Approximate Bayesian Computational methods for the inference of unknown parameters

Yuqin Ke and Tianhai Tian

Abstract Recent advances in biology, economics, engineering and physical sciences have generated a large number of mathematical models for describing the dynamics of complex systems. A key step in mathematical modelling is to estimate model parameters in order to realize experimental observations. However, it is difficult to derive the analytical density functions in the Bayesian methods for these mathematical models. During the last decade, approximate Bayesian computation (ABC) has been developed as a major method for the inference of parameters in mathematical models. A number of new methods have been designed to improve the efficiency and accuracy of ABC. Theoretical studies have also been conducted to investigate the convergence property of these methods. In addition, these methods have been applied to a wide range of deterministic and stochastic models. This chapter gives a brief review of the main ABC algorithms and various improvements.

1 Introduction

Since more and more natural and social science problems involve the uncertainty in observations, statistical models and parameter inference play an important role in the development of mathematical methods for studying real-world problems. In particular, the era of big data has generated huge amount of data whose volume is increasing at a very fast speed. Mathematical and statistical models are becoming more and more complex in terms of the network size and regulatory relationships. Thus effective and efficient methods are strongly needed to infer unknown parameters in these models in order to reduce the simulation errors against the experimental data.

There are two major types of inference methods, namely the optimization methods and Bayesian statistical methods. The optimization methods are designed to

School of Mathematical Sciences, Monash University, Clayton VIC 3800, Australia
e-mail: yuqin.ke@monash.edu, tianhai.tian@monash.edu

minimize an objective function by searching for parameters within a given parameter space in a directed manner. The inferred set of parameters produces the best fit to the experimental data [30]. A variety of effective approaches have been developed in recent years. Among them, the genetic algorithm is a popular and effective approach and has been widely applied to various models [50]. These methods all share two main ingredients: a cost function for specifying the distance between simulated data and experimental data and an optimization algorithm for searching for parameters in order to optimize the cost function. However, when the landscape of the cost function is complex, it is difficult for these methods to find the global optimum. To tackle this challenge, the global optimization methods have been proposed to explore the complex surfaces as widely as possible. Comparison studies have been conducted to examine the efficiency of several global optimization algorithms for the test models [21].

Compared with the optimization methods, the Bayesian inference methods can estimate the probability distributions of parameters by using the Bayes' rule to update the prior probability estimates. In addition, Bayesian methods are more robust in dealing with stochastic models and/or experimental data with noise [22, 55]. In recent years Bayesian methods have been successfully used in a diverse range of fields and provide the promise to applications [48]. The recent advances in approximate Bayesian computation (ABC) provide effective methods without any restriction on the requirement of the likelihood function. This chapter provides a brief review for the recent development in ABC, including the rejection ABC, regression ABC, Markov chain Monte Carlo (MCMC) ABC and sequential Monte Carlo (SMC) ABC. We also discuss the relevant improvements and extensions of these methods, such as the choice of summary statistics.

2 Principle of Bayesian Inference

For the Bayesian inference problems, model parameters are treated as random quantities along with the observation data. The Bayesian inference involves the estimation of the posterior probability

$$p(\theta|y) = \frac{p(y|\theta)\pi(\theta)}{p(y)} \propto p(y|\theta)\pi(\theta), \tag{1}$$

where y is the observation data and the parameter vector of the model is θ ($\theta \in \Theta \subseteq \mathbb{R}^q, q \geq 1$). In addition, $\pi(\theta)$ is the prior distribution representing the prior beliefs about the parameters under investigation, and $p(y|\theta)$ is a likelihood function of parameter θ. The marginal distribution, defined by

$$p(y) = \int_{\theta \in \Theta} p(y|\theta)\pi(\theta)d\theta \tag{2}$$

often involves a high-dimensional integral, and $p(\theta|y)$ is the posterior probability distribution which expresses the uncertainty regarding θ conditional on the observed experimental data y. All the Bayesian inference about θ will be based on the estimated $p(\theta|y)$. However, the integrations which produce the Bayesian quantities of interest (such as marginal posteriors, marginal moments, and probablitiy intervals) can only be performed analytically when the density function $p(y|\theta)$ is available, which can be achieved only for relatively simple cases.

When the density function $p(y|\theta)$ is available, the classic Metropolis-Hasting algorithm is applied to find a Markov chain of the parameters, which is given below.

Algorithm 1: Metropolis-Hasting algorithm

Given the observation data y_{obs}, proposal distribution $q(\cdot)$, and an initial sample from the prior distribution $\theta^{(0)} \sim \pi(\theta)$.

At iteration $i \geq 0$

1. Generate a sample from the proposal distribution $\theta^{'} \sim q(\theta|\theta^{(i)})$.
2. Draw a sample from the uniform distribution $\mu \sim U(0,1)$ and calculate the ratio
$$\alpha = \min(1, \frac{\pi(\theta^{'})p(y_{obs}|\theta^{'})q(\theta^{(i)}|\theta^{'})}{\pi(\theta^{(i)})p(y_{obs}|\theta^{(i)})q(\theta^{'}|\theta^{(i)})}). \tag{3}$$
3. If $\mu \leq \alpha$, accept the sample as $\theta^{(i+1)} = \theta^{'}$; otherwise reject the sample.
4. Repeat steps $1 \sim 3$ until the requeired number of posterior samples is obtained.

Based on the classic Metropolis-Hasting algorithm, a number of more sophisticated methods have been designed, such as the Markov chain Monte Carlo (MCMC), the importance sampling (IS), and the sequential Monte Carlo (SMC) [20, 42]. The MCMC sampling methods usually break a high-dimensional problem into a number of smaller dimensional problems and generate a sample of dependent or correlated draws which can be treated as a realizetion of a Markov chain with equilibrium distribution equal to $p(\theta|y)$. Once the convergence to $p(\theta|y)$ occurs, any subsequent simulated value can be viewed as a sample from $p(\theta|y)$ and all these samples are used to estimate the posterior quantities of interest.

An alternative approach is the Gibbs sampling if the marginal distribution of each parameter is available. In the basic version, the Gibbs sampling is a special case of the Metropolis-Hastings algorithm. However, in its extended versions, these methods can be considered as a framework for sampling each variable (or more generally, each group of variables) from a number of variables in turn. It can also be incorporated into the Metropolis-Hastings algorithm (or other methods) to implement in one or more sampling steps. The detail of this algorithm is given below.

Algorithm 2: Gibbs Sampling

Given the observation data y_{obs}, and an initialize sample from the prior distribution $\theta^{(0)} = (\theta_1^{(0)}, \cdots, \theta_p^{(0)})^T \sim \pi(\theta)$.

At iteration $i \geq 0$

1. Generate a sample for $\theta_1^{(i+1)}$ using

$$\theta_1^{(i+1)} \sim \pi(\theta_1 | \theta_2^{(i)}, \cdots, \theta_p^{(i)}, y_{obs}).$$

2. For $j = 2, ..., p-1$, sample for $\theta_j^{(i+1)}$ using

$$\theta_j^{(i+1)} \sim \pi(\theta_j | \theta_1^{(i+1)}, ..., \theta_{j-1}^{(i+1)}, \theta_{j+1}^{(i)}, \cdots, \theta_p^{(i)}, y_{obs}).$$

3. Generate a sample for $\theta_p^{(i+1)}$ using

$$\theta_p^{(i+1)} \sim \pi(\theta_p | \theta_1^{(i+1)}, \theta_2^{(i+1)}, \cdots, \theta_{p-1}^{(i+1)}, y_{obs}).$$

4. Repeat steps $1 \sim 3$ until the requeired number of posterior samples is obtained.

Although these Bayesian inference methods are effective, they are based on the availability of the likelihood function. However, it may be difficult to derive the likelihood function directly for many complex models. For example, the analytical density function may not be available, or it may be expensive to calculate the likelihood. In some cases, the observed experimental data are insufficient to obtain a tractable likelihood. This intractability prohibits the direct implementation of a generic MCMC algorithm.

3 Rejection ABC method

To deal with complex models without analytical likelihood, a number of algorithms have been developed during the past two decades, which are referred to as the likelihood-free inference or Approximate Bayesian Computation (ABC). The ABC method is based on the following intuition: namely if a sample of the unknown parameter produces the simulation that matches the observed dataset, this sample should be close to the exact value of the parameter. Conversely, if the simulated dataset differs from the observed data substantially , this sample should not be considered as the estimate of the parameter. Thus the method strongly relies on the metric to determine the distance between simulated dataset and observed dataset. In the late 90's, ABC was first introduced as a rejection technique bypassing the computation of the likelihood function [49]. Later, Pritchard et al. proposed a generalisation based on an approximation of the target [39]. In recent years, the ABC methods have been proposed with various improvements and have been applied to a

wide range of application fields, such as population genetics, ecology, epidemiology and systems biology.

The ABC spirit is based on the following algorithm [44].

Algorithm 3: Likelihood-free rejection sampling

Given the observation data y_{obs}, and prior distribution $\pi(\theta)$.

1. Generate a sample from the prior distribution $\theta' \sim \pi(\theta)$.
2. Simulate the model using θ' to get a dataset $x \sim p(x|\theta')$.
3. Accept the sample θ' if $x = y_{obs}$, otherwise reject it.
4. Repeat the above steps until the required number of posterior samples is obtained.

In this paper Rubin just exhibited this algorithm as an intuitive way to understand the posterior distributions from a frequentist perspective rather than using it for interring models where the likelihood function was not available [44]. Then Tavaré *et al.* proposed an implementation of the rejection algorithm for the Bayesian inference of parameters in population genetics. When the data are discrete and of low dimension, this algorithm is effective. However, the probability of acceptance for a sample is usually very low.

As mentioned earlier, the rejection algorithm is dependent on a metric to measure the distance between the simulation and observation data. For inference problems with continuous distributions, or the datasets are high dimensional, it may be necessary to use summary statistics to reduce the dimensionality. Pritchard *et al.* suggested the prototype rejection-ABC algorithm as follows in a population genetics setting [39].

Algorithm 4: Rejection ABC method

Given the observation data data y_{obs}, prior distribution $\pi(\theta)$, summary statistics $s(\cdot)$, tolerance level $\varepsilon > 0$, and distance function $\rho(\cdot,\cdot)$

1. Generate a sample from the prior distribution $\theta' \sim \pi(\theta)$.
2. Simulate the model using θ' to get a dataset $x \sim p(x|\theta')$.
3. Calculate the distance between the simulation and experimental data $\rho(s(x), s_{obs})$ based on the given summary statistics $s(\cdot)$.
4. Accept the sample θ' if
$$\rho(s(x), s_{obs}) < \varepsilon.$$
Otherwise reject the sample.
5. Repeat steps $1\sim 4$ until the required number of posterior samples is obtained

The basic idea of ABC is to use summary statistics with a small tolerance to produce a good proximation of the posterior distribution. The output is the samples of parameters from the distribution $p(\theta|\rho(s(x), s_{obs}) \leq \varepsilon)$. The choice of summary statistics is very important which we will discuss later. In addition, the tolerance ε in

Algorithm 4 may determine the efficiency of ABC. The basic ABC rejection algorithm may result in long computing time when a prior distribution is far away from posterior distribution. In addition, there is no learning process in this algorithm; and thus no information could be obtained from the previous accepted samples of parameters. When the search space is complex, the convergence rate of this algorithm may be very slow.

4 Regression ABC

To improve the efficiency of the rejection-ABC algorithm, Beaumont *et al.* [4] introduced the regression approach by explicitly modeling the discrepancy between the simulated summary statistics and that of the observed data through the following algorithm.

Algorithm 5: Regression ABC

Given the observation data y_{obs}, prior distribution $\pi(\theta)$, summary statistics $s(\cdot)$, tolerance level ε, and distance function $\rho(\cdot)$.

1. Generate a sample from the prior distribution $\theta^{(i)} \sim \pi(\theta)$.
2. Simulate the model using $\theta^{(i)}$ to get a dataset $x^{(i)} \sim p(x|\theta^{(i)})$ and compute the summary statistics $s^{(i)} = s(x^{(i)})$.
3. Repeat steps 1 and 2, until N pairs $\{\theta^{(i)}, s^{(i)}\}$ are obtained.
4. Associate each pair $(\theta^{(i)}, s^{(i)})$ with a weight $\omega^{(i)} \propto K_\varepsilon(\rho(s^{(i)} - s_{obs}))$. The weighted kernel can be selected as:

$$K_\varepsilon(t) = \begin{cases} \varepsilon^{-1}(1-(t/\varepsilon)^2) & t \le \varepsilon, \\ 0 & t > \varepsilon. \end{cases}$$

5. Apply a regression model to the n points, which have nonzero weights to obtain an estimate of $E(\theta|s(x) = s^{(i)})$, denoted as $\hat{m}(s^{(i)})$.
6. Adjust each sample to

$$\theta^{*(i)} = \hat{m}(s_{obs}) + (\theta^{(i)} - \hat{m}(s^{(i)})).$$

7. Use $\{\theta^{*(i)}, \omega^{(i)}\}$ to aproximate the posterior distribution.

Here the samples $\theta^{(i)}$ are adjusted with weights $\omega^{(i)} > 0$ to account for the difference between simulated summary statistics and that of the observed data. Beaumont *et al.* [4] suggested a local linear model in the region of s_{obs}, given by

$$\begin{aligned} \theta^{(i)} &= m(s^{(i)}) + e^{(i)}, \\ m(s^{(i)}) &= \alpha + \beta^T (s^{(i)} - s_{obs}), \end{aligned}$$

where $e^{(i)}$ are zero-mean random variates with common variance, $m(s)$ is the conditional expectation of θ given s.

In this approach, the choice of ε involves a bias-variance trade-off, namely the increase of ε will reduce the variance because of a larger sample size for fitting the regression. However, this will also increases bias arising from the departure from the linearity and homoscedasticity [8].

When the number of samples is not very large due to the computational constraints, the homoscedastic assumption is no longer valid, because the neighbourhood of samples where $\omega^{(i)} \neq 0$ is too large. Thus Blum *et al.* [9] extended this algorithm to a nonlinear and heteroscedastic model, given by

$$\theta^{(i)} = m(s^{(i)}) + \sigma(s^{(i)})e^{(i)},$$

where $\sigma(s^{(i)}) = Var(\theta|s^{(i)})$ denotes the conditional variance. The variance is then estimated by using a second regression model for the logarithm of the squared residuals, given by

$$log(\theta^{(i)} - \hat{m}(s^{(i)}))^2 = log(\sigma(s^{(i)})) + \eta^{(i)},$$

where $\eta^{(i)}$ are independent, zero-mean variates with common variance. The parameter adjustment then can be performed as follows:

$$\theta^{*(i)} = \hat{m}(s_{obs}) + (\theta^{(i)} - \hat{m}(s^{(i)})) \times \frac{\hat{\sigma}(s_{obs})}{\hat{\sigma}(s^{(i)})}, \tag{4}$$

where $\hat{\sigma}(s)$ denotes the estimator of $\sigma(s)$. Here e plays the same role as for homoscedastic model, but it has more flexibility on deviations from homoscedasticty.

5 MCMC-ABC algorithm

In the Rejection-ABC and Regression-ABC algorithms, parameter values are sampled from the prior distribution. Thus the acceptance rate may be low if the prior and posterior distributions are quite different. In fact, using samples from a non-informative prior is very inefficient because this scheme does not account for the data at the proposal stage and thus may lead to proposed values located in low posterior probability regions. To address this issue, Marjoram *et al.* [33] introduced the following MCMC-ABC algorithm.

Algorithm 6: MCMC-ABC algorithm

Given the observation data y_{obs}, summary statistics $s(\cdot)$, tolerance level ε, distance function $\rho(\cdot)$, and proposal distribution $q(\cdot)$.

Initialize the first sample from the prior distribution $\theta^{(0)} \sim \pi(\theta)$.

At iteration $i \geq 0$

1. Generate a sample from the proposal distribution $\theta^{'} \sim q(\theta|\theta^{(i)})$.
2. Simulate the model using $\theta^{'}$ to get a dataset $x \sim p(x|\theta^{'})$.
3. Draw a sample from the uniform distribution $\mu \sim U(0,1)$, and calculate the ratio
$$\alpha = \min(1, \frac{\pi(\theta^{'})q(\theta^{(i)}|\theta^{'})}{\pi(\theta^{(i)})q(\theta^{'}|\theta^{(i)})} \times I(\rho(s(x), s_{obs}) \leq \varepsilon)).$$
Here $I(A)$ is an indicator function.
4. If $\mu \leq \alpha$, accept the sample $\theta^{(i+1)} = \theta^{'}$; otherwise $\theta^{(i+1)} = \theta^{(i)}$.
5. Repeat steps $1\sim 4$ until the required number of posterior samples is obtained.

This algorithm has a similar structure as that of the standard MCMC. Both algorithms use a proposal distribution and prior distribution to calculate the ratio. The difference is that the density function is used in MCMC for computing the ratio, while in MCMC-ABC we treat the ratio of density function as one if the simulation error satisfies the criterion. Thus the performance of MCMC-ABC strongly depends on the selection of proposal distribution and prior distribution.

A potential drawback of MCMC-ABC is the selection of tolerance level ε and proposal distribution $q(\theta|\theta^{(i)})$ that may lead to expensive pilot runs [47, 26]. The convergence property of the generated chain $(\theta^{(1)}, \cdots, \theta^{(n)})$ is important because MCMC algorithm may suffer if the proposal distribution is poorly chosen [14]. A potential issue is that the chain may get stuck in a low probability region of the posterior and lead to a poor approximation [18]. Since the proposed sample $\theta^{'}$ must meet two criteria, the rejection rate of the MCMC ABC may be extremely high.

6 SMC ABC

To tackle the challenges in MCMC-ABC, sequential Monte Carlo sampling techniques have been introduced to ABC. Sequential Monte Carlo sampling differs from the MCMC approach by using the technique of particle filtering. Rather than drawing one candidate sample $\theta^{'}$ at a step, this algorithm considers a pool with a large number of samples $(\theta_1^{'}, \cdots, \theta_N^{'})$ simultaneously and treats each sample as a particle. Sisson *et al.* [45] proposed a method which embed ABC simulation steps in Sequential Monte Carlo algorithm based on the theoretical work in [15]. This method generates sample from a sequence of approximate ABC posteriors under successively smaller acceptance tolerances[5, 46, 51]. SMC-ABC concentrates on simulating a dataset from the parameter regions with relatively high acceptance probabilities and can adapt tuning choices such as acceptance tolerances during the computation, which has potential advantages over the Rejection-ABC or MCMC-ABC. Here we illustrate the algorithm of Beaumont *et al.* [5]:

Algorithm 7: SMC-ABC

Given the observation data y_{obs}, summary statistics $s(\cdot)$, distance function $\rho(\cdot)$, tolerance thresholds $\varepsilon_1 \geq \cdots \geq \varepsilon_T$, and density kernel $K(\cdot)$.

1. At iteration $t = 1$,
   a. For $i = 1, \ldots, N$, repeat:
      i. Sample sample $\theta_i^{(1)} \sim \pi(\theta)$, and simulate dataset $x \sim p(x|\theta_i^{(1)})$.
      ii. Accept $\theta_i^{(1)}$ if $\rho(s(x), s_{obs}) \leq \varepsilon_1$, otherwise reject this sample.
      iii. Set weight $\omega_i^{(1)} = 1/N$.
   b. Take τ_2^2 as twice the empirical variance of the $\theta_i^{(1)}$s
2. At iteration $2 \leq t \leq T$
   a. For $i = 1, \ldots, N$, repeat:
      i. Pick θ_i^* from $\theta_j^{(t-1)}$s with probabilities $\omega_j^{(t-1)}$
      ii. Generate sample $\theta_i^{(t)} \sim K(\theta|\theta_i^*, \tau_t^2)$, and simulate a dataset $x \sim p(x|\theta_i^{(t)})$.
      iii. Accept $\theta_i^{(t)}$ if $\rho(s(x), s_{obs}) \leq \varepsilon_t$, otherwise reject this sample.
      iv. Set the weight of this accepted particle as

$$\omega_i^{(t)} \propto \frac{\pi(\theta_i^{(t)})}{\sum_{j=1}^{N} \omega_j^{(t-1)} K(\theta_i^{(t)}|\theta_j^{(t-1)}, \tau_t^2)}.$$

   b. Take τ_{t+1}^2 as twice the weighted empirical variance of the $\theta_i^{(t)}$s.

At the first iteration, this algorithm draws samples from the prior distribution $\pi(\theta)$, simulates the model using the sample, calculate summary statistics, and select N samples that satisfy the error criterion. This step actually is the rejection-ABC algorithm. However, at the subsequent iterations, samples are drawn from a density kernel $K(\theta)$ based on the previous particle population. A Gaussian kernel is used in Beaumont *et al.* [5], given by

$$K(\theta_i^{(t)}|\theta_j^{(t-1)}, \tau_t^2) = \varphi\{\tau_t^{-1}(\theta_i^{(t)} - \theta_j^{(t-1)})\},$$

where $\varphi(\cdot)$ is the density of a normal distribution. This algorithm effectively performs the repeated importance sampling technique, which is also known as population Monte Carlo [12]. Similar algorithms have been proposed by using different formulas to calculate the weights and different kernel functions [5, 46, 51].

SMC-ABC has addressed a potential drawback of the rejection and regression approaches. If the data are informative, the posterior distribution may be very narrow compared with the prior, then the rejection and regression algorithms may become inefficient. Thus repeatedly sampling from a gradually improving approximation of the posterior will make the distribution of summary statistics become closer to the posterior distribution, and increase the density of samples whose summary statistics is located in the vicinity of the target [6].

7 Choice of Summary Statistics

As discussed in previous sections, the posterior distribution of dataset $p(\theta|y_{obs})$ is approximated by

$$p(\theta|s_{obs}) \propto p(s_{obs}|\theta)\pi(\theta),$$

where s_{obs} is the summary statistics which usually has lower dimension than that of the data y_{obs}. If s_{obs} is sufficient,

$$p(\theta|s_{obs}) = p(\theta|y_{obs}).$$

When s_{obs} is highly informative, $p(\theta|s_{obs}) \approx p(\theta|y_{obs})$ is a good approximation. However, for many practical problems, it is hard to derive sufficient statistics or even a highly informative statistics. An appropriate choice of summary statistics is required to balance the informativeness and low-dimensionality. In some application fields, there has been a history of the development of summary statistics within a model-based framework in recent years. However, it is also possible that empirical summaries can be used without any strong theory to support them. Thus the selection of informative summary statistics is one of the important steps in the application of ABC. In recent years a number of methods have been proposed regarding the selection of summary statistics [19, 38].

Joyce and Marjoram [24] first proposed the ε-sufficiency concept and score of statistics for selecting an additional summary statistic s_k from the candidate set, when the model already has summary statistics $s_1, \ldots, s_{k-1}$. Later three methods regarding the choice of summary statistics have been used in application [31], namely

1. selection of a subset of the summary statistics that maximizes prespecified criteria such as the Akaike Information Criterion [10] or the entropy of a distribution [35];
2. partial least square regression to get linear combinations of the original summary statistics that are maximally decorrelated and highly correlated with the parameters [54]; and
3. summary statistics are chosen by minimizing a loss function under the assumption of a statistical model between parameters and transformed statistics of simulated data [1, 19].

Blum *et al.* [10] provided a comprehensive review of the principal methods. However, this topic still remains as a challenging problem in Bayesian inference.

8 Early rejection ABC

To reduce the simulation time, a number of inference methods have been proposed based on the idea of early rejection. For example, the delayed ABC divides a method into two stages [13]. In the first stage, a sample of parameters may be rejected or accepted by using an approximated posterior distribution. If it is accepted, a standard

ABC method will be applied in the second stage to evaluate the discrepancy between the observation data and simulation. This idea has been used in the MCMC-ABC for inferring stochastic differential equation models, in which the prior distribution and proposal distribution are used in the first stage for early rejection [41]. Based on the MCMC-ABC [41], a sample is rejected if the following ratio is less than a sample $\omega \sim U(0,1)$ by using the same notations in Eq. (3)

$$\omega > \frac{\pi(\theta^*)\pi(\theta_i|\theta^*)}{\pi(\theta_r)\pi(\theta^*|\theta_i)}. \tag{5}$$

In this approach, the kernel density function $p(y|\theta)$ is removed from the ratio above. Thus the performance of this early-rejection technique is fully dependent on the choice of the proposal density function $\pi(\theta^*|\theta_i)$.

A recently published approach is the Lazy ABC, which proposes a random stopping rule to abandon simulations with unsatisfactory accuracy [40]. This method makes ABC more scalable to applications where simulation is expensive. The detailed algorithm is given below

Algorithm 8: Lazy ABC

Input: prior density $\pi(\theta)$ and importance density $g(\theta)$, observation data y_{obs}, summary statistics $s(\cdot)$, tolerance level ε, distance function $\rho(\cdot,\cdot)$, proposal distribution $q(\cdot)$, and a continuous probability function $\alpha(\theta,x)$.

At iteration $i = 1 : N$

1. Generate a sample from importance sampling $\theta^* \sim g(\theta)$.
2. Simulate the model to get a dataset $x^* \sim p(x|\theta^*)$ and let $\alpha^* = \alpha(\theta^*,x^*)$.
3. With probability α^* continue to step 4. Otherwise perform early rejection: namely let $l^* = 0$ and go to step 6.
4. Simulate the model to get dataset $Y^* \sim p(Y|\theta^*,x^*)$.
5. Set $l^*_{ABC} = \mathbb{1}[d(s(y^*),s(y_{obs})) < \varepsilon]$ and $l^* = l^*_{ABC}/\alpha^*$.
6. Set $w^* = l^*\pi(\theta^*)/g(\theta^*)$.
7. Repeat steps $1\sim 6$ until the required number of posterior samples is obtained.

Output: A set of N pairs of (θ^*, w^*) values.

The detailed information of Lazy importance sampling and multiple stopping decision can be found in [40].

9 ABC software packages

A number of computer software packages have been designed in recent years to implement ABC in different platforms using various computer languages. A software package, BioBayes, provides a framework for Bayesian parameter estimation

and evidential model ranking over models of biochemical systems using ordinary differential equations. This package is extensible allowing additional modules to be included [53]. A Python package, ABC-SysBio, implements parameter inference and model selection for dynamical systems in the ABC framework [29]. This package combines three algorithms: ABC rejection sampler, SMC ABC for parameter inference, and SMC ABC for model selection. It is designed to work with models written in Systems Biology Markup Language (SBML). Deterministic and stochastic models can be analyzed in ABC-SysBio. In addition, a computational tool SYSBIONS has been designed for model selection and parameter inference using nested sampling [23]. Using a data-based likelihood function, this package calculates the evidence of a model and the corresponding posterior parameter distribution. This is a C-based, GPU-accelerated implementation of nested sampling that is designed for biological applications.

Also in the R platform, a number of software packages have been designed. Among them, package abc implements Rejection ABC with many methods of regression post-processing; while EasyABC implements a wide suite of ABC algorithms but not post-processing [36]. Package abctools has been designed to complement the existing software provision of ABC algorithms by focusing on tools for tuning them. It implements many previous unavailable methods from literature and makes them easy available to the research community [36]. In addition, there are also two ABC packages implemented as MATLAB toolbox. EP-ABC has been designed for state space models and related models, and ABC-SDE for inferring parameters in stochastic differential equations [37]. There are still some other software packages that have been reviewed in [36], including ABCreg, ABCtoolbox, Bayes SSC, DIY-ABC, and PopABC.

10 Conclusion

In this chapter, we have reviewed a number of algorithms of ABC, together with the relevant improvements, from choice of summary statistics to early rejection, aiming at increasing the statistical accuracy and computational efficiency. In addition, we give a few of the widely used software packages for the practical use of ABC algorithms. In recently years, the ABC methods have been applied to a wide range of inference problems in biology, economics, engineering and physical sciences. These applications have also raised more challenging questions for parameter inference, such as high-dimensional data [34] [28] and stochastic modeling [56], which provides interesting topics for future research.

References

1. Aeschbacher, S., Beaumont, M. A., Futschik, A.: A novel approach for choosing summary statistics in approximate Bayesian computation. Genetics. **192**, 1027-1047 (2012)
2. Barthelmé, S., Chopin, N.: Expectation propagation for likelihood-free inference. J. Am. Stat. Assoc. **109**, 315-333 (2014)
3. Bazin, E., Dawson, K. J., Beaumont,M. A.: Likelihoodfree inference of population structure and local adaptation in a Bayesian hierarchical model. Genetics. **185**, 587-602 (2010)
4. Beaumont, M. A., Zhang, W., Balding, D. J.: Approximate Bayesian computation in population genetics. Genetics. **162**, 2025-2035 (2002)
5. Beaumont, M. A.: Adaptive approximate Bayesian computation. Biometrika. **96**, 983-990 (2009)
6. Beaumont, M. A.: Approximate Bayesian computation in evolution and ecology. Annual Review of Ecology, Evolution and Systematics. **41**, 379-406 (2010)
7. Biau G., Crou F., Guyader A.: New insights into approximate Bayesian computation. Ann. I. H. Poincar B. **51**, 376-403 (2015)
8. Blum, M. G. B.: Approximate Bayesian computation: A nonparametric perspective. J. Amer. Statist. Assoc. **105**, 1178-1187 (2010)
9. Blum, M. G. B., François, O.: Non-linear regression models for Approximate Bayesian computation. Stat. Comput. **20**, 63-73 (2010)
10. Blum, M. G. B., Nunes, M. A., Prangle, D., Sisson, S. A.: A comparative review of dimension reduction methods in Approximate Bayesian computation. Statistical Science. **28**(2), 189-208 (2013)
11. Blum, M. G. B.: Regression approaches for Approximate Bayesian computation. arXiv:1707.01254v1 (2017)
12. Cappé, O., Guillin, A., Marin, J-M., Robert, C.P.: Population Monte Carlo. J. Comput. Graph. Stat., **13**(4), 907-929 (2004)
13. Christen, J. A. and Fox, C.: Markov chain Monte Carlo Using an Approximation. Journal of Computational and Graphical Statistics, **14**, 795-810 (2005)
14. Csilléry, K., Blum, M.G. B., Gaggiotti, O., Franois, O.: Approximate Bayesian computation (ABC) in practice. Trends Ecol. Evol. **25**(7), 410-418 (2010)
15. Del Moral, P., Doucet, A., Jasra, A.: Sequential Monte Carlo samplers. J. R. Stat. Soc. Ser. B. **68**, 411-436 (2006)
16. Del Moral, P., Doucet, A, Jasra, A.: An adaptive sequential Monte Carlo method for approximate Bayesian computation. Stat Comput. **22**(5),1009-1020 (2012)
17. Deng, Z., Tian, T.: A continuous optimization approach for inferring parameters in mathematical models of regulatory networks. BMC Bioinf. **15**, 256 (2014)
18. Drovandi, C. C., Pettitt, A. N.: Estimation of parameters for macroparasite population evolution using approximate Bayesian computation. Biometrics. **67**(1), 225-233 (2011)
19. Fearnhead, P., Prangle, D.: Constructing summary statistics for approximate Bayesian computation: semi-automatic approximate Bayesian computation. J. R. Stat. Soc. Ser. B. **74**, 419-474 (2012)
20. Gelman, A., Carlin, J. B., Stern, H. S., Rubin, D. B.: Bayesian Data Analysis. Chapman and Hall/CRC Press. London. (2003)
21. Goel, G., Chou, I.C., Voit, E.O.: System estimation from metabolic time-series data. Bioinformatics, **24**(21),2505-2511 (2008)
22. Green, P. J., Łatuszyński, K., Pereyra, M. , Robert, C. P.: Bayesian computation: a summary of the current state, and samples backwards and forwards. Stat. Comput. **25**, 835-862 (2015)
23. Johnson, R., Kirk, P., Stumpf, MPH. SYSBIONS: nested sampling for systems biology. Bioinformatics. **31**(4), 604-605 (2015)
24. Joyce, P., Marjoram, P.: Approximately sufficient statistics and Bayesian computation. Statistical Applications in Genetics and Molecular Biology. **7**(1), Article 26 (2008)
25. Kousathanas, A., Leuenberger, C., Helfer, J., Quinodoz, M., Foll,M., Wegmann, D.:Likelihood-Free Inference in High-Dimensional Models. Genetics. **203**, 893-904 (2016)

26. Kypraios, T., Neal, P. and Prangle, D.: A tutorial introduction to Bayesian inference for stochastic epidemic models using Approximate Bayesian computation. Mathematical Biosciences. **000**, 1-12 (2016)
27. Lenormand, M., Jabot, F., Deffuant, G.: Adaptive approximate Bayesian computation for complex models. Comput Stat. **28**(6), 2777-2796 (2013)
28. Li, J., Nott, D. J., Fan, Y., Sisson, S. A.: Extending approximate Bayesian computation methods to high dimensions via a Gaussian copula model. Computational Statistics and Data Analysis. **106**, 77-89 (2017)
29. Liepe, J., et al.: A framework for parameter estimation and model selection from experimental data in systems biology using approximate Bayesian computation. Nat Protoc. **9**(2):439-456 (2014)
30. Lillacci, G., Khammash, M.: Parameter estimation and model selection in computational biology. PLoS Comput Biol. **6**(3), e1000696 (2010)
31. Lintusaari, J., Gutmann, M., Dutta, R., Kaski, S., Corander, J.: Fundamentals and Recent Developments in Approximate Bayesian Computation. Syst. Biol. **66**(1), e66-e82 (2017)
32. Marin, J.-M., Pudlo, P., Robert, C. P., Ryder, R. J.: Approximate Bayesian computational methods, Stat. Comput. **22**, 1167-1180 (2012)
33. Marjoram, P., Molitor, J., Plagnol, V., Tavaré, S.: Markov chain Monte Carlo without likelihoods. Proceedings of the National Academy of Sciences of the United States of America. **100**(26),15324-15328 (2003)
34. Nott, D. J., Fan, Y., Marshall, L., Sisson, S. A.: Approximate Bayesian computation and Bayes's linear analysis: toward high-dimensional ABC. Journal of Computational and Graphical Statistics. **23**(1),65-86 (2014)
35. Nunes, M.A., Balding, D.J.: On optimal selection of summary statistics for approximate Bayesian computation. Stat. Appl. Genet.Mol. Bio. **9**, Article 34(2010)
36. Nunes M. A., Prangle D.: abctools: an R package for tuning approximate Bayesian computation analyses. R J. **7**(2), 189-205 (2015)
37. Picchini, U.: Inference for SDE models via approximate Bayesian computation. J. Comput. Graph. Stat. **23**(4), 1080-1100 (2014)
38. Prangle, D.: Summary statistics in Approximate Bayesian computation. arXiv:1512.05633 (2015)
39. Pritchard, J.K., Seielstad, M. T., Perez-Lezaun, A., Feldman, M. W.: Population growth of human Y chromosomes: a study of Y chromosome microsatellites. Mol. Biol. Evol. **16**(12), 1791-1798 (1999)
40. Prangle D.: Lazy ABC. Statistics and Computing. **26**, 171-185 (2016)
41. Picchini U.: Inference for SDE models via approximate Bayesian computation. J Comput Graph Stat. **23**, 1080-1100 (2014)
42. Robert, C. P., Casella, G.: Monte Carlo Statisical Methods. Springer. New York. (2004)
43. Robert, C. P.: Approximate Bayesian Computation: a survey on recent results. Monte Carlo and Quasi-Monte Carlo Methods. Springer. pp 185-205 (2016)
44. Rubin, D. B.: Bayesianly justifiable and relevant frequency calculations for the applies statistics. Annals of Statistics. **12**(4), 1151-1172 (1984)
45. Sisson, S. A., Fan, Y., Tanaka, M. M.: Sequential Monte Carlo without likelihoods. Proceedings of the National Academy of Sciences of the United States of America. **104**(6),1760-1765 (2007)
46. Sisson, S. A., Fan, Y., Tanaka, M. M.: Sequential Monte Carlo without likelihoods. Proceedings of the National Academy of Sciences of the United States of America. **106**(39),16889-16889 (2009)
47. Sisson, S. A., Fan, Y.: Likelihood-free MCMC. Handbook of Markov Chain Monte Carlo. CRC Press. 313-335 (2011)
48. Sunnaker, M., et al.: Approximate Bayesian computation. PLoS Comput Biol. **9**(1), e1002803 (2013)
49. Tavaré, S.,Balding,D., Griffith, R., Donnelly, P.: Inferring coalescence times from DNA sequence data. Genetics. **145**, 505-518 (1997)

50. Tian, T., Smith-Miles, K.: Mathematical modeling of GATA-switching for regulating the differentiation of hematopoietic stem cell. BMC Syst Biol. **8** Suppl 1:S8 (2014)
51. Toni, T., Welch, D., Strelkowa, N., Ipsen, A., Stumpf, M. P. H.: Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. J. R. Soc. Interface. **6**, 187-202 (2009)
52. Turner, B. M., Van Zandt, T.: A tutorial on approximate Bayesian computation. J Math Psychol. **56**(2):69-85 (2012)
53. Vyshemirsky, V., Girolami, M.: BioBayes: a software package for Bayesian inference in systems biology. Bioinformatics. **24**(17), 1933-1934 (2008)
54. Wegmann, D., Leuenberger, C., Excoffier, L.: Efficient approximate Bayesian computation coupled with Markov chain Monte Carlo without likelihood. Genetics. **182**,129-141 (2009)
55. Wilkinson, D.J.: Bayesian methods in bioinformatics and computational systems biology. Brief Bioinform. **8**(2):109-116 (2007)
56. Wu Q., Smith-Miles, K., Tian, T.: Approximate Bayesian computation schemes for parameter inference of discrete stochastic models using simulated likelihood density. BMC Bioinf.15:S3. (2014)