OA Genetics
Open Access

Bioinformatics

Critical review

Approximation Bayesian computation

P Marjoram*

Abstract

Introduction

Approximate Bayesian computation is an analysis approach that has arisen in response to the recent trend to collect data of very high dimension. This has led to many existing methods become intractable because of difficulties in calculating the likelihood function. Approximate Bayesian computation circumvents this issue by replacing calculation of the likelihood with a simulation step in which it is estimated in one way or another. In this review, we give an overview of the approximate Bayesian computation approach, by giving examples of some of the more popular specific forms of approximate Bayesian computation. We then discuss some of the areas of most active research and application in the field, specifically, choice of low-dimensional summaries of complex datasets and metrics for measuring similarity between observed and simulated data. Next, we consider the question of how to do model selection in an approximate Bayesian computation context. Finally, we discuss an area of growing prominence in the approximate Bayesian computation world, use of approximate Bayesian computation methods in genetic pathway inference.

Conclusion

We expect the rise of approximate Bayesian computation methods to continue, and we hope this will include the continued development of theory and machinery to guide the user in making some of the key choices discussed above.

*Corresponding author
Email: pmarjora@usc.edu

Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, California, USA

Introduction

At a time in which our ability to collect data is growing at great rates, it is also the case that new challenges arise when attempting analysis of these data. Given data, D, a model, M, that attempts to explain the data, and a set of model parameters, θ, our analysis task often depends upon calculation of the likelihood, $f(D \mid \theta)$, either as a direct component of a frequentist analysis, or as a step towards calculating the posterior distribution $f(\theta \mid D)$ in the Bayesian paradigm (and our perspective in this article will be Bayesian). Using Bayes' theorem the posterior distribution is calculated as $f(\theta \mid D) \propto f(D \mid \theta)\pi(D)$, where $\pi(\cdot)$, the prior distribution, captures our beliefs about θ before the data is collected. However, as complexity or volume of data increases, calculation of the likelihood (and, therefore, also the posterior) often becomes impossible, either because it is computationally intractable or because closed-form expressions are not derivable. This conflict has led to the rise of an alternative approach called approximate Bayesian computation (ABC).

ABC methods borrow intuition from likelihood estimation, introduced by Diggle and Gratton[1]. There, large-scale Monte Carlo simulation is used to directly approximate the likelihood of D given θ (and all expressions here are implicitly also dependent upon the model M) as the proportion of times in which simulation of data, D′, using parameter θ, results in D′ = D. However, as data complexity grows, the probability of observing D′ = D typically becomes vanishingly small, even when the correct value of θ is used. This has led to the appearance of ABC versions of rejection methods. This review discusses the approximate Bayesian computational approach.

Discussion

The author has referenced some of his own studies in this review. The protocols of these studies have been approved by the relevant ethics committees related to the institution in which they were performed.

ABC rejection methods

The simplest form of ABC, that based on rejection methods, supposes the existence of a set of summary statistics, S, that capture key features of D, and adopts the intuition of likelihood approximation within the following algorithm:

- Set, i = 0.
- Sample θ′ from the prior π().
- Simulate data D′ using model M with parameter θ′. Calculate a set of summary statistics S′ from D′.
- If S′ = S accept θ′.
- Set i = i + 1. If i < N, go to 1; else stop.

Here, N is a predetermined large number. The resulting set of accepted θ-values form a sample from the distribution $f(\theta \mid S)$. In the best case scenario in which the set of statistics S are sufficient for θ, we have (by definition of sufficiency) $f(\theta \mid S) = f(\theta \mid D)$. However, in most contexts exact matching even of summary statistics is relatively unlikely, in which case we introduce a distance measure, $d(\cdot, \cdot)$, a tolerance threshold ε, and we replace step 4 above with

4′ If $d(S', S) < \varepsilon$ accept θ′.

Now we obtain independent samples from a distribution that we will call $\varphi(\theta \mid S)$. One of the important

FOR CITATION PURPOSES: *Marjoram P.* Approximation Bayesian computation. OA Genetics 2013 May 01;1(1):3.

Competing interests: none declared. Conflict of interests: none declared.
All authors contributed to the conception, design, and preparation of the manuscript, as well as read and approved the final manuscript.
All authors abide by the Association for Medical Ethics (AME) ethical rules of disclosure.

OA Genetics
Open Access

Critical review

caveats of an ABC analysis is that, in general, it is not possible to state the degree of agreement between the distribution one wanted to calculate, $f(\theta \mid D)$, and the distribution from which one obtains a sample, $\varphi(\theta \mid S)$. This is currently often assessed via simulation study, but is an area of active research in the ABC community.

Rejection methods work well provided there is good overlap between prior and posterior parameter distributions. However, when this is not the case efficiency is low since much time is wasted sampling potential θ-values from parts of the prior distribution that are poorly supported by the posterior. Problems also arise when the dimension of the parameter spaces is large. For this reason, a number of other methods have arisen, previous to the existence of ABC, that are more efficient. Many of these ideas have now been adapted into the ABC context. An early example is the adoption of Metropolis–Hastings Markov chain Monte Carlo (MCMC)[2,3], into what has become known as ABC-MCMC (or the 'no-likelihoods' MCMC method).

ABC-MCMC

The ABC-MCMC algorithm[4], starts from an arbitrarily chosen θ-value and proceeds as follows:

1. If now at θ, propose a move to θ′ according to a proposal distribution $q(\theta \rightarrow \theta')$.
2. Simulate a dataset, D′ using θ′.
3. If $D' \sim D$ proceed to 4; else, output θ and return to 1.
4. Calculate the Hastings Ratio (HR):
$$h = \min\{1, \pi(\theta')q(\theta' \rightarrow \theta)/\pi(\theta)q(\theta \rightarrow \theta')\},$$
5. Accept, and output, the new θ′ with probability h. Else, return to, and output, θ. Go to 1.

Here, q() is a user-defined transitional kernel that controls how we propose new θ-values. Once the chain of θ-values has reached stationarity, outputs from the chain have the required distribution. ABC-MCMC differs from traditional MCMC in that the calculation of the ratio of likelihoods for new and old parameter values has been replaced by a step in which we simulate a single dataset, D′, using θ′, and then proceed to calculate the rest of the HR only if $D' \sim D$. Thus, the intractable likelihood has again been replaced by a simulation step, thereby recovering tractability. However, ABC-MCMC has been shown to mix relatively poorly, compared to traditional MCMC, in the tails of the posterior. The reason for this is simple. In traditional MCMC, if we propose a θ′-value in the tail of $f(\theta \mid D)$ it will be the case that $f(D \mid \theta')$ is likely to be very small. However, provided the transitional kernel q() proposes small changes to θ when generating θ′, at least some of the time, it will also be the case that $f(D \mid \theta)$ will be small, and that the ratio $f(D \mid \theta')/f(D \mid \theta)$ will typically be of order 1. Thus, the HR will, all other things being equal, not take too small a value. This encourages good mixing. In ABC-MCMC, we have replaced the ratio of likelihoods term with the generation of a dataset for θ′ only. Thus, in the tails of the posterior for θ, the probability of generating a $D' \sim D$ may be vanishingly small, and is not countered by similar behaviour of $P(D \sim D' \mid \theta)$. There are several possible responses to this, if mixing becomes problematic. First, use a proposal kernel that sometimes proposes large changes to θ, thereby retaining the possibility of proposing θ′-values out of the tail of the posterior, whatever is the currently accepted value of θ. Second, Andrieu C, Roberts, have shown that we can run a generalised version of the ABC-MCMC algorithm in which we simulate data to approximate the likelihood of $(D \mid \theta)$ for both new and old parameter values[5]. However, it is important to note that when one estimates $f(D \mid \theta)$ in the denominator of the traditional MCMC HR this way, one must recycle the estimate that was used when accepting θ, rather than re-estimate it. Otherwise biases are introduced. We note in passing that the ABC-MCMC algorithm above can be viewed as a version of this latter approach in which we use a single dataset, and an indicator function that takes the value 1 if $D' \sim D$, as a crude estimate of $f(D' \sim D \mid \theta')$ and $f(D' \sim D \mid \theta)$.

An alternative response to these mixing issues results in another popular ABC algorithm, Sequential Monte Carlo ABC (ABC-SMC).

ABC-SMC

ABC-SMC uses a population of θ-values, rather than a single θ-value, at any given time[6,7]. While some of these may be in the tails of the posterior, others will likely not, thus improving mixing properties. The algorithm is a form of 'importance sampling'[8]. It iterates through T generations, proceeding as follows (we base our description on that of Secrier et al.[7]):

1. Define tolerances $\varepsilon_1, \ldots, \varepsilon_T$. Tolerance ε_t is used in generation t. Define the initial 'posterior' parameter distribution, f_1, to be equal to the prior distribution π. Set the population count to t = 1, and define a target number of acceptances per population, N.
   1.A. Set the particle indicator to i = 0.
   1.B. Sample a parameter-value, θ, from f. If t > 1 perturb the sampled parameter value (e.g., by adding a normal random variable).
   1.C. Simulate data $D_{t,i}$ using θ. If the distance between $D_{t,i}$ and the observed data is greater than εt return to step 1.B; otherwise, set i = i + 1, and calculate a 'weight' for the accepted parameter value θ. This weight is an 'importance sampling' weight that

FOR CITATION PURPOSES: *Marjoram P.* Approximation Bayesian computation. OA Genetics 2013 May 01;1(1):3.

Competing interests: none declared. Conflict of interests: none declared.
All authors contributed to the conception, design, and preparation of the manuscript, as well as read and approved the final manuscript.
All authors abide by the Association for Medical Ethics (AME) ethical rules of disclosure.

OA Genetics
Open Access

Critical review

corrects for the fact that θ was sampled from f_t rather than π.

1.D. If i < N go to 1.B; otherwise construct a new 'posterior' distribution, f_{i+1}, from the set of weights of accepted parameter values.

2. If t < T , set t = t + 1 and go to 1.A.

We have omitted many of the technical details, but the intuition is that the algorithm performs a rejection method in which, rather than sampling from the prior, we sample from an importance sampling distribution formed from the posterior distribution calculated in the previous 'generation', but adding noise to sampled parameter values to allow the generation of new values. As such, it is a form of importance sampling in which the importance sampling distribution changes over time. The algorithm has now been used in a number of applications[7,9,10,11], and is implemented in the ABC-SysBio package[12].

ABC—Data summaries and match tolerance

A number of decisions need to be made when performing an ABC analysis. Principal among them, perhaps, is the needs to measure the match between observed and simulated data. This is often achieved through the adoption of a set of summary statistics that are designed to capture key features of the data. In the early days of ABC, these were often chosen using 'investigator intuition'. More recently a number of studies have appeared in which more principled methods are proposed. Joyce and Marjoram[13] developed a sequential scheme for scoring statistics according to whether their use in the analysis substantially improved the quality of inference, as measured by changes to the posterior distribution (the addition of uninformative statistics should not be expected to substantially change the posterior distribution that results). Nunes et al.[14] proposed a similar scheme designed to minimise the average squared error of the posterior distribution. Fearnhead and Prangle[15] showed how to construct statistics in a semi-automatic manner. Jung and Marjoram[16] develop a method to choose both a subset of statistics and weights that should be applied to each statistic in the subsequent calculation of similarity with observed data.

In other related work, Beaumont et al.[17] discarded the concept of 'rejection' and instead included all simulated iterations in the estimation of the posterior for θ, but now weighting each iteration by the distance between observed and simulated statistic values after fitting a local linear regression of θ on S. Blum et al.[18] generalised this to use non-linear regression, using an importance sampling scheme to refine the fit. Wegmann et al.[19] aimed to reduce the dimensionality of the analysis, and thereby increase efficiency, by reducing the number of data-points considered in the analysis, and so raise the acceptance rate. One might hope to do this simply by calculating principal components of the values the data take over a large number of simulated datasets. However, principal components often perform rather poorly in ABC analyses, since they are designed to return orthogonal directions for which the variation in the data is greatest, whereas ABC performs best when projections of the data concisely capture variation in the parameters. Wegman's method uses a partial least squares approach to choose orthogonal axes that have maximum correlation with the parameters of interest. These axes are analogous to the results of a principal component analysis, but the partial least square approach ensures that the axes have good utility in predicting parameter values. In Wegman et al.'s[19] study, the method was applied to an analysis of time of divergence of two populations in an ABC-MCMC context.

In an alternative approach, Hamilton et al.[20] took an existing set of statistics and chose weights for them using a scheme in which large numbers of dataset were simulated, with only those that were similar to the observed data being retained. Using those data, regress the S_i on each parameter in θ in turn, recording the model-fit R^2 in each case. A set of weights are then calculated to measure the degree of informativeness of statistic i on parameter j. (In fact, rather than weighting the statistics directly, Hamilton defines a weighted Euclidean distance metric to measure the difference between observed and simulated statistic values, but the effect is the same.) The scheme was applied to an analysis of evolutionary parameters in models of human demography.

A number of these methods were compared by Barnes et al.[21], in which a further new, improved method was proposed (see below).

ABC and model selection

One of the most active areas of research in ABC is its application to model selection. Here, we suppose we are trying to decide between two models, M_1 and M_2, (the following generalises in an obvious way if there are more than two models). In a Bayesian paradigm, evidence for M_1 compared to M_2 is weighed in terms of the Bayes Factor, $BF = f(D \mid M_1)/f(D \mid M_2)$, the ratio of the posterior and priors odds in favour of M_1[22]. In an ABC context, it has been common to use an approximation to the BF, $BF_{ABC} = f(S \mid M_1)/f(S \mid M_2)$. Research in this area was perhaps provoked by a study of Templeton[23] that attacked an ABC analysis of Fagundes et al.[24] in which several possible models for early human evolution were compared. It was later shown that Rogers was in fact attacking

FOR CITATION PURPOSES: *Marjoram P.* Approximation Bayesian computation. OA Genetics 2013 May 01;1(1):3.

Competing interests: none declared. Conflict of interests: none declared.
All authors contributed to the conception, design, and preparation of the manuscript, as well as read and approved the final manuscript.
All authors abide by the Association for Medical Ethics (AME) ethical rules of disclosure.

OA Genetics
Open Access

the Bayesian method itself, rather than the ABC approach to Bayesian analysis[25], but a series of studies have subsequently emerged in which complications involved with ABC in a model selection context have been discussed. Fundamentally, the issue is that the ABC approximation to the Bayes Factor, BF_{ABC}, is related to the actual Bayes Factor, BF, in the following way: $BF = BF_{ABC} \times f(D \mid S,M_1)/f(D \mid S,M_2)$. However, it is not necessarily the case that $f(D \mid S,M_1)/f(D \mid S,M_1) = 1$. Most interestingly, as pointed out by Robert et al.[26], we do not necessarily have $f(D \mid S,M_1)/f(D \mid S,M_2) = 1$, even when the statistics S are sufficient for parameter estimation in M_1 and M_2 individually. Robert et al.[26] give a nice example in which count data might arise from Poisson or Geometric distributions. They show that the ratio $f(D \mid S,M_1)/f(D \mid S,M_2)$ is not equal to 1 even when S is formed from the union of statistics that are sufficient for inference in the two models separately.

There have been two responses to this issue. First, it has been noted that provided one works with the data, rather than summary statistics thereof, the problem is avoided. In this context, with a slight abuse of notation, the BF is approximated as $BF'_{ABC} = f(d(D',D) < \varepsilon \mid ABC\ M_1)/f(d(D',D) < \varepsilon \mid M_2)$, and we note that, as $\varepsilon \to 0$ we have $BF'_{ABC} \to BF$. Of course, as we have noted, choice of ε represents a compromise between accuracy and tractability, so achieving a result sufficiently close to the limiting behaviour may be practically difficult. The second approach introduces the concept of statistics that are sufficient for model selection (SFM), in the sense that if S is SFM, then $f(D \mid S,M_1)/f(D \mid S,M_2) = 1$. This was introduced by Barnes et al.[21], in a study in which they present an algorithm that attempts to choose a set of statistics that appear to be SFM, and which generalises and improves the methods of Joyce and Marjoram[13] for choosing approximately sufficient statistics in a non-model-selection context.

ABC and genetic networks

An area of growing application of ABC methods is that of inference of genetic networks. Here, the goal is to infer parameters of a known network relating expression of a set of genes, possibly related to some phenotypes of interest. Alternatively, we might wish to construct the network from scratch, aiming to infer which genes are involved and how they interact with each other. ABC methods are of interest here because as the complexity of networks grows, computational intractability becomes an issue (again either because exact solutions are impossible, or because networks contain genuinely stochastic components, or because numeric algorithms become too slow to perform well) (see Marjoram et al.[27] for an overview of this perspective).

The leading exponents of ABC in this field are the group of Stumpf et al., who have written a number of papers on the subject[e.g.,7,28] and have also produced a software package (ABC SysBio) that makes implementations of ABC methods relatively straightforward in this context, and which integrates with the widely used SysBio systems biology software package[12].

The ABC SysBio method is for analysis of known networks. A recent study by Rau et al.[29] addressed the issue of how to build networks from the ground up in an ABC context. Their method uses time-course data to test for linear relationships between pairs of genes, arguing that many networks can be well approximated using linear components. The complexity of the search space is kept manageable by supposing limits on the number of genes that can directly affect the behaviour of another gene.

It remains to be seen how popular applications of ABC analyses will become in the context of gene networks, but the growing view that such networks might be used to leverage the power of genome-wide association studies, suggests that there is a powerful need for methods that remain tractable for relatively complex networks.

Conclusion

In the modern era, we are collecting data that are bigger, generally by orders of magnitude, than data that were collected previously. This means that more detailed inference can be made, often using models that are more complex than before. A consequence of this is that standard statistical analysis methods often become intractable. There are two common responses to the intractability of the likelihood: (1) simplify the model so that the likelihood function can, once again, be calculated; or (2) add an approximation step to the analytic method itself. At this point, we recall a quote attributed to George Box: 'All models are wrong, but some are useful'. While approach (1) above is possible, it may lead to a model so divorced from reality that conclusions drawn from it cannot be considered particularly informative. The American statistician John Tukey said 'Far better an approximate answer to the right question, which is often vague, than an exact answer to the wrong question, which can always be made precise'. ABC methods embrace this spirit, allowing tractable analysis of large, modern datasets. Consequently, there is an increasing tendency for investigators to turn to ABC methods in answer to the challenges of analysis of modern data sets. As such, the rise of ABC has been rapid—from essentially no studies prior to 2000, to over a hundred per year most recently.

Competing interests: none declared. Conflict of interests: none declared.
All authors contributed to the conception, design, and preparation of the manuscript, as well as read and approved the final manuscript.
All authors abide by the Association for Medical Ethics (AME) ethical rules of disclosure.

OA Genetics
Open Access

In this review, we surveyed ABC methods and illustrated some of the key decisions that need to be made in an ABC analysis. We also pointed to areas of active research in the ABC community. We expect the rise of ABC methods to continue, and we hope this will include the continued development of theory and machinery to guide the user in making some of the key choices discussed above.

Acknowledgement

This study was funded by NSF award DMS 1101060 and NIH awards R01MH100879 and 1U01GM103804. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NSF or NIH.

References

1. Diggle PJ, Gratton RJ. Monte Carlo methods of inference for implicit statistical models. J Roy Statist Soc B. 1984;46:193–227.
2. Hastings WK. Monte Carlo sampling methods using Markov chains and their applications. Biometrika. 1970 Apr;57(1):97–109.
3. Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. Equations of state calculations by fast computing machines. J Chem Phys. 1953 Jun;21(6):1087–91.
4. Marjoram P, Molitor J, Plagnol V, Tavaŕe S. Markov chain Monte Carlo without likelihoods. Proc Natl Acad Sci U.S.A. 2003 Dec;100(26):15324–8.
5. Andrieu C, Roberts GO. The pseudo-marginal approach for efficient Monte Carlo computations. Ann Stat. 2009;37(2): 697–725.
6. Del Moral P, Doucet A, Jasra A. An adaptive sequential Monte Carlo method for approximate Bayesian computation. Stat Comput. 2011 May.1–12.
7. Secrier M, Toni T, Stumpf MPH. The ABC of reverse engineering biological signalling systems. Mol Bio Syst. 2009 Dec;5(12):1925–35.
8. Ripley BD. Stochastic simulation. New York: John Wiley & Sons, Inc.; 1982.
9. Del Moral P, Doucet A, Jasra A. Sequential Monte Carlo samplers. J R Statist Soc B. 2006 Jun;68(3):411–36.
10. Sisson SA, Fan Y, Tanaka MM. Sequential Monte Carlo without likelihoods. Proc Natl Acad Sci. 2007 Feb;104(6):1760–5.
11. Jasra A, Martin J, McCoy E, Singh SS. Filtering via approximate Bayesian computation. Technical report. Cambridge University Engineering Department. 2010.
12. Liepe J, Barnes C, Cule E, Erguler K, Kirk P, Toni T, et al. ABC-SysBio–approximate Bayesian computation in Python with GPU support. Bioinformatics. 2010 Jul;26(14):1797–9.
13. Joyce P, Marjoram P. Approximately sufficient statistics and Bayesian computation. Stat App Genet Mol Biol. 2008 Aug;7(1):Article26.
14. Nunes MA, Balding DJ. On optimal selection of summary statistics for Approximate Bayesian Computation. Stat Appl Genet Mol Biol. 2010;9:a34.
15. Fearnhead P, Prangle D. Semi-automatic approximate Bayesian computation. Arxiv. 2010 Apr;arXiv:1004.1112v1.
16. Jung H, Marjoram P. Choice of summary statistic weights in approximate Bayesian computation. Stat Appl Genet Mol Biol. 2011 Jan;10(1):Article 45.
17. Beaumont MA, Zhang W, Balding DJ. Approximate Bayesian computation in population genetics. Genetics. 2002 Dec;162(4):2025–35.
18. Blum MGB, Francois O. Non-linear regression models for approximate Bayesian computation. Stat Comput. 2010 Jan;20(1):60–73.
19. Wegmann D, Leuenberger C, Excoffier L. Efficient approximate Bayesian computation coupled with Markov chain Monte Carlo without likelihoods. Genetics. 2009 Aug;182(4):1207–18.
20. Hamilton G, Currat M, Ray N, Heckel G, Beaumont M, Excoffier L. Bayesian estimation of recent migration rates after a spatial expansion. Genetics. 2005 May;170(1):409–17.
21. Barnes CP, Filippi S, Stumpf MPH, Thorne T. Considerate approaches to constructing summary statistics for ABC model selection. Stat Comput. 2012 Jun;22(6):1181–97.
22. Kass R, Raftery A. Bayes factors. J Am Stat Assoc. 1995;90(430):773–95.
23. Templeton AR. Coherent and incoherent inference in phylogeography and human evolution. Proc Natl Acad Sci. 2010 Apr;107(14):6376–81.
24. Fagundes NJR, Ray N, Beaumont M, Neuenschwander S, Salzano FM, Bonatto SL, et al. Statistical evaluation of alternative models of human evolution. Proc Natl Acad Sci. 2007 Nov;104(45):17614–19.
25. Beaumont MA, Nielsen R, Robert C, Hey J, Gaggiotti O, Knowles L, et al. In defence of model-based inference in phylogeography. Mol Ecol. 2010 Feb;19(3): 436–46.
26. Robert CP, Cornuet J-M, Marin J-M, Pillai NS. Lack of confidence in approximate Bayesian computational (ABC) model choice. PNAS (Open Access). 2011 Sep;108(37):15112–17.
27. Marjoram P, Zubair A, Nuzhdin SV. Post-GWAS: where next? more samples, more SNPs or more biology? Heredity (Edinb). 2104 Jan;112(1):79–88.
28. Toni T, Welch D, Strelkowa N, Ipsen A, Stumpf MPH. Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. J R Soc Interface. 2009 Feb;6(31):187–202.
29. Rau A, Jaffr´ezic F, Foulley J-L, Doerge RW. Reverse engineering gene regulatory networks using approximate Bayesian computation. Stat Comput. 2012 Nov;22(6):1257–71.