplines such as biology, ecology, genetics, hydrology and finance, there has been a demand for the development of new computational tools for performing statistical inference.

One useful approach that has had considerable success involves noting that whilst the observed data likelihood \( p(y|\theta) \) may be intractable, the so-called complete data likelihood \( p(y, z|\theta) \) may be feasible to compute, where \( z \in \mathcal{Z} \) represent the unobserved (or artificially introduced) variables of the model. This is often referred to as data augmentation (Tanner and Wong, 1987). A Bayesian approach that samples from the joint distribution \( p(\theta, z|y) \propto p(y|z, \theta)p(\theta)p(z|\theta) \) defined on \( \Theta \times \mathcal{Z} \) by default also samples from the \( \theta \)-marginal.
by ignoring the \( z \) samples. A common method for sampling from \( p(\theta, z|y) \) involves the construction of a component-wise MCMC algorithm that iteratively samples from the conditional posteriors \( p(\theta|z, y) \) and \( p(z|\theta, y) \) in some fashion. The data augmentation approach may not be successful for applications where \( z \) is very high dimensional. Firstly, it may be impossible to explore the joint posterior space of \( z \) and \( \theta \) in a feasible amount of time, especially if there is posterior dependence between \( z \) and \( \theta \). Secondly, substantial mathematical and algorithmic expertise may be required to determine an MCMC algorithm with relatively efficient mixing properties.

More recently, a suite of methods called exact-approximate MCMC (or pseudo-marginal) algorithms (Andrieu and Roberts, 2009) have been developed to overcome some of the issues of data augmentation MCMC. Pseudo-marginal methods attempt to mimic an ideal sampler on the marginal space of \( \theta \) by replacing the intractable likelihood \( p(y|\theta) \) with an unbiased estimator. The resulting MCMC algorithm maintains the target posterior \( p(\theta|y) \) as its limiting distribution. The main advantage is that MCMC proposals can be made directly on the space of \( \theta \), independently of \( z \). The main issue is that pseudo-marginal methods are only successful if the variance of the likelihood estimator is reduced to a suitable level (Doucet et al., 2015). In challenging applications, it may not be computationally feasible to achieve such a low variance estimator.

When the proposed model for explaining the observed data \( y \) is too complex for the aforementioned approaches, ABC can be applied, provided that it is relatively cheap to simulate data from the model. The drawback of ABC is that it generally produces an approximation to the true posterior distribution \( p(\theta|y) \). The purpose of this paper is to assist the reader in understanding the approximation behaviour of ABC and provide advice on the successful implementation of ABC. We will also describe the current challenges facing ABC methods that may provide some insight to future research directions for ABC.

2 Putting the ‘A’ in ABC

ABC avoids evaluation of the likelihood \( p(y|\theta) \) by instead simulating data \( x \in \mathcal{Y} \) from the model. In essence, if a proposed \( \theta \) value generates \( x \) that is ‘close enough’ to \( y \), then \( \theta \) is
included in the posterior sample. If $y$ is high dimensional, it is computationally impractical to simulate $x$ very close to $y$, even when a $\theta$ value with high posterior support is proposed. For this reason, the first step in ABC approaches often involves reducing the full dataset $y$ down to a set of summary statistics $s_y = S(y)$ where $S(\cdot) : \mathcal{Y} \rightarrow \mathcal{S} \subseteq \mathbb{R}^d$ and $d \geq p$ is the number of summary statistics. Thus ABC attempts to target the posterior conditional on the summary

$$p(\theta|s_y) \propto p(s_y|\theta)p(\theta).$$

The posteriors $p(\theta|s_y)$ and $p(\theta|y)$ are equivalent only when $s_y$ is a sufficient statistic. Unfortunately, a low dimensional sufficient statistic is generally unavailable in applications where ABC is used. Thus the first source of error in ABC is the loss of information in reducing $y \rightarrow s_y$, but is often a necessary price to pay to reduce a second source of error, the ‘close enough’ criterion, that we discuss in more detail later. A brief account of dimension reduction methods used in ABC is provided in Section 5.

Unfortunately, the summary statistic likelihood $p(s_y|\theta)$ is also typically intractable. Instead, ABC approximates the intractable likelihood with the following ABC likelihood

$$p_\epsilon(s_y|\theta) = \int_{\mathcal{Y}} K_\epsilon(\rho(s_x, s_y)) p(x|\theta) dx.$$  \hspace{1cm} (2)

In this expression, $\rho(s_x, s_y)$ is a function that compares the simulated data with the observed data, where $s_x = S(x)$. Here $K_\epsilon(\cdot)$ is a kernel weighting function with bandwidth $\epsilon$, which is designed to give higher weight when $\rho$ is small. The distance function $\rho$ and the ABC tolerance $\epsilon$ both need to be selected by the practitioner. The integral in (2) is analytically intractable, but can be estimated unbiasedly by taking $n$ iid simulations from the model, \( \{x_i\}_{i=1}^n \overset{\text{iid}}{\sim} p(x|\theta), \) computing the corresponding summary statistics \( \{s_i\}_{i=1}^n \) where $s_i = S(x_i)$, and calculating

$$p_{\epsilon,n}(s_y|\theta) = \frac{1}{n} \sum_{i=1}^n K_\epsilon(\rho(s_i, s_y)).$$  \hspace{1cm} (3)

Using the above unbiased estimator of (2) is generally sufficient to obtain a Bayesian algorithm that targets $p_\epsilon(\theta|s_y) \propto p_\epsilon(s_y|\theta)p(\theta)$. Based on (3), ABC can be viewed as pro-
ducing a non-parametric estimator of the density $p(s_y|\theta)$ (Blum, 2010). This highlights the second source of error in ABC methods. A non-zero value of $\epsilon$ introduces a bias into the approximation of $p(s_y|\theta)$. The smaller the value of $\epsilon$, the smaller the bias. However, the variance of the estimator (3) increases and can only be reduced by increasing $n$, thus raising the computational cost. Intuitively, a small value of $\epsilon$ enforces stricter matching between observed and simulated data, increasing accuracy, but the probability of generating simulated data close enough reduces, hence increasing the computational cost.

The user-specified parameters $\epsilon$, $\rho$ and $K_\epsilon$ can all have an impact on the ABC posterior. The value of $\epsilon$ should be chosen as small as possible such that the computation is still manageable. Regarding the distance function, Prangle (2016) investigate in detail the weighted Euclidean distance, so that $\rho$ is not dominated by a subset of the summaries. Other works (e.g. Vo et al. (2015)) have considered the Mahalanobis distance for $\rho$, which also attempts to account for correlations between summaries. The choice of $K_\epsilon$ appears to be less critical in practice. A convenient choice in many circumstances is the indicator kernel, $K_\epsilon(\rho(s_x, s_y)) \propto I(\rho(s_x, s_y) \leq \epsilon)$. Another option is the Gaussian kernel, $K_\epsilon(\rho(s_x, s_y)) \propto \exp(-\rho(s_x, s_y)/2\epsilon)$, which Bonassi and West (2015) demonstrate can provide some mild computational improvements as it leads to higher weight when $\rho(s_x, s_y)$ is smaller.

In summary, the ABC posterior is only equivalent to the true posterior when a sufficient statistic (e.g. the full dataset $y$) is used and when perfect matching can be performed on the basis of the simulated and observed sufficient statistic (i.e. $\epsilon = 0$). However, it is often necessary to reduce $y$ down to a non-sufficient $s_y$ (loss of information) and it is generally infeasible to perfectly match $s_x$ with $s_y$ (introduction of $\epsilon$). As in most Bayesian analyses, expectations are estimated by sampling from the (approximate) posterior. Thus the third source of error in ABC is Monte Carlo error. We discuss approaches to sample from the ABC posterior next.
Here we discuss algorithms to sample from the ABC posterior distribution \( p_\epsilon(\theta|s_y) \) for a particular choice of \( s_y, \rho, K_\epsilon \), and \( \epsilon \). The choice of ABC algorithm is often an important consideration. Implementation of a highly efficient algorithm can be required as generating an \( s_x \) sufficiently close to \( s_y \) so that \( p_\epsilon(\theta|s_y) \approx p(\theta|s_y) \) can be considered as a rare event, even when the proposed \( \theta \) has high posterior support. The number of proposed parameter values in IS or the length of the chain in MCMC generally needs to be significantly larger when performing ABC compared to a regular Bayesian algorithm that uses likelihood evaluations.

An importance sampling (IS) ABC method is shown in Algorithm 1. The method samples \( N \) independent draws from an importance distribution \( g(\theta) \) and produces a weighted sample \( \{W_i, \theta_i\}_{i=1}^N \) from \( p_\epsilon(\theta|s_y) \). It differs from standard IS in that simulation from the model is required to avoid likelihood evaluations. Since ABC uses a stochastic likelihood estimate \( K_\epsilon(\rho(s_i, s_y)) \), the importance weights exhibit significant variability relative to a regular IS procedure that uses the actual likelihood, meaning that the value \( N \) must be significantly increased to generate a comparable effective sample size.

Algorithm 1 General ABC IS algorithm.

**Input:** The importance distribution \( g(\theta) \), the target ABC posterior distribution \( p_\epsilon(\theta|s_y) \) and the number of importance samples \( N \).

**Output:** A weighted sample \( \{W_i, \theta_i\}_{i=1}^N \) from \( p_\epsilon(\theta|s_y) \).

1: Sample \( \{\theta_i, x_i\}_{i=1}^N \) iid \( \sim p(y|\theta)g(\theta) \).
2: For each \( i = 1, \ldots, N \) compute \( s_i = S(x_i) \) and calculate \( \rho_i = \rho(s_i, s_y) \).
3: For each \( i = 1, \ldots, N \) compute the unnormalised importance weight \( w_i = \frac{K_\epsilon(\rho(s_i, s_y))p(\theta_i)}{g(\theta_i)} \).
4: For each \( i = 1, \ldots, N \) normalise the weights \( W_i = w_i/\sum_{j=1}^N w_j \).

An ABC algorithm that is commonly used in the applied literature is ABC rejection, which is a special case of ABC IS. In ABC rejection, we set \( g(\theta) \) as the prior \( p(\theta) \) and the kernel to be the indicator \( K_\epsilon(\rho(s_x, s_y)) \propto \mathbb{I}(\rho(s_x, s_y) \leq \epsilon) \). In this case, the importance weights are either zero or proportional to a constant. With this implementation, we do not need to pre-specify the value of \( \epsilon \) by choosing a priori the effective number of samples in the ABC posterior. The method is shown in Algorithm 2, which is due to Pritchard et al.
Firstly, a large collection of samples is drawn from the prior predictive distribution. Then, for each simulation, a discrepancy value is calculated. The approximation to the ABC posterior consists of all the samples from the prior \( \theta \) where \( i \in \{1, \ldots, N\} \) such that \( \rho(s_i, s_y) \leq \epsilon \) where \( \epsilon \) is taken to be the empirical \( \alpha \)-quantile of the collection of discrepancy values \( \{\rho(s_i, s_y)\}_{i=1}^{N} \). A smaller value of \( \alpha \) leads to a smaller value of \( \epsilon \) (less bias) but a fewer effective number of samples (increased Monte Carlo error).

**Algorithm 2 ABC rejection algorithm of Pritchard et al. (1999).**

**Input:** The prior distribution \( p(\theta) \), a summary statistic function \( S(\cdot) \) such that \( s_y = S(y) \), the discrepancy function \( \rho \), the approximate proportion of samples to keep in the ABC posterior \( \alpha \) and the number of importance samples \( N \).

**Output:** A sample \( \{\theta_i\}_{i=1}^{N} \) from \( p_{\epsilon}(\theta|s_y) \) where \( \epsilon \) is also determined from the algorithm.

1: Sample \( \{\theta_i, x_i\}_{i=1}^{N} \) iid \( p(y|\theta)p(\theta) \)
2: For each \( i = 1, \ldots, N \) compute \( s_i = S(x_i) \) and calculate \( \rho_i = \rho(s_i, s_y) \)
3: Determine \( \epsilon \) as the empirical \( \alpha \)-quantile of the discrepancies
4: Keep all the \( \theta_i \) such that the corresponding \( \rho_i \leq \epsilon \)

There are several appealing features of the ABC IS and rejection approaches. Firstly, the algorithms are easily implemented on parallel computing architectures. Further, step 1 of these algorithms is independent of the observed data, which means that these large collection of simulations can be re-used for different choices of the summary statistic, discrepancy function, datasets from the same model and performing simulation studies to validate a particular choice of summary statistic and/or discrepancy function. However, when the posterior is substantially different to the prior, ABC rejection can be highly inefficient. For ABC IS, it is generally difficult to specify a \( g(\theta) \) that is close to \( p(\theta|s_y) \).

In order to improve the efficiency of ABC algorithms, several advanced sampling strategies have been developed. Only a brief coverage of this literature will be provided here. Marjoram et al. (2003) develop an MCMC approach to ABC, which allows proposals to be made locally in high posterior support regions. Here a proposal is made on the space \( \Theta \times Y \). More specifically, the MCMC ABC proposal is \( q(\theta^*, x^*|\theta, x) = p(x^*|\theta^*)q(\theta^*|\theta) \). With this choice of proposal, it is easy to show that any likelihood evaluations cancel in the Metropolis-Hastings ratio. Assuming the indicator kernel, the MCMC acceptance probability is

\[
\min \left\{ 1, \frac{p(\theta^*)q(\theta|\theta^*)}{p(\theta)q(\theta^*|\theta)} \mathbb{1}(\rho(s_x, s_y) \leq \epsilon) \right\},
\]
but of course any kernel can be used. A more comprehensive review of MCMC ABC is provided in Sisson and Fan (2011). Unfortunately, the MCMC ABC approach has some drawbacks. It is prone to stickiness in the tails of the posterior and can suffer from slow convergence (Sisson and Fan, 2011). Further, it inherits the drawbacks of standard MCMC approaches generally. MCMC is not an embarrassingly parallel algorithm and it can have difficulty exploring posteriors that are multi-modal and/or have complicated landscapes.

SMC ABC methods can help to overcome the issues of MCMC ABC, and have been successfully applied in many different areas. Most commonly, SMC ABC methods work by traversing a set of weighted samples (or particles) through a sequence of ABC posterior distributions defined by a sequence of non-increasing tolerances

\[ p_{\epsilon,t}(\theta, x|s_y) \propto p(\theta)p(x|\theta)I(\rho(s_x, s_y) \leq \epsilon_t), \]

where \( \epsilon_1 \geq \epsilon_2 \geq \cdots \geq \epsilon_T \) and \( \epsilon_T \) is the target tolerance. Proposals for \( \theta \) at target \( t + 1 \) can be informed by the population of particles \( \{\theta_{t,i}, W_{t,i}\}_{i=1}^{N} \) at target \( t \). Further, it is possible to determine the sequence \( \{\epsilon_t\}_{t=1}^{T} \) adaptively and the stopping rule can be based on whether \( \epsilon_T \) has been reached or the acceptance rate of the method becomes unacceptably low. For these reasons SMC ABC methods require little tuning. Moreover, since SMC uses a population of particles across the parameter space, the algorithm is easy to implement with parallel computing and it is better equipped to deal with more irregular posterior distributions. Various SMC ABC methods can be found in, but not limited to, Sisson et al. (2007), Beaumont et al. (2009), Drovandi and Pettitt (2011), Del Moral et al. (2012) and Vo et al. (2015). One drawback of SMC ABC is the samples from the intermediate posteriors \( t = 1, \ldots, T - 1 \) are typically not of interest and thus the algorithm is somewhat wasteful, especially if many intermediate distributions are required to bridge the prior and the posterior.

Most ABC algorithms are relatively simple to implement once the code to simulate from the model has been developed.
4 Regression Adjustment

The aim of regression adjustment is to predict the ABC posterior for $\epsilon = 0$ (effectively $p(\theta|s_y)$) by exploiting any remaining relationship between the parameter and the corresponding simulated summary statistic in the ABC posterior sample, which we denote as $\{\theta_i, s_i\}_{i=1}^N$, where $N$ is the number of samples. The first regression adjustment approach is due to the seminal ABC publication Beaumont et al. (2002), which uses local linear regression. Here an extension based on neural networks, developed in Blum and François (2010), is described. For simplicity suppose $\theta$ is univariate. Blum and François (2010) consider the model

$$\theta_i = \mu(s_i) + \sigma(s_i)\eta_i,$$

where $\eta_i, i = 1, \ldots, N,$ are independently distributed errors with zero mean and variance one. The $\mu(s)$ and $\sigma(s)$ are flexible mean and standard deviation functions, which are fitted using neural networks. We denote the fitted functions as $\hat{\mu}(s)$ and $\hat{\sigma}(s)$. Let $\hat{\eta}_i$ denote the empirical residual $\hat{\eta}_i = \hat{\sigma}(s_i)^{-1}(\theta_i - \hat{\mu}(s_i))$. Estimating the posterior distribution $p(\theta|s_y)$ by the fitted regression model at $s_y$ and the empirical residuals gives that

$$\theta^a_i = \hat{\mu}(s_y) + \hat{\sigma}(s_y)\hat{\eta}_i$$

$$= \hat{\mu}(s_y) + \hat{\sigma}(s_y)\hat{\sigma}(s_i)^{-1}(\theta_i - \hat{\mu}(s_i)),$$

$i = 1, \ldots, N$ comprise an approximate sample from $p(\theta|s_y)$ if the regression model is correct. For multivariate $\theta$, estimates of the univariate ABC posterior distributions can be obtained by applying the regression adjustment approach to each component of $\theta$ separately. Alternatively, it is possible to consider multivariate regression models, but these are more difficult to fit. Transformations may need to be applied prior to regression adjustment to ensure that the adjusted values remain within prior limits.

5 Selecting Summary Statistics

In the majority of ABC applications, a good choice of summary statistic is low dimensional whilst maintaining the majority of information contained in $y$. Most approaches begin
with some initial large collection of potential summaries that may be informative for one
or more of the components of $\theta$. Then, a subsetting procedure (e.g. Nunes and Balding
(2010)) or a dimension reduction method is applied. The best subset procedure can be
expensive as it, in the worse case, involves enumerating all possible combinations of the
summary subsets (Blum et al., 2013). Thus here we describe some dimension reduction
techniques, with a more comprehensive review provided in Blum et al. (2013).

One approach that has been highly cited and found success (e.g. Vo et al. (2015) and
Cameron and Pettitt (2012)) in the literature is the semi-automatic method of Fearnhead
and Prangle (2012). Here estimates of the posterior means are used as summary statistics.
These are obtained from regressing the parameters individually on a large collection of
summaries based on the output of a short pilot run of ABC. Fearnhead and Prangle (2012)
use parametric regressions but non-parametric machine learning regression can also be used
(Aeschbacher et al., 2012).

Summary statistics can also be formed based on ideas from indirect inference (Gourieroux
et al., 1993). This can involve using the parameter estimate or score of an alternative
parametric model that has a tractable likelihood and fits the data as the summary statistic
(see Drovandi et al. (2015) and some references therein).

Some other dimension reduction approaches applied to ABC include partial least squares
(Wegmann et al., 2009), neural networks (Blum and François, 2010) and ridge regression
(Blum et al., 2013).

6 Challenges for ABC Methods

Despite the immense success of ABC methods in providing a principled method for statisti-
cal inference of complex models in important applications, there remain several statistical
challenges. Some of these existing challenges are described here.

Methods that can handle a high-dimensional summary statistic without having to resort
to dimension reduction or subsetting are of interest. Price et al. (2017) develop a Bayesian
version of the synthetic likelihood (Wood, 2010), which assumes a multivariate normal dis-
tribution for the summary statistic with a mean and covariance matrix that depends on $\theta$. 
Price et al. (2017) demonstrate that the synthetic likelihood is capable of handling higher dimensional summary statistics relative to ABC. However, the normality assumption may not always be reasonable.

There is also a lack of ABC methods for dealing with a high-dimensional parameter, which also implies a high-dimensional summary statistic. Nott et al. (2014) develop a marginal adjustment method where each univariate posterior distribution can be accurately estimated by focusing only on statistics that are relevant for each component of the parameter vector \( \theta \). Li et al. (2017) extend this idea by obtaining all univariate and bivariate ABC posteriors and combining them with a Gaussian copula representation of the joint ABC posterior. Further research is required in this area.

A typical ABC analysis may require hundreds of thousands or millions of simulations from the model. This may be prohibitive for some applications even when the model simulation is only moderately expensive. The synthetic likelihood approach does help to reduce the number of model simulations, at the expense of the normality assumption (Price et al., 2017). Continued efforts in efficient algorithm development are required. An alternative approach to reduce the number of model simulations is emulation. For example, Gutmann and Corander (2016) fit a Gaussian process to the discrepancy function \( \rho \) at a set of ‘training’ points. For a proposed value of \( \theta \), the discrepancy value can be predicted from the Gaussian process without having to simulate the model, provided there is enough certainty in the prediction. However, there are several difficulties with the emulation approach. The approach does not scale efficiently with an increase in the number of training points, which naturally needs to increase as the dimension of \( \theta \) grows. Further, it is difficult to devise an appropriate method for where to place the training simulations.

7 Related Articles

Bayesian Analysis and Markov Chain Monte Carlo Simulation; Bayesian Inference; Bayesian Inference of Markov Processes; Convergence and Mixing in Markov Chain Monte Carlo: Advanced Algorithms and Latest Developments; Data Augmentation; Generalized Method-of-Moments Estimation; Kernel Density Estimation; Markov Chain Monte Carlo Algo-
rithms; Markov Chain Monte Carlo (MCMC); Monte Carlo Methods; Monte Carlo Methods, Sequential; The Metropolis-Hastings Algorithm

Acknowledgments

The author was supported by an Australian Research Council’s Discovery Early Career Researcher Award funding scheme (DE160100741).

References

Aeschbacher, S., Beaumont, M. A., and Futschik, A. (2012). A novel approach for choosing summary statistics in approximate Bayesian computation. Genetics, 192(3):1027–1047.

Andrieu, C. and Roberts, G. O. (2009). The pseudo-marginal approach for efficient Monte Carlo computations. The Annals of Statistics, 37(2):697–725.

Beaumont, M. A., Cornuet, J.-M., Marin, J.-M., and Robert, C. P. (2009). Adaptive approximate Bayesian computation. Biometrika, 96(4):983–990.

Beaumont, M. A., Zhang, W., and Balding, D. J. (2002). Approximate Bayesian computation in population genetics. Genetics, 162(4):2025–2035.

Blum, M. G. B. (2010). Approximate Bayesian computation: a non-parametric perspective. Journal of the American Statistical Association, 105(491):1178–1187.

Blum, M. G. B. and François, O. (2010). Non-linear regression models for Approximate Bayesian Computation. Statistics and Computing, 20(1):63–73.

Blum, M. G. B., Nunes, M. A., Prangle, D., and Sisson, S. A. (2013). A comparative review of dimension reduction methods in approximate Bayesian computation. Statistical Science, 28(2):189–208.

Bonassi, F. V. and West, M. (2015). Sequential Monte Carlo with adaptive weights for approximate Bayesian computation. Bayesian Analysis, 10(1):171–187.
Brooks, S., Gelman, A., Jones, G., and Meng, X.-L. (2011). *Handbook of Markov Chain Monte Carlo*. CRC press.

Cameron, E. and Pettitt, A. (2012). Approximate Bayesian computation for astronomical model analysis: a case study in galaxy demographics and morphological transformation at high redshift. *Monthly Notices of the Royal Astronomical Society*, 425(1):44–65.

Del Moral, P., Doucet, A., and Jasra, A. (2006). Sequential Monte Carlo samplers. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(3):411–436.

Del Moral, P., Doucet, A., and Jasra, A. (2012). An adaptive sequential Monte Carlo method for approximate Bayesian computation. *Statistics and Computing*, 22(5):1009–1020.

Doucet, A., Pitt, M. K., Deligiannidis, G., and Kohn, R. (2015). Efficient implementation of Markov chain Monte Carlo when using an unbiased likelihood estimator. *Biometrika*, 102(2):295–313.

Drovandi, C. C. and Pettitt, A. N. (2011). Estimation of parameters for macroparasite population evolution using approximate Bayesian computation. *Biometrics*, 67(1):225–233.

Drovandi, C. C., Pettitt, A. N., and Lee, A. (2015). Bayesian indirect inference using a parametric auxiliary model. *Statistical Science*, 30(1):72–95.

Fearnhead, P. and Prangle, D. (2012). Constructing summary statistics for approximate Bayesian computation: Semi-automatic ABC (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 74(3):419–474.

Gourieroux, C., Monfort, A., and Renault, E. (1993). Indirect inference. *Journal of Applied Econometrics*, 8(S1):S85–S118.

Gutmann, M. U. and Corander, J. (2016). Bayesian optimization for likelihood-free inference of simulator-based statistical models. *Journal of Machine Learning Research*, 17(125):1–47.
Li, J., Nott, D. J., Fan, Y., and Sisson, S. A. (2017). Extending approximate Bayesian computation methods to high dimensions via Gaussian copula. *Computational Statistics & Data Analysis*, 106:7789.

Marjoram, P., Molitor, J., Plagnol, V., and Tavaré, S. (2003). Markov chain Monte Carlo without likelihoods. *Proceedings of the National Academy of Sciences*, 100(26):15324-15328.

Nott, D. J., Fan, Y., Marshall, L., and Sisson, S. A. (2014). Approximate Bayesian computation and Bayes linear analysis: toward high-dimensional ABC. *Journal of Computational and Graphical Statistics*, 23(1):65–86.

Nunes, M. A. and Balding, D. J. (2010). On optimal selection of summary statistics for approximate Bayesian computation. *Statistical Applications in Genetics and Molecular Biology*, 9(1):34.

Prangle, D. (2016). Adapting the ABC distance function. *To appear in Bayesian Analysis*.

Price, L. F., Drovandi, C. C., Lee, A., and Nott, D. J. (2017). Bayesian synthetic likelihood. *To appear in Journal of Computational and Graphical Statistics*.

Pritchard, J. K., Seielstad, M. T., Perez-Lezaun, A., and Feldman, M. W. (1999). Population growth of human Y chromosomes: a study of Y chromosome microsatellites. *Molecular Biology and Evolution*, 16(12):1791–1798.

Sisson, S. A. and Fan, Y. (2011). *MCMC handbook*, chapter Likelihood-free Markov chain Monte Carlo, pages 313–335. Chapman & Hall.

Sisson, S. A., Fan, Y., and Tanaka, M. M. (2007). Sequential Monte Carlo without likelihoods. *Proceedings of the National Academy of Sciences*, 104(6):1760–1765.

Tanner, M. A. and Wong, W. H. (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association*, 82(398):528–540.

Vo, B. N., Drovandi, C. C., Pettitt, A. N., and Pettet, G. J. (2015). Melanoma cell colony expansion parameters revealed by approximate Bayesian computation. *PLOS Computational Biology*, 11(12):e1004635.
Wegmann, D., Leuenberger, L., and Excoffier, L. (2009). Efficient approximate Bayesian computation coupled with Markov chain Monte Carlo without likelihood. *Genetics*, 182:1207–1218.

Wood, S. N. (2010). Statistical inference for noisy nonlinear ecological dynamic systems. *Nature*, 466:1102–1107.