Approximate Bayesian Computation in Evolution and Ecology

Mark A. Beaumont

Department of Mathematics and School of Biological Sciences, University of Bristol, Bristol BS8 1TNW, United Kingdom; email: m.beaumont@bristol.ac.uk

Annu. Rev. Ecol. Evol. Syst. 2010. 41:379–406

First published online as a Review in Advance on August 10, 2010

The *Annual Review of Ecology, Evolution, and Systematics* is online at ecolsys.annualreviews.org

This article's doi:
10.1146/annurev-ecolsys-102209-144621

1543-592X/10/1201-0379$20.00

Key Words

Bayesian inference, Monte Carlo, simulation

Abstract

In the past 10 years a statistical technique, approximate Bayesian computation (ABC), has been developed that can be used to infer parameters and choose between models in the complicated scenarios that are often considered in the environmental sciences. For example, based on gene sequence and microsatellite data, the method has been used to choose between competing models of human demographic history as well as to infer growth rates, times of divergence, and other parameters. The method fits naturally in the Bayesian inferential framework, and a brief overview is given of the key concepts. Three main approaches to ABC have been developed, and these are described and compared. Although the method arose in population genetics, ABC is increasingly used in other fields, including epidemiology, systems biology, ecology, and agent-based modeling, and many of these applications are briefly described.

1. INTRODUCTION

Models abound in biology. It is often easy to construct models that might plausibly describe our observations; it is also often straightforward to simulate artificial data sets for given parameters in the model. Yet, it is generally difficult to simulate the parameter values (or, indeed, models) that could have given rise to a given data set. This largely unremarked asymmetry lying at the heart of scientific enquiry has become more acute because our knowledge of complex and complicated systems has increased, and so too has our ability to simulate them on clusters of computers.

A naturally powerful, rich, and flexible framework within which to address these problems is provided by the Bayesian paradigm. Indeed, the notion of simulating parameter values only really makes sense in a Bayesian context. Until recently, however, Bayesian inference has required that we can compute the likelihood function—the probability of obtaining the observations given some parameter value. For many problems, although it may be straightforward to write a computer program to simulate data, it may actually be very difficult to work out the likelihood function. Typically, this arises if the model has many hidden states and the probability of the data depends on summing probabilities over all possible states. Examples would be an unobserved genealogy in population genetics, or an unobserved trajectory of census size in an ecological model, or the many hidden states in an agent-based model. Arising from population genetics, a group of techniques—often dubbed approximate Bayesian computation, or likelihood-free inference—has been developed that avoids the need for a likelihood function. It is now becoming clear that the scope of these methods is potentially much broader than in population genetics alone, and the aim of this review is to provide a description of the general approach with examples of applications in a number of different fields.

2. BAYESIAN INFERENCE AND APPROXIMATE BAYESIAN COMPUTATION

Broadly, Bayesian computation involves the estimation of a conditional probability density and functions thereof. Typically, the modeler defines the joint density of parameter values θ and data x, $p(x, \theta) = p(x \mid \theta)\pi(\theta)$, by specifying, respectively, in the right-hand side of the equation above, the likelihood function and the prior. The task is then to compute the posterior,

$$p(\theta \mid x) = \frac{p(x \mid \theta)\pi(\theta)}{p(x)},$$

which is typically difficult because the marginal likelihood,

$$p(x) = \int p(x \mid \theta)\pi(\theta)d\theta,$$

often involves a high dimensional integral. Provided the likelihood can be evaluated up to a normalizing constant, Monte Carlo methods such as Markov chain Monte Carlo (MCMC), importance sampling (IS), or sequential Monte Carlo (SMC) can be used (Gelman et al. 2003, Robert & Casella 2004). In this case, even though the posterior distribution is not available as a function that can be evaluated, samples can be taken from it. That is, a scheme has been devised for simulating parameter values given the data. As models become more complicated, the likelihood function becomes more difficult to define and compute, and it is easier to simulate data samples from the model given the value of a parameter (or parameter vector). Unfortunately, we then have no control over what data is simulated. If the data are discrete and of low dimension, then it is possible to sample from the posterior density of the parameter without an explicit likelihood function, and without approximation, by the following algorithm Rubin (1984):

Given observation y, repeat the following until N points have been accepted:

1. Draw $\theta_i \sim \pi(\theta)$.
2. Simulate $x_i \sim p(x \mid \theta_i)$.
3. Reject θ_i if $x_i \neq y$.

These are sampled from $p(\theta \mid y)$.

The posterior distribution, from which these points are sampled, gives the probability distribution of the parameter value that gave rise to the observations. A variety of summaries of this distribution can be obtained from the simulated points. For univariate or bivariate θ, it is possible to use kernel density methods (or simply draw a histogram) based on the points to visualize the distribution. For multivariate θ, marginal distributions are typically visualized and summarized. Given a kernel density, it is then possible to estimate the mode (corresponding to the maximum likelihood estimate when the prior is uniform). Alternatively the posterior mean, or median, can be estimated and various quantities related to the spread of parameter values, such as the 0.025 and 0.975 quantiles (giving the 95% equal-tail credible intervals) or, alternatively, the highest posterior density limits [see Gelman et al. 2003 for a very readable account of Bayesian analysis in general].

The x_i in the algorithm above are sampled from the distribution of the data marginal to the parameters, $p(x)$, often called the marginal likelihood. The observed acceptance rate in the algorithm above gives an estimate of the marginal likelihood $p(x)$ evaluated at the observation y. The prior predictive distribution is the distribution of any summaries $S(x)$ computed from the data

$$p(S(x)) = \int p(S(x) \mid \theta)\pi(\theta)d\theta,$$

and therefore includes the marginal likelihood $p(x)$ as a special case. In the context of ABC, because the data are summarized (see below), the two names are usually interchangeable. The prior predictive distribution is often used in model validation and assumes a particular importance in ABC analysis in which the aim of the modeling is to simulate data that are similar to the observations (see Section 6). Related to the prior predictive distribution, and also used in model validation, is the posterior predictive distribution,

$$\int p(S(x) \mid \theta)p(\theta \mid S(y))d\theta,$$

samples of which can be straightforwardly drawn by sampling $\theta_i \sim p(\theta \mid S(y))$ and then simulating $x_i \sim p(x \mid \theta)$.

When the data are of low dimension but continuously distributed, step 3 of the Rubin (1984) algorithm can be modified as

Reject θ_i if $\rho(x_i, y) > \epsilon$,

where $\rho(\cdot)$ is a function measuring the distance between simulated and observed points. Once the data become high dimensional (a sample consisting of many uni- or multivariate measurements, for example), then it may become necessary to reduce the dimensionality through the use of summary statistics, thus giving the following prototype rejection-ABC algorithm:

Given observation y, repeat the following until N points have been accepted:

1. Draw $\theta_i \sim \pi(\theta)$.
2. Simulate $x_i \sim p(x \mid \theta_i)$.
3. Reject θ_i if $\rho(S(x_i), S(y)) > \epsilon$.

Here, as with x, y, and θ, the function $S(\cdot)$ may be vector-valued. Bayesian computation in terms of this modified algorithm is illustrated in **Figure 1**. One can regard the computation of summary statistics as a set of mappings from a high dimension to a low dimension. Typically information

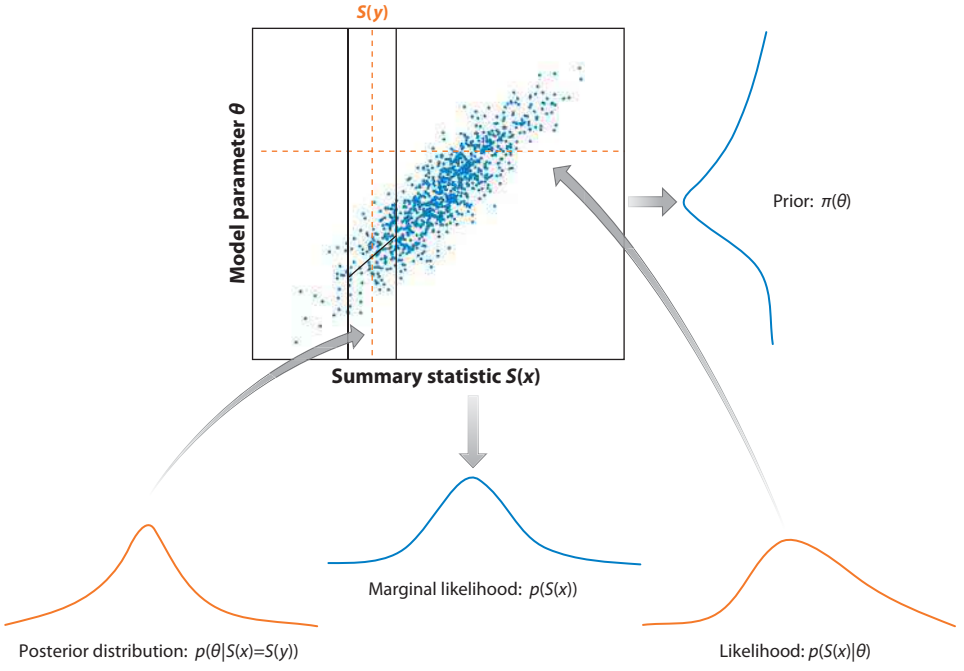

Figure 1

This figure illustrates rejection- and regression-based approximate Bayesian computation (ABC). Points are sampled from the joint distribution, which is a product of the prior, and the likelihood, $p(\theta, S(x)) = p(S(x) \mid \theta)\pi(\theta)$. The prior, $\pi(\theta)$ is the marginal distribution on the right-hand side of the figure. The marginal likelihood, or prior predictive distribution, $p(S(x))$, is the marginal distribution at the bottom of the figure. The aim is to approximate the posterior distribution, $p(\theta \mid S(y))$, evaluated at the point $S(y)$. This can be achieved by rejecting all points that lie outside the region enclosed by the two straight lines, for which $\rho(S(x), S(y)) > \epsilon$. Alternatively, potentially more accurate conditional density estimation may be used. This is illustrated here by the straight line representing $E[\theta \mid S(x)]$, with which the simulated points are adjusted in regression ABC. Adapted with permission from Beaumont & Rannala 2004 and Beaumont 2008.

is lost, but, with enough of these low dimensional summaries much of the information in the high-dimensional data may be captured. The aim is for the summary statistics to satisfy

$$p(\theta \mid x) = p(\theta \mid S(x))$$

for all priors $\pi(\theta)$, which is known as Bayes sufficiency (Kolmogorov 1942).

3. A BRIEF HISTORY

As noted above, there have been early methods proposed within the statistical literature that are very much within the ABC spirit (Diggle & Gratton 1984, Rubin 1984). However, the recent interest (and the name) primarily derives from four papers (Beaumont et al. 2002, Marjoram et al. 2003, Pritchard et al. 1999, Tavaré et al. 1997), all of which were motivated by problems in population genetic analysis. The antecedents to those papers are briefly discussed, and some subsequent developments are highlighted.

Although likelihood-based estimation had been introduced into population genetics at an early stage (Cavalli-Sforza & Edwards 1967, Ewens 1972), most methods of inference prior to the

1990s used moment estimators. Subsequently, a series of papers developed computer-intensive techniques for likelihood-based analysis (Griffith & Tavaré 1994, Kuhner et al. 1995), leading the way to the current preoccupation with likelihood and Bayesian inference in population genetics (Beaumont & Rannala 2004, Marjoram & Tavaré 2006). However, these approaches are hard to implement in genealogical modeling, and there has always been a strong motivation to consider alternative, more approximate, approaches. The use of a rejection algorithm for Bayesian inference of population genetic parameters was proposed by Tavaré et al. (1997). A simplified version of the algorithm, keeping the salient features, is:

Given observation $S(y)$, likelihood function $p(S(x) \mid \theta)$, and $K = \max_{\theta} p(S(y) \mid \theta)$, repeat the following until N points have been accepted:

1. Draw $\theta_i \sim \pi(\theta)$,
2. Simulate $u_i \sim U(0, 1)$,
3. Reject θ_i if $p(S(y) \mid \theta_i)/K \leq u_i$.

These are sampled from $p(\theta \mid S(y))$.

At first glance, their algorithm appears rather different from the prototype rejection-ABC algorithm above in that an analytical function, $p(S(x) \mid \theta)$, giving the likelihood of the summary could be obtained; rejection is then performed simply by devising an algorithm in which proposed θ_i are accepted in proportion to this likelihood. The less efficient version of the algorithm, in which $S(x_i)$ is randomly simulated under the model for given θ_i, and then all θ_i are rejected when $S(x_i) \neq S(y)$, corresponds to the rejection-ABC algorithm. However, the price of efficiency is less flexibility, because some analytical function is needed for the rejection step. More similar to the rejection-ABC algorithm, but in a non-Bayesian framework, are the techniques proposed by Fu & Li (1997) and Weiss & von Haeseler (1998):

Summarize the observed data as $S(y)$.

Define a grid of parameter values.

1. Choose θ_i from grid.
2. Set $N = 0$; repeat the following M times:
   a. Simulate $x_i \sim p(x \mid \theta_i)$.
   b. If $\rho(S(x_i), S(y)) \leq \epsilon$ then $N = N + 1$.
3. Compute approximation to likelihood, $\mathcal{L}(\theta_i) \propto (N/M)$.

Such an approach has been used in a number of other papers (e.g., Wall 1999). Note the similarity to the rejection-ABC algorithm. If, instead, θ_i is sampled from a uniform distribution defined by the bounds of the grid, rejection is performed according to the rejection-ABC algorithm, and then some density estimation procedure is applied to the distribution of accepted θ_i, then an equivalent result would be obtained.

In the first clear example of what would be called an ABC approach, introducing the rejection-ABC algorithm described above, Pritchard et al. (1999) summarized up to 445 Y-chromosome gene copies assayed at eight microsatellites into just three numbers: the number of distinct haplotypes, the mean across loci of the variance in repeat number, and heterozygosity. These were chosen on the basis of past theoretical investigations. Their distance function $\rho(\cdot)$ was chosen to be a Chebyshev distance of the form

$$\max_j \mid S_j(x) - S_j(y) \mid \quad \text{for} \quad j = 1 \ldots s \text{ summary statistics.}$$

Pritchard et al. (1999) normalized the summaries by dividing by $S(y)$, giving

$$\max_j \left| \frac{S_j(x)}{S_j(y)} - 1 \right| \quad \text{for} \quad j = 1 \ldots s \text{ summary statistics.}$$

The Pritchard et al. (1999) paper was introduced at a time when a number of full-likelihood methods in population genetics were being described, and there was heightened interest in the potential and power of such approaches. However, it rapidly became clear that scaling of MCMC and IS methods to large data sets and complex problems was challenging, and a number of different techniques were proposed, unrelated to ABC, based on approximations to likelihoods (Li & Stephens 2003, Nielsen 2000). Meanwhile, an interest in summary-based methods continued, and one of the issues that needed to be confronted was the so-called curse of dimensionality when using more than a few summary statistics. To put this into context, suppose that the summary statistics are all uncorrelated and their prior predictive distribution uniformly distributed between 0 and 2 for each statistic. If the target value is 1 for each summary statistic, and we require a 10% tolerance, using the Chebyshev distance used by Pritchard et al. (1999), then we reject 90% of the points for one summary statistic, 99% of the points for two summary statistics, 99.9% of the points for three summary statistics, etc. Obviously, this example is extreme, but if we have a complex model in which parameters are not highly correlated in the posterior distribution (that is, are identifiable in the model), then for the summaries to be approximately jointly sufficient, a number of these summaries of at least similar magnitude to the number of identifiable parameters in the model are required. The curse of dimensionality clearly represents a major problem.

Subsequent developments for tackling this issue are discussed in more detail in sections below, but a brief outline is given here. One approach was suggested by Beaumont et al. (2002) who noted that the procedure by Pritchard et al. (1999) can be viewed as a method of conditional density estimation: Points are sampled from the joint distribution of parameters and summary statistics $p(x, \theta)$, and then rejection is used to approximate, for observation y, the conditional density, $p(\theta \mid y)$ (**Figure 1**). They proposed a method based on local linear regression and showed that a much wider tolerance interval could be used (that is, larger proportion of simulated points accepted) for the same degree of error in a standard rejection algorithm. Indeed, for multivariate normal summaries and parameters, the method should approximate the posterior well if all simulated points are accepted. An alternative method was suggested by Marjoram et al. (2003), who noted that it is very inefficient to simply sample trial parameter values from the prior. If the proposal mechanism is embedded in an MCMC sampler, then the proposal mechanism becomes more adaptive because one is effectively sampling from the posterior with some added noise. Thus, the distribution of proposed summary statistics (marginal to the state of the chain) is close to the posterior predictive distribution, which, if the data are informative, should occupy a much narrower proportion of the space. The approach taken by Beaumont et al. (2002) is a postsampling correction method, whereas that of Marjoram et al. (2003) attempts to obtain the samples more efficiently in the first place. The MCMC approach may have difficulty mixing (converging to the target distribution), as discussed in Section 4.2, and Sisson et al. (2007) introduced the third main class of ABC method, based on SMC. In this approach, one starts with a wide tolerance interval, sampling parameters from the prior and rejecting those parameters that give summary statistics outside the tolerance interval. But then the estimates are successively refined by resampling from parameter values that have already been obtained. Again, as with MCMC, the distribution of summary statistics in the proposals is close to the posterior predictive distribution at the end.

Further refinements are associated with the studies highlighted above. For conditional density estimation, a significant advance has been proposed by Blum & François (2010), who have used nonlinear regression and additional enhancements. Another approach to density estimation has been proposed by Leuenberger & Wegmann (2010). The MCMC method has been improved by Bortot et al. (2007) to allow for a variable tolerance, allowing the MCMC to mix better. A method for choosing among a set of summary statistics in a systematic way has been suggested by Joyce & Marjoram (2008), and the use of ABC to validate models is described by Ratmann et al. (2009).

Useful recent reviews of ABC have been provided by Csillery et al. (2010) and Bertorelle et al. (2010).

4. APPROXIMATE BAYESIAN COMPUTATION TECHNIQUES

The three main groups of methods are briefly described here. In addition, two particular special cases that frequently arise in Bayesian inference, model choice, and hierarchical modeling pose particular challenges for ABC, and the various approaches that have been developed for tackling these issues are discussed.

4.1. Regression-Based Conditional Density Estimation

As noted in Beaumont et al. (2002), another way to look at the rejection-ABC algorithm (introduced in Section 2) is as a method of conditional density estimation. Once it is viewed in this way, then other methods that have been developed for such problems can be applied. Beaumont et al. (2002) proposed the following algorithm based on local linear regression (Ruppert & Wand 1994):

1. Given observation y, repeat the following until M points have been generated:
   a. Draw $\theta_i \sim \pi(\theta)$.
   b. Simulate $x_i \sim p(x \mid \theta_i)$.
2. Compute k_j, the empirical standard deviation of the $S_j(x)$.
3. Define $\rho(S(x), S(y))$: $\sqrt{\sum_{j=1}^{s} (S_j(x)/k_j - S_j(y)/k_j)^2}$.
4. Choose tolerance ϵ such that the proportion of accepted points $P_\epsilon = N/M$.
5. Weight the simulated points $S(x_i)$ using $K_\epsilon(\rho(S(x_i), S(y)))$, where
$$K_\epsilon(t) = \begin{cases} \epsilon^{-1}\left(1 - (t/\epsilon)^2\right) & t \le \epsilon \\ 0 & t > \epsilon \end{cases}.$$
6. Apply weighted linear regression to the N points that have nonzero weight to obtain an estimate of $\hat{E}(\theta \mid S(x))$.
7. Adjust $\theta_i^* = \theta_i - \hat{E}[\theta \mid S(x_i)] + \hat{E}[\theta \mid S(y)]$.
8. The θ_i^*, with weights $K_\epsilon(\rho(S(x_i), S(y)))$, are taken to be random draws from an approximation to the posterior distribution $p(\theta \mid y)$.

The assumption behind step 7 in the algorithm is that only the mean θ varies with $S(x)$, and the distribution of residuals remains constant in θ. It should be noted that although the use of local linear regression to obtain estimates of $E[\theta \mid x]$ is a well-developed approach, the regression-adjustment step for conditional density estimation appears not to have been used outside ABC. Typically, regression adjustment is performed on one parameter at a time to obtain estimates of marginal posterior distributions. However, standard multivariate multiple regression gives the same estimates of the expected value of θ, regardless of whether regression is applied marginally or jointly. This is because the least-squares machinery assumes that the residuals are uncorrelated between parameters. That is, if θ has components θ_1 and θ_2, then $\{\hat{E}[\theta_1 \mid S(y)], \hat{E}[\theta_2 \mid S(y)]\} = \hat{E}[\{\theta_1, \theta_2\} \mid S(y)]$. Thus, sequentially applying the algorithm above to each parameter gives an approximation to the joint posterior distribution for all parameters. In practice, this appears to work quite well (Bazin et al. 2010); however, in reality, the residuals are generally correlated between parameters, and there is undoubtedly scope for improving on the least-squares approximation in a multivariate setting.

Concerning the weighting scheme (step 5), R.D. Wilkinson (submitted, arXiv 0811.3355) notes that if the observations are viewed as arising under a model with some error (which could be the Epanechnikov above, or some Gaussian error kernel), then a rejection algorithm similar to that

discussed by Tavaré et al. (1997) (Section 3), based on the error density, effectively gives the same results. In this case, under pure rejection without regression adjustment, the weighting scheme can be viewed as sampling without approximation (other than in choice of summary statistic) from some alternative model that is a convolution of the error distribution and original model.

Regression-based methods are not without problems. A difficulty occurs when the observed summary statistics lie outside the prior predictive distribution, which is a symptom of a misspecified model. In this case, regression is then extrapolating rather than interpolating. Even if linearity assumptions hold true, it is a well-known feature of linear regression that the error on the regression increases rapidly with increasing distance from the centroid. However, for most problems, the relationship between summary statistics and parameters is highly nonlinear, further exacerbating problems in the the use of regression.

Blum & François (2010) introduced a number of improvements to the regression-ABC algorithm. Rather than using linear regression to compute $\hat{E}[\theta \mid S(y)]$, they use a feed-forward neural network model implemented in R (the nnet package). Additionally, they model the log of the squared residuals of the regression using nnet to obtain an estimate of $\hat{\sigma}[\theta \mid S(x)]$. It is then possible to modify step 7 in the regression-ABC algorithm as

$$\theta_i^* = \frac{\hat{\sigma}[\theta \mid S(y)]}{\hat{\sigma}[\theta \mid S(x_i)]}(\theta_i - \hat{E}[\theta \mid S(x_i)]) + \hat{E}[\theta \mid S(y)].$$

The intuition here is that now both the expectation and variance of θ are allowed to vary with $S(x)$, and if, for example, a residual is in a low-variance part of the space relative to the variance at $S(y)$, the magnitude of the residual is expanded. A further refinement by Blum & François (2010) is to add an additional SMC step (see Section 4.3). With these additional steps, Blum & François are able to demonstrate enhanced accuracy compared to the original method by Beaumont et al. (2002).

An additional difficulty with regression-ABC, discussed by Beaumont et al. (2002) and highlighted by Leuenberger & Wegmann (2010), arises when the regression adjustment projects points that are outside the prior bounds when a bounded prior is used. This problem occurs when the observations lie at the boundaries of the prior-predictive distribution. Transformations can be used to keep the points within the support of the product of the prior and likelihood function—for example, a log or logistic transformation preceding regression adjustment, followed by back-transformation—however, these methods are somewhat ad hoc. Certainly a log-transformation may make sense for a parameter whose support in the likelihood is bounded by zero, such as the mutation rate. However, when bounded priors are used, a logistic transformation may produce a spike of high density at the boundary and may also introduce strong nonlinearity. A more accurate procedure is probably to discard points that are projected outside a bounded prior that does not coincide with the support of the likelihood.

Some of these issues are avoided by Leuenberger & Wegmann's (2010) approach, in which they consider the inverse problem of modeling the $S(x_i) = f(\theta_i) + a + e_i$, where $f()$ is a linear function, a is a vector constant, and the e_i are drawn from a multivariate normal whose variance-covariance matrix is also to be estimated. The key behind this approach is that because a multivariate normal model is assumed, it is then possible to compute the likelihood $p(S(x) \mid \theta)$ directly from the fitted model. Because the prior is also known analytically, it is then also possible to compute the posterior distribution $p(\theta \mid S(x))$ directly. Because the method is modeling $S(x)$, there are no problems in the estimation of θ, which remains within the support of the model. Furthermore, because the likelihood takes a simple form, it is possible to integrate over parameters to obtain a marginal likelihood that can then be compared between models. The key drawback of the method is the

assumption of multivariate normality. However, Leuenberger & Wegmann (2010) find that it is not a major restriction in practice.

4.2. Markov Chain Monte Carlo

A disadvantage of rejection and regression ABC is that parameter values are sampled from the prior, and if the data are informative, the posterior distribution will be more concentrated than the prior. Thus, the great majority of simulated parameter values does not give rise to summary statistics that are similar to those in the data. Conceivably an efficient conditional density estimation procedure could use the information from all simulated points, but in practice this is unlikely to be achieved and, hence, many points will need to be rejected or given negligible weight. We therefore need some procedure whereby the parameters are sampled from a distribution that is closer in shape to the posterior than from the prior. There are two main ways to do this. One approach is via MCMC simulation, which is discussed in this section, and the other is through SMC, which is discussed in the next section. The MCMC method is a standard approach to solving many Bayesian problems, and it is beyond the scope of this review to enter into the details (see Gelman et al. 2003, Robert & Casella 2004). An MCMC-ABC algorithm, based on that of Marjoram et al. (2003), is as follows:

Initialize by sampling $\theta^{(0)} \sim \pi(\theta)$.
At iteration $t \geq 1$,

1. Simulate $\theta' \sim K(\theta \mid \theta^{(t-1)})$.
2. Simulate $x \sim p(x \mid \theta')$.
3. If $\rho(S(x), S(y)) < \epsilon$,
   a. $u \sim \mathcal{U}(0, 1)$;
   b. if $u \leq \pi(\theta')/\pi(\theta^{(t-1)}) \times K(\theta^{(t-1)} \mid \theta')/K(\theta' \mid \theta^{(t-1)})$,
   $$\theta^{(t)} = \theta';$$
   c. otherwise
   $$\theta^{(t)} = \theta^{(t-1)};$$
4. otherwise
   $$\theta^{(t)} = \theta^{(t-1)}.$$

In this algorithm, one starts by sampling from the prior $\pi(\theta)$ (although, if the chain truly converges, any starting point will do). Then a new value of θ is proposed using a proposal distribution $K(\cdot)$ that depends on the current value of θ. Marjoram et al. (2003) show that this leads to samples from $p(\theta \mid \rho(S(x), S(y)) < \epsilon)$. At convergence, the marginal proposal distribution (that is, the average distribution of proposed θ') is typically dominated by the posterior distribution itself, thus addressing the criticism of the plain rejection method. It is further possible to apply any of the regression-adjustment methods on the MCMC sample in order to obtain a more accurate estimate (Wegmann et al. 2009).

A potential drawback in this application of the MCMC machinery to ABC is that the acceptance rate of the algorithm is proportional to the frequency with which data are simulated such that $\rho(S(x), S(y)) < \epsilon$ (itself proportional to the likelihood in the limit of small ϵ). This is in contrast to MCMC with explicit likelihoods, where the acceptance rate is a function of the ratio of the likelihood for the proposed and current state. A consequence is that the rejection rate is highest, and therefore the rate of mixing is slowest, in the tails of the distribution, which can lead to poor mixing of the MCMC (Sisson et al. 2007; but see Ratmann 2010), especially if the starting point is chosen far into the tails of the posterior distribution.

A number of techniques, however, can be used to try to circumvent this problem. For example the tolerance ϵ can be initially high and then reduced during the burn-in phase of the MCMC so that sampled parameter values are within the vicinity of the mode in the final production run of the chain (Ratmann et al. 2007). An alternative approach is to augment the simulated Markov chain with the tolerance parameter itself (Bortot et al. 2007). In this case a tight prior, for example, an exponential distribution, is used to keep ϵ close to zero while allowing larger values to be occasionally used. It should be noted that this is not an adaptive procedure (one is not finding an optimum tolerance), but is more of a device for allowing the tolerance to vary while keeping detailed-balance. Although the interpretation of the resulting marginal posterior distribution for ϵ is rather unclear, it is possible to condition on particular values of ϵ and also to examine how the posterior distribution varies with ϵ. Small values of ϵ generally have low acceptance rates and therefore may not be frequently sampled. Thus, this procedure is not a magic bullet for obtaining convergence within narrow tolerance bounds, but at least has the advantage of producing converged MCMC chains within ABC.

As noted by Becquet & Przeworski (2007), there is a close relationship between the algorithm by Marjoram et al. (2003) and another algorithm that has become known as the pseudo-marginal method (Andrieu & Roberts 2009) introduced by Beaumont (2003). In the algorithm by Marjoram et al. (2003), the rejection step (**step 3**, above) can be regarded as a Monte Carlo estimate, based on a sample of size 1, of the approximated likelihood function. Andrieu & Roberts (2009) show that, more generally, Monte Carlo estimates of the likelihood based on any size of sample (including ABC approximations) can be used in place of the true likelihood in the Metropolis-Hasting accept-reject step provided their expectation corresponds to the true likelihood. Irrespective of the sampling error, the MCMC converges to the true posterior distribution (or, in the context of ABC, to the approximated posterior distribution). This is not a free lunch, however, and estimates with a high error converge poorly. The relevance for ABC is that the convergence properties of the original Marjoram et al. (2003) algorithm can be improved by the use of multiple samples from the same parameter values [as demonstrated by Becquet & Przeworski (2007)].

4.3. Sequential Monte Carlo

Sisson et al. (2007) proposed a method for iteratively improving on an ABC approximation. Their approach consisted of two main features: weighted resampling from the set of points already drawn and a successive reduction in the tolerance ϵ. A bias has been noted in the original algorithm by Sisson et al. (2007), which has been corrected (Beaumont et al. 2009, Sisson et al. 2009, Toni et al. 2009). This bias arises because Sisson et al. (2007) aimed to make their proposed SMC algorithm broadly applicable and based it on a very general algorithm described by Del Moral et al. (2006). Unfortunately, it turns out (Beaumont et al. 2009) that the generality of the Del Moral result does not apply in the ABC setting.

A straightforward SMC algorithm is the following [taken from Beaumont et al. (2009), but very similar to that of Toni et al. (2009) and Sisson et al. (2009)]:

Given a decreasing sequence of tolerance thresholds $\epsilon_1, \ldots, \epsilon_T$,

1. At iteration $t = 1$,
   for $i = 1, \ldots, N$,
      until $\rho(S(x), S(y)) < \epsilon_1$
        simulate $\theta_i^{(1)} \sim \pi(\theta)$ and $x \sim p(x \mid \theta_i^{(1)})$.
   Set $\omega_i^{(1)} = 1/N$.
   Take τ_2^2 as twice the empirical variance of the $\theta_i^{(1)}$'s.

2. At iteration $2 \leq t \leq T$,

for $i = 1, \ldots, N$,
until $\rho(S(x), S(y)) < \epsilon_t$
pick $\theta_i^\star$ from the $\theta_j^{(t-1)}$'s with probabilities $\omega_j^{(t-1)}$;
generate $\theta_i^{(t)} \sim K(\theta \mid \theta_i^\star; \tau_t^2)$ and $x \sim p(x \mid \theta_i^{(t)})$.
Set $\omega_i^{(t)} \propto \pi(\theta_i^{(t)}) / \sum_{j=1}^N \omega_j^{(t-1)} K(\theta_i^{(t)} \mid \theta_j^{(t-1)}; \tau_t^2)$.

Take τ_{t+1}^2 as twice the weighted empirical variance of the $\theta_i^{(t)}$'s.

In the first step, one simulates parameter values from the prior, simulates data and summary statistics, and retains the closest N points, as in rejection-based ABC. This set of points can be regarded as having been drawn from an approximation to the posterior distribution. One then fits a density kernel $q(\theta) = \sum_{j=1}^N \omega_j^{(t-1)} K(\theta \mid \theta_j^{(t-1)}; \tau_t^2)$ around the points (and it can be shown that a density with twice the empirical variance of the simulated points is a good choice; Beaumont et al. 2009) and resamples parameter values from this. A Gaussian kernel is typically used, but other distributions such as the t-distribution may also be useful. The tolerance ϵ is reduced. Data are then simulated and summary statistics computed, and the closest N points are retained. Each point is given a weight (importance weight) $\pi(\theta)/q(\theta)$ that takes into account that the points are not sampled from the prior. This weighted set of points then gives an improved approximation to the posterior because the tolerance interval is smaller. The process is then continued for further iterations. The algorithms proposed by Toni et al. (2009) and Beaumont et al. (2009), and the revised algorithm proposed by Sisson et al. (2009) are very similar and are based on the IS ideas above. An alternative method has been suggested by Del Moral and colleagues (P. Del Moral, A. Doucet, and A. Jasra, in revision, available at **http://www.math.u-bordeaux1.fr/~delmoral/delmoral_doucet_jasra_smcabc.pdf**), based on Del Moral et al. (2006), but avoids the bias as described by Sisson et al. (2007). As with the MCMC-ABC method of Becquet & Przeworski (2007), Del Moral and colleagues show that performance can be improved by the use of repeated samples with the same parameter values, and their SMC procedure appears to improve on the IS algorithm.

The attractive feature of SMC-ABC is that it addresses a drawback of the rejection/regression approach in that the latter is very inefficient if the posterior is very narrow in relation to the prior (that is, when the data are informative). By repeatedly sampling from a steadily improving approximation to the posterior, the distribution of summary statistics becomes closer to the posterior predictive distribution (Section 2), and the density of simulated data points with summaries in the vicinity of the target is increased. However, in the limit that one is sampling parameter values from $p(\theta \mid \rho(S(x), S(y)) < \epsilon)$, there is little scope for improving the proportion of simulated points that are within the tolerance regions. Thus, one gets a brick-wall effect in that, with a schedule of decreasing ϵ, the proportion of points that are accepted becomes very small, without major improvement in the approximation of the posterior. The SMC-ABC algorithms can then be potentially further enhanced to address this, as with MCMC-ABC, by including a conditional density estimation step (using, for example, regression) on the final distribution of points (Blum & François 2010, Lopes & Beaumont 2010). Indeed, it can potentially be used at every step in an IS-based SMC-ABC algorithm, although the stability of the algorithm may be compromised if there are inaccuracies in the regression adjustment that lead to a strongly skewed distribution of weights.

4.4. Bayesian Model Choice

Bayesian inference can also be straightforwardly applied to models as well as parameters themselves. Thus, given a series of models $\mu_1, \mu_2, \ldots, \mu_\Omega$ with some prior probabilities $\sum_i \pi(\mu_i) = 1$,

it is of interest to compute the posterior probability $p(\mu_i \mid x)$ and also the likelihood ratio, or Bayes factor, between two models i and j:

$$\frac{p(\mu_i \mid x)}{p(\mu_j \mid x)} \div \frac{p(\mu_i)}{p(\mu_j)} = \frac{p(x \mid \mu_i)}{p(x \mid \mu_j)}.$$

The right-hand side of the above equation is the ratio of marginal likelihoods (Section 2). There are some occasional misunderstandings that arise about Bayesian model choice, and it is perhaps worth reiterating some of the major aspects here. With special exceptions, Bayes factors between models can be computed only if the parameters within the models have priors that integrate to one (these are termed proper priors; for example, one cannot assume a flat prior between $\pm\infty$). Bayesian model choice may be strongly affected by the priors placed on the parameters within the models. Bayesian model choice automatically penalizes models with many parameters, therefore one does not need to account for different numbers of parameters between models (see MacKay 2003, Chapter 28, for an illuminating discussion of Bayesian model choice in the context of Occam's razor).

Pritchard et al. (1999) describe the use of model choice with ABC for distinguishing between different mutation models. The idea is that the average (marginal) acceptance rate is proportional to the marginal likelihood. If exactly the same tolerance interval, ϵ, is used for each model, then the constant of proportionality is the same, and an estimate of the ratio of marginal likelihoods is given by the ratio of acceptance rates. Given relative acceptance rates and priors on the models, posterior probabilities can readily be obtained. This approach has been widely used (e.g., Estoup et al. 2004, Miller et al. 2005). It is also possible to use the regression framework to estimate posterior probabilities directly (Beaumont 2008, Fagundes et al. 2007). In this approach, the model indices are treated as categorical variables in a multinomial (polychotomous) regression. With two models, this is logistic regression. The approach draws $M \sim \pi(\mu)$ and estimates $p(M = \mu_i \mid S(y))$.

Comparisons of the regression and rejection approaches to model choice approaches are given in Fagundes et al. (2007), Beaumont (2008), and Guillemaud et al. (2009). A feature that is highlighted by these studies is that the rejection approach is generally only accurate when P_ϵ (see Section 4.1) is very small: In these examples, the closest 100–200 points out of 5×10^5–10^6 draws from the prior. By contrast, the regression method is relatively insensitive to the choice of tolerance, and the value of ϵ for parameter estimation is often used. With very small P_ϵ, the rejection method tends to give similar answers to regression (Fagundes et al. 2007, Guillemaud et al. 2009). However, care must be taken that the observations $S(y)$ are not an outlier under the prior predictive distribution; otherwise substantial differences between regression- and rejection-based inference may occur. Typically, with the rejection approach, the observations are ranked by their Euclidean distance from the target, and then a running average estimate of the posterior probability is computed. The DIY-ABC software package (Cornuet et al. 2008) performs these computations, allowing the user to compare regression and rejection approaches to model choice.

It is also possible to perform model choice within the MCMC and SMC frameworks. For SMC, the approach is essentially the same as for parameter estimation. The model indicator is treated as a categorical parameter, and importance weighting (see Section 4.3) is used as before. Toni et al. (2009) use model choice in a systems biology context to compare different epidemiological models. With MCMC, one method would be to use reversible-jump MCMC (RJMCMC) (Green 1995). However, model choice has not yet been implemented within an MCMC-ABC context.

4.5. Hierarchical Bayesian Models

Bayesian models have great flexibility, and one aspect is the use of hierarchical structures in the model. Hierarchical models are normally used when there are a large number of similar units (loci,

populations, individuals, etc.), and there is uncertainty whether these units should be parameterized identically or independently. The basic idea is that the parameter, Θ of the prior $\pi_\Theta(\theta)$, is not a given but is itself inferred. For example, the mutation rate may not be the same at each locus. Thus, we model each locus as having mutation rate θ_j in a model of the form:

$$p(\theta \mid x) \propto \left[\prod_j p(x_j \mid \theta_j) p(\theta_j \mid \Theta) \right] \pi(\Theta).$$

In this case, the hyperparameter Θ might be the vector (ξ, σ) giving the location and scale. As $\sigma \to 0$, $\theta_j \to \xi$, and the model reduces to a nonhierarchical model,

$$p(\xi \mid x) \propto \left[\prod_j p(x_j \mid \xi) \right] \pi(\xi),$$

with all loci having the same mutation rate, $\theta_j = \xi$. At the other extreme, if σ is sufficiently large such that $p(\theta_j \mid \Theta)$ is the same for all loci, then we have

$$p(\theta \mid x) \propto \left[\prod_j p(x_j \mid \theta_j) \pi(\theta_j) \right],$$

and all the mutation rates are estimated independently. That is, the data at one locus has no effect on the mutation rate inferred at a different locus.

A hierarchical model allows these two extremes to be spanned in a flexible way, and at the same time allows us to examine whether a model of identical parameters is reasonable or not. The more evidence there is favoring small σ, as in the example above, the more each locus contributes to the estimation of Θ, which is a property known as borrowing strength. The hierarchical framework can also be viewed as an alternative to Bayesian model choice (Gelman et al. 2003): In the example above, we might have estimated posterior probabilities of models with mutation rates that are all different or all the same. Gelman et al. (2003) argue strongly in favor of the benefits of hierarchical models.

An alternative to the hierarchical models above is the empirical Bayes method in which, in the context of the example above, the hyperparameter Θ is estimated by other, often classical, ways from the data (e.g., a moment-based estimator of the mean and variance in mutation rate among loci). This approach was taken by Pritchard et al. (1999) in an ABC analysis using a model with varying microsatellite mutation rates. Once the prior is specified, the mutation rates can then be drawn and ABC performed. A fully hierarchical model for mutation rates was used in an ABC analysis by Excoffier et al. (2005).

A potential difficulty with the use of hierarchical models in an ABC context is that summary statistics should capture information from each unit so that the hyperparameters can be well inferred. However, if used naively, this would result in a plethora of statistics, scaling with the number of units. This aspect has been discussed in a number of papers by Hickerson and coworkers (Hickerson et al. 2006, 2010; Hickerson & Meyer 2008), who have emphasized the utility of ABC in hierarchical models (abbreviated to HABC in those studies). There are two related problems. First, as noted above, if there are many units (e.g., loci) then there are potentially many summary statistics, and it is unlikely that simulated data will closely match the observations. Second, even if there are only a few units, there is still the problem of how best to order the summary statistics for comparison with the observations (Hickerson et al. 2006, Sousa et al. 2009). Typically, the units are exchangeable (that is, the likelihood does not depend on their label or ordering), and therefore

it is quite arbitrary which simulated unit should be matched with which observation. If there are a number of summary statistics measured for each unit, the problems become magnified.

In the more general case with many units, it will never be feasible to match summary statistics from each unit, and Bazin et al. (2010) describe two algorithms for performing HABC with large data sets. The basic insight is that in many hierarchical models the posterior distribution factorizes as

$$p(\Theta, \theta \mid x) = \left[\prod_{i=1}^{L} p(\theta_i \mid x_i, \Theta)\right] p(\Theta \mid x),$$

where Θ is the hyperparameter, and θ_i and x_i are, respectively, the parameter and observation at the ith unit (population, locus, etc.). This suggests that inference of the hyperparameters can be performed independently of inference of the parameters. Bazin et al. (2010) distinguish between symmetric and unit-specific summary statistics, $S(x)$ and $U(x_i)$. The former—means, variances, higher moments of summaries computed for each unit—are invariant to ordering and can be used to infer the hyperparameters independently of the parameters. These can be substituted for the data,

$$p(\Theta, \theta \mid x) \approx \left[\prod_{i=1}^{L} p(\theta_i \mid U(x_i), \Theta)\right] p(\Theta \mid S(x)).$$

To infer each parameter, Bazin et al. (2010) show that only the symmetric summary statistics and those summary statistics for the unit in question are needed. In one algorithm, the symmetric and unit-specific summary statistics are all output in a single ABC run. In the other, a two-step procedure is used, similar to SMC, where in the first pass only the symmetric summary statistics are output and only the hyperparameters are inferred. The hyperparameters are then resampled from this approximate posterior and the joint distribution of parameters and summary statistics is simulated for a single unit, which is then used to infer the unit-specific paramaters. As noted in Bazin et al. (2010), this latter algorithm involves a small approximation, but the effect decreases with increasing number of units. There is widespread scope for the analysis of complicated models both in population genetics and ecology through hierarchical approaches, and it can be anticipated that there will be many future applications.

5. CHOICE OF SUMMARY STATISTICS

An undoubtedly rather arbitrary area of ABC modeling lies in the choice of summary statistics for a problem. In some fields, such as population genetics, there is a history of the development of summary statistics within a model-based framework. For example, the number of segregating sites in a sample of DNA sequences and the average pairwise difference between any two sequences are jointly informative about past growth rates in the population (Tajima 1989), but each summary statistic by itself is not. However, it is also possible that data summaries are used without necessarily a strong theory relating these summaries to parameters in a model. The effect of summary statistics on inferences is an area that has not been systematically studied. A related issue is whether summary statistics chosen by consideration of one particular model might then bias the outcome of model choice because these summary statistics may have little relation to parameters in other models. Again, this is a topic that would benefit from further study.

The main concept is that as more and more summary statistics are used, they should be jointly sufficient for the likelihood (Section 2). In addition, with many summaries of varying degrees of correlation to each other and the parameters, it is hoped that the influence of any one on the outcome may be diminished. The major argument against this strategy is that the accuracy and stability of ABC decreases rapidly with increasing numbers of summary statistics (the curse

of dimensionality discussed in Section 3). Joyce & Marjoram (2008) suggested an approach for choosing summary statistics from among a large candidate set. In their method, the ratio of estimated posterior densities is computed for candidate sets with/without a particular summary. Departures from one by more than a threshold amount in this ratio are taken to indicate that the excluded summary is important. By sequentially applying this procedure, subsets of informative summaries may be found.

The method above effectively gives an all-or-nothing weight to particular summary statistics, depending on whether they are chosen or not. An alternative idea is to weight different summaries according to the magnitude of some measure of association with the parameter values. Hamilton et al. (2005a) used weights based on the marginal regression coefficients of each summary with the parameter of interest and showed that this yielded an improvement over a study conducted by Beaumont et al. (2002). The idea here is that if a summary statistic has little relationship to a parameter value, then we should tolerate large deviations in the summary statistic from the target by giving it a lower weight in the calculation of the Euclidean distance. Otherwise there is a danger of rejecting points where informative summary statistics are close to the target because of noise induced by the uninformative summary statistics. Another way of looking at this is that in Beaumont et al.'s (2002) method, the tolerance boundary is a hypersphere around the target. The modification by Hamilton et al. (2005a) makes this boundary a hyperellipse, the width of whose axes are proportional to the correlation with the parameter values. More generally, in local linear modeling, the issue of weighting is subsumed under the choice of multivariate bandwidth (Ruppert & Wand 1994), and this is often chosen by cross-validation. Thus, it can be seen that the choice of metric, particularly in a multivariate setting, is also an important consideration.

The number of summary statistics can also be reduced through multivariate dimension reduction techniques by rotating the distribution of summary statistics. For example, Wegmann et al. (2009) have used partial least-squares (PLS), where orthogonal axes are chosen that have the highest correlation to the parameter of interest. Alternatively, principal components analysis (PCA) can be used to rotate the axes, where the orthogonal axes are chosen on the basis of maximizing the variance (Bazin et al. 2010). Because it uses information from the independent variable, PLS often shows a modest improvement over PCA in high dimensional regressions.

6. MODEL VALIDATION AND TESTING

An important component of Bayesian modeling is validation and testing. Validation usually refers to the process of assessing goodness of fit of the model and comparing between models, rather than to debugging and error checking. ABC simulation can often be run in parallel on a cluster and, hence, can be relatively fast to run. In addition, the rejection-/regression-based approaches only need generate one large sample from the joint distribution of parameters and summary statistics, which can then be applied to many simulated data sets in order to test the accuracy of the method.

Because ABC involves approximation, it is useful to try to distinguish errors due to the approximation from errors caused by the choice of model. Note that there is no way of checking that the underlying simulations are correctly implemented. All the tests discussed here use the same simulation procedure both to generate test data sets and to perform the ABC. One way of checking the underlying simulations is to generate test data sets using a different simulation procedure. Alternatively, in a field that is well studied, it may be possible to compare the expected value of summary statistics against their theoretical expectations. Often in frequentist analyses, validation and testing have included estimation of power and coverage properties as well as bias. These have often been carried over into the analyses of ABC methods (Beaumont et al. 2002, Excoffier et al. 2005). Many of the earlier tests of ABC looked at coverage, bias, and mean square error by repeated

simulation with a single parameter set, as is typically done in frequentist analysis. However, unlike frequentist methods, which typically involve asymptotic approximations, Bayesian credible intervals are exact under repeated sampling from the prior, but are not exact under repeated sampling from a single point. Thus, repeated sampling from a point, because it is subject to error anyway, will not provide a good test of the approximation inherent in ABC. More recently, tests of ABC methods under repeated sampling from the prior have been performed (Rosenblum et al. 2007, Wegmann et al. 2009), and these more accurately diagnose problems. However, even if a method passes these tests, there is still uncertainty as to how closely the inferences match those that would have been obtained if a tractable likelihood function had been available.

In ABC, where we treat the summary statistics as data, the prior predictive distribution of these summary statistics is the same as the marginal likelihood (Section 2; see also **Figure 1**). Often the marginal distributions for each summary statistic (or pairs of summaries) are visualized and compared with the corresponding target statistic (e.g., Bazin et al. 2010, Itan et al. 2009). If the target is outside this distribution, which could be summarized by a p-value, for example, then this would suggest a problem in the model. A related test, more commonly used in Bayesian model validation, is to compare the observations with the posterior predictive distribution (Section 2). This can be realized approximately by simulating data with parameters drawn randomly from the current posterior sample. From a Bayesian perspective, it is difficult to justify any procedure that judges the fit of a model without reference to another. For these reasons, Gelman et al. (2003) caution that p-values should be interpreted more liberally. Yet, from a more pragmatic standpoint, these distributions are, in practice, very useful. Often the method is applied to summary statistics marginally, yet the joint distribution may be more informative. For example, in population genetics, the joint distribution of the number of segregating sites and pairwise difference may be more diagnostic of a failure of a stable population model than either marginally.

Unlike Bayesian methods based on a likelihood function, examining the prior predictive distribution of summary statistics serves a more important role than simply examining model fit. The entire reliability of the procedure is predicated on the assumption that the distribution of parameter values conditional on the observed summary statistics is sufficiently similar also for simulated summary statistics that are either close enough to these observations or can be adjusted through some regression procedure. However, if the observations lie well outside the simulated distribution, this assumption becomes more tenuous and is particularly problematic for regression, as discussed in Section 4.1. This is in contrast to methods based on a likelihood function for which reliable estimation is always possible, however unlikely the observations under the model. These issues are discussed in depth by Ratmann et al. (2009), who examine the distribution of discrepancies ($S_j(x) - S_j(y)$) for each summary statistic to diagnose problems in model specification.

7. APPLICATIONS OF APPROXIMATE BAYESIAN COMPUTATION

The initial applications of ABC have, reflecting its origins, been mainly in the context of population genetic analysis. Most of these have used rejection or regression. However, in the past few years a number of other areas in ecology, epidemiology, and systems biology have seen an increase in the use of ABC for inference, and the majority of these applications have been based on MCMC or SMC. The following sections attempt to give a broad overview of the current literature.

7.1. Population Genetics

In many areas of population genetics, the data consist of the frequencies of alleles or haplotypes from one or several populations. The goal is often to understand the demographic history of the

populations (in terms of migration rates, changes in population sizes, and times of demographic events such as colonizations or admixture events). Recently, there has been interest in finding evidence of natural selection at the genomic level. A variety of approaches have been taken, including the use of full-likelihood methods and various approximations, including ABC.

7.1.1. The demography of colonization. After the initial paper by Pritchard et al. (1999), an early application using the same algorithm came from Estoup et al. (2001), who modeled the invasion of the Pacific islands by the cane toad. The parameters in this model consisted of the invasion time, bottleneck duration, and population sizes. A further paper (Estoup & Clegg 2003) looked at the colonization history of silvereyes in the Pacific, inferring similar parameters and, again, using the plain rejection algorithm. A relatively small number of summary statistics was used, although more than was used by Pritchard et al. (1999) (four in the case of the toads, and seven for the silvereyes), and rejection was based on the Chebyshev distance. As a consequence of the curse of dimensionality noted above, with seven summary statistics, a rather larger relative tolerance was required and the rejection rate was more than 20-fold higher. Later papers dealing with the demographic history of invasions have been based on the regression approach by Beaumont et al. (2002) and have typically used more summary statistics. An example is the modeling of the invasion by cane toads of northern Australia by Estoup et al. (2004), in which more than 60 summary statistics were used (mostly made of pairwise measures of differentiation between populations) to infer parameters relating to the founder numbers and rates of migration. Other studies that have used regression-based ABC to analyze colonization scenarios have been conducted by Miller et al. (2005; invasion of Europe by western corn rootworm), Guillemaud et al. (2009; model comparison using western corn rootworm), Pascual et al. (2007; introduction history of *Drosophila subobscura*), Rosenblum et al. (2007; colonization of novel habitat by a lizard), and Legrand et al. (2009; demographic origins of *Drosophila sechellia*). Rejection-based ABC has been used by Thornton & Andolfatto (2006), to quantify the magnitude and timing of demographic bottlenecks in *Drosophila melanogaster*, and also by Mardulyn et al. (2009), to look at the colonization history of a leaf beetle. François et al. (2008), in modeling the colonization history of European *Arabidopsis thaliana*, found the regression method by Beaumont et al. (2002) gave poor results and used a weighted rejection scheme. These studies have used a variety of simulation techniques, often written specifically for the problem at hand, or have used programs such as *ms*. More explicitly, spatial models have also been used with regression-based ABC. For example, Hamilton et al. (2005a) modeled spatially distributed data as arising from a history of colonization followed by expansion with migration. Neuenschwander et al. (2008) modeled the colonization history of bullheads, and Estoup et al. (2010) performed a more detailed analysis of cane toads in Australia. However, studies by Neuenschwander et al. (2008) and Estoup et al. (2010) have used the SPLATCHE program (Currat et al. 2004, Neuenschwander 2006), which allows for quite realistic modeling of spatial problems. The landscape is divided into a lattice, and local densities and rates of movement can be specified. The study by Estoup et al. (2010) introduces a technique for approximating diffusion on a continuous space.

7.1.2. Human demographic history. The analysis of human demographic history has motivated the development of a wide number of statistical techniques based on full likelihood (e.g., Hey 2005), likelihood-based approximations (Hellenthal et al. 2008), and a variety of approaches using summary statistics (Belle et al. 2006, Pluzhnikov et al. 2002), including a few that have used ABC methods (Fagundes et al. 2007, Hamilton et al. 2005b, Patin et al. 2009), based on regression. The study by Hamilton et al. (2005b) inferred differential rates of maternal and paternal gene flow in Thai populations from a comparison of Y and mtDNA diversity using a range expansion model similar to that used by Hamilton et al. (2005a). Fagundes et al. (2007) modeled worldwide human

genetic diversity surveyed at 50 autosomal sequence loci from African, Asian, and Native American samples. They used these data to compare a number of different models of recent human evolution, including various versions of the out-of-Africa model and the multiregional hypothesis. They used Beaumont's (2008) regression-based procedure to obtain the posterior probability of different models, and this enabled them to conclude that an out-of-Africa model with population expansion best explained the data. A similar method of analysis has recently been used by Patin et al. (2009) to compare different demographic models that explain the genetic differentiation within different populations of African Pygmies, as well as between Pygmies and African agricultural populations.

7.1.3. Phylogeography. Statistical phylogeography is another area in which a number of full-likelihood methods are available [reviewed by Nielsen & Beaumont (2009)]. Given that phylogeographic models tend to be relatively simple (branching models of population divergence, with or without gene flow), it may often be the case that a full-likelihood method, which should be more powerful, is more appropriate for the problem under study. However, large sample size, including many loci, may lead to poor convergence of MCMC methods and may be more amenable to an ABC approach. Recombination, also, is a phenomenon that is not handled well by full-likelihood methods, and this is addressed in Becquet & Przeworski (2007). Their technique uses a modification of the MCMC-ABC method of Marjoram et al. (2003) (Section 4.2). For nonrecombining data, Becquet & Przeworski (2007) were able to show comparable performance to the IM software program of Hey & Nielsen (2004), which is a full-likelihood method that can be used for nonrecombining sequences. Becquet & Przeworski (2007) analyzed up to 68 loci (sequences of length around 700 bp) in chimpanzees and a smaller number from gorillas and orangutans to obtain estimates of divergence times, migration rates, and effective population sizes. These analyses have been performed with pairs of taxa, but the ABC method has also been used with up to five taxa to compare different topologies (Palero et al. 2009) using model choice with regression. Hickerson et al. (2006) studied the problem of inferring concordant divergence times of multiple paired taxa in the context of a hierarchical Bayesian model fitted by regression ABC (Section 4.5). In this model, it is assumed that each pair of taxa has information concerning some environmental phenomenon (closure of the Panama isthmus, for example) that leads to vicariance. A hierarchical approach is appropriate because one can obtain information on the variance in divergence times among taxa. Hickerson et al. (2006) applied their method to the distribution of divergence times of eight echinoid sister taxon pairs separated by the emergence of the Isthmus of Panama. Related analyses for different scenarios are given in Hickerson & Meyer (2008) and Leache et al. (2007), and a discussion of HABC approaches is given in Hickerson et al. (2010). Rejection-based ABC has been used by Jakobsson et al. (2006) and Putnam et al. (2007) to estimate divergence times and population sizes in phylogeographic models. The modeling of admixture is closely related to the models of vicariance discussed above. In this case, population lineages fuse at certain times, rather than diverge, and the interest is to date the time of admixture, admixture proportions, and effective population sizes. As with the modeling of population divergence, admixture has also been well studied in the context of likelihood-based models, and there have also been some ABC-based studies that have made close comparisons with the likelihood-based methods (Excoffier et al. 2005, Sousa et al. 2009). More general migration models have not been extensively analyzed with ABC: Foll et al. (2008) introduce a method, mainly in the context of ascertainment modeling, and Bazin et al. (2010) infer migration rates, but in the context of detecting selection (discussed below).

7.1.4. Ancient DNA and temporally spaced samples. With the increasing prevalence of ancient DNA samples, there has been much interest in making more fine-scaled, and better calibrated,

inferences about past demographic events using full-likelihood approaches (Drummond et al. 2002). On an ecological timescale, it is also possible to use archived material to obtain estimates of changes in population size with simpler models that ignore mutation (Wang 2005). The full-likelihood methods are typically relatively limited in scope to single populations that are changing size through time, and there has been interest in using ABC methods to model more detailed scenarios pertaining to the problem at hand. An early study in this area is by Chan et al. (2006), who used the Bayesian Serial Simcoal program to infer the strength and date of a population bottleneck within the past 10,000 years from samples of the tuco-tuco (*Ctenomys sociabilis*) using regression-based ABC. Interestingly, although a number of subsequent studies have used the BayeSSC program to simulate scenarios (e.g., Belle et al. 2009, Ramakrishnan & Hadly 2009) and to choose among them by comparing simulated and observed summary statistics under a goodness-of-fit procedure, it does not appear to have been used within an ABC framework per se. For data on an ecological timescale, Tallmon et al. (2004) have introduced an ABC method (using regression) for inferring effective population size (N_E) based on two samples from microsatellites. The main advantage of this method over full-likelihood methods is that, for microsatellites, it allows for a more realistic prior on the initial gene frequencies. A related ABC application, although based on a single sample, is to infer current population sizes using patterns of pairwise relatedness and linkage disequilibrium (Ramakrishnan et al. 2004, Tallmon et al. 2008).

7.1.5. Detecting selection and inferring recombination. An early application of ABC was that by Tishkoff et al. (2001), who modeled selection at the G6PD locus in humans. They were able to infer selection on an allele at the target G6PD locus and mutation and recombination rate at a nearby microsatellite marker locus using rejection. In their method, they used a Poisson branching process model to simulate a genealogy with selection. An alternative approach is to use an approximation based on the coalescent, in which selection at one allele is modeled as an expanding population and recombination at a marker locus acts like migration. This has been used by Przeworski (2003) to develop a method for inferring relevant parameters relating to selective sweeps; that is, selection coefficient, time of onset of selection, recombination rate, mutation rate at the neutral locus, and effective population size. The illustrative example studied by Przeworski (2003) used data from the *tb1* gene in maize, but the method can be widely applied and is based on the rejection algorithms of Tavaré et al. (1997) and Pritchard et al. (1999). Regression-based ABC has been used to infer relevant parameters in a model of recurrent selective sweeps in *Drosophila melanogaster* using forward simulations (Jensen et al. 2008). The same approach has also been used to model recurrent sweeps in the neo-X chromosome of *Drosophila miranda* (Bachtrog et al. 2009). All the above models are of positive selection in the genome. Positive selection can also be approximated using codon models with different rates of synonymous and nonsynonymous mutation. The use of ABC to infer parameters in a model that includes recombination has been described by Wilson et al. (2009). Another application of ABC for inferring recombination is given by Tiemann-Boege et al. (2006). An approach to detecting evidence of locally adaptive selection is by identifying loci that have an increased level of genetic differentiation (Foll & Gaggiotti 2008). For loci with high and variable mutation rates, current likelihood-based methods may yield false positives, and an ABC method for inferring parameters in a hierarchical Bayesian model has been introduced by Bazin et al. (2010), which is competitive with full-likelihood methods when the assumptions of the latter are met and exceed its performance when loci have variable mutation rates. ABC has also been used to model parameters relating to selection on lactase persistence in a spatially explicit modeling framework (Gerbault et al. 2009, Itan et al. 2009).

7.2. Ecology, Epidemiology, and Systems Biology

Aspects of ecology, epidemiology, and systems biology are structurally very similar, as noted by Toni et al. (2009). Many features are captured by deterministic or stochastic compartment models, the dynamics of which can be modeled by systems of partial or ordinary differential equations (o.d.e.s) or stochastic differential equations (s.d.e.s). Data often consist of multivariate time series or snapshot spatial data taken at one or several points in time. The goal is then to compare between hypothesized dynamic models that could explain the patterns and to infer the parameters in candidate models or averaged over models.

7.2.1. Systems biology. The study by Toni et al. (2009) introduces a number of different modeling scenarios, including the Lotka-Volterra example from ecology and also epidemiological models. They address inference problems using the SMC framework, first introduced by Sisson et al. (2007), correcting for the bias in the latter's original algorithm [see also the review by Secrier et al. (2009)]. The data are usually a time series of counts or concentrations, and they give examples based on o.d.e.s and the equivalent model as a system of s.d.e.s. Rejection is typically based on the Euclidean distance between the target and simulated time series. Toni et al. (2009) also use their approach for model choice, enabling them to obtain posterior probabilities of different transmission models. A further example of the use of model choice by SMC-ABC is conducted by Toni & Stumpf (2010). In the first application of ABC to an agent-based model (ABM), Sottoriva & Tavaré (2010) have used a rejection algorithm to infer parameters related to differentiation of cells and tissues based on methylation data from cancerous cells. Another area of systems biology within which ABC methods have been used is in the study of protein interaction networks (PINs). There is now substantial data on the topology of such networks, and the interest is in using the data to discriminate between different models of network evolution. Although likelihood-based methods have been developed (Wiuf et al. 2006), estimation is often intractable (Wiuf & Ratmann 2009), and it has proved efficient to simulate graphs from models of network evolution and to compare those graphs to those observed using MCMC-ABC (Ratmann et al. 2007, 2009). An application of ABC in a systems biology context, but relevant also to genealogical models, is the study of Tavaré (2005), who describes a method using rejection ABC for inferring parameters in a branching process model from data on tumor mutations.

7.2.2. Ecology and palaeontology. Within ecology there have been relatively few applications of ABC methods, although Toni et al. (2009) give an example using a Lotka-Volterra system with time series data on abundances. There is currently substantial interest in the validation of agent-based models in ecology (Piou et al. 2009), and this may become a fertile area of application for ABC. The MCMC-ABC method has been used to analyze species abundance data under a sequential broken stick model (Solow & Smith 2009). In this case, the summary statistics were chosen, on the basis of earlier studies, to be the number of distinct species in a sample and the frequency of the commonest and second-commonest species. ABC has also been used to infer regional speciation rates and local immigration of species under a neutral ecological model (Jabot & Chave 2009). The study by Rabosky (2009) uses MCMC-ABC to infer extinction rates from phylogenetic data. Here the phylogeny is assumed to be known without error from a tree estimation procedure, and the parameters of a model with heritable extinction rates are fitted using data from North American wood warblers. A related example that uses rejection ABC is provided by Wilkinson & Tavaré's (2009) study (see also Wilkinson 2007), in which fossil count data are used to infer parameters in a branching process model. Wilkinson et al. (2010) also integrate molecular information with fossil count data, using a variant of MCMC-ABC.

7.2.3. Epidemiology. There have been a number of epidemiological analyses that have used the ABC method, mostly taking either the SMC or MCMC approach. Tanaka et al. (2006) used MCMC-ABC to fit parameters of an epidemiological model of tuberculosis transmission. Their data consists of counts of different genotypes (identical in structure to the Solow & Smith (2009) data, discussed above, in a different context); the rationale is that the pattern of clustering gives information on the underlying transmission dynamics. The count data is summarized as gene diversity (heterozygosity) and number of alleles. The study by Shriner et al. (2006) looks at the evolution of HIV within the host using a metapopulation model, fitted by regression ABC. Walker et al. (2010) use the SMC-ABC method to study a SARS coronavirus outbreak in Hong Kong. They use the standard SEIR (susceptible-exposed-infected-removed) epidemic model but arrange it on a lattice so that individuals infect neighbors and also a number of more distant individuals. McKinley et al. (2009) provide a detailed comparison of the behavior of SMC-ABC with full-likelihood MCMC in the analysis of a SEIR model and show that very similar results can be obtained. Luciani et al. (2009) use SMC-ABC to infer parameters in a birth-death-mutation model of tuberculosis transmission with the aim of estimating parameters relating to the development of drug resistance.

8. CONCLUSIONS: THE FUTURE OF APPROXIMATE BAYESIAN COMPUTATION

When a likelihood function can be formulated and efficiently evaluated, there is no advantage to using ABC as an alternative. For many of the applications considered here, the likelihood function can be evaluated in principle, but in practice it is computationally too expensive. An alternative to ABC in these cases is to devise approximations that are more tractable. The current incarnation of ABC arose from population genetics, yet it is unlikely to be the dominant method of analysis in this field. A variety of analytical approximations have been developed here (Li & Stephens 2003, Nielsen 2000) and are increasingly applied to large-scale genomic data from humans (Hellenthal et al. 2008, Nielsen et al. 2005). As any field of research becomes more intensely studied, it is likely to be the case that the patterns in the data are primarily determined by dominating processes that are amenable to approximation (Wakeley 2004).

From a practical point of view, many of the scenarios that interest evolutionary biologists and ecologists can be realized in terms of a data-generating stochastic process. It is intuitive and straightforward for many biologists to devise schemes for simulating data, whereas encapsulating the same model in terms of a likelihood function that can be evaluated remains within the purview of a rather more specialized few. Thus, there is likely to be a creative tension between the many for whom the ABC procedure is an eminently practical way forward and the few for whom the same model could potentially be written down as a likelihood function.

ABC is mainly useful in the initial exploratory phase of research. There is often a need for ad hoc modeling of particular cases, particularly in areas of applied biology. Also, it may well be the case that for some problems the structure of the likelihood is very well understood but computationally intractable, and an ABC solution is suitable and adequate (Grelaud et al. 2009). A particular field in which the ABC approach may prove to be compelling is in the area of agent-based modeling (Sottoriva & Tavaré 2010), where there is little hope of obtaining a likelihood, and there is insufficient common structure for any approximations developed in one area of application to be transferred to another.

DISCLOSURE STATEMENT

The author is not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

I am very grateful to David Balding, Lounès Chikhi, Mark Kirkpatrick, Olli Ratmann, Richard Sibly, Simon Tavaré, and Ian Wilson for their helpful comments on a previous draft. This review was written while at the School of Biological Sciences at the University of Reading in the United Kingdom.

LITERATURE CITED

Anderson CNK, Ramakrishnan U, Chan YL, Hadly EA. 2005. Serial SimCoal: a population genetics model for data from multiple populations and points in time. *Bioinformatics* 21:1733–34

Andrieu C, Roberts GO. 2009. The pseudo-marginal approach for efficient Monte Carlo computations. *Ann. Stat.* 37:697–725

Bachtrog D, Jensen JD, Zhang Z. 2009. Accelerated adaptive evolution on a newly formed X chromosome. *PLoS Biol.* 7:712–19

Bazin E, Dawson KJ, Beaumont MA. 2010. Likelihood-free inference of population structure and local adaptation in a Bayesian hierarchical model. *Genetics.* 185:587–602

Beaumont MA. 2003. Estimation of population growth or decline in genetically monitored populations. *Genetics* 164:1139–60

Beaumont MA. 2008. Joint determination of topology, divergence time and immigration in population trees. In *Simulations, Genetics and Human Prehistory, McDonald Inst. Monogr*, ed. S Matsumura, P Forster, C Renfrew, pp. 135–54. Cambridge, UK: McDonald Inst. Archaeol. Res.

Beaumont MA, Cornuet J-M, Marin JM, Robert CP. 2009. Adaptive approximate Bayesian computation. *Biometrika* 96:983–90

Beaumont MA, Rannala B. 2004. The Bayesian revolution in genetics. *Nat. Rev. Genet.* 5:251–61

Beaumont MA, Zhang W, Balding DJ. 2002. Approximate Bayesian computation in population genetics. *Genetics* 162:2025–35

Becquet C, Przeworski M. 2007. A new approach to estimate parameters of speciation models with application to apes. *Genome Res.* 17:1505–19

Belle EMS, Benazzo A, Ghirotto S, Colonna V, Barbujani G. 2009. Comparing models on the genealogical relationships among Neandertal, Cro-Magnoid and modern Europeans by serial coalescent simulations. *Heredity* 102:218–25

Belle EMS, Ramakrishnan U, Mountain JL, Barbujani G. 2006. Serial coalescent simulations suggest a weak genealogical relationship between Etruscans and modern Tuscans. *Proc. Natl. Acad. Sci. USA* 103:8012–17

Bertorelle G, Benazzo A, Mona S. 2010. ABC as a flexible framework to estimate demography over space and time: some cons, many pros. *Mol. Ecol.* 19:2609–25

Blum MGB, François O. 2010. Non-linear regression models for approximate Bayesian computation. *Stat. Comput.* 20:63–73

Bortot P, Coles SG, Sisson SA. 2007. Inference for stereological extremes. *J. Am. Stat. Assoc.* 102:84–92

Bray TC, Sousa V, Parreira B, Bruford MW, Chikhi L. 2010. 2BAD: an application to estimate the parental contributions during two independent admixture events. *Mol. Ecol. Resour.* 3:538–41

Cavalli-Sforza LL, Edwards AWF. 1967. Phylogenetic analysis. Models and estimation procedures. *Am. J. Hum. Genet.* 19:233–57

Chan YL, Anderson CNK, Hadly EA. 2006. Bayesian estimation of the timing and severity of a population bottleneck from ancient DNA. *PLoS Genet.* 2:e59

Cornuet J-M, Santos F, Beaumont MA, Robert CP, Marin JM, et al. 2008. Inferring population history with DIY ABC: a user-friendly approach to approximate Bayesian computation. *Bioinformatics* 24:2713–19

Csillery K, Blum MGB, Gaggiotti F, François O. 2010. Approximate Bayesian computation (ABC) in practice. *Trends Ecol. Evol* 25:410–18

Currat M, Ray N, Excoffier L. 2004. SPLATCHE: a program to simulate genetic diversity taking into account environmental heterogeneity. *Mol. Ecol. Notes* 4:139–42

Del Moral P, Doucet A, Jasra A. 2006. Sequential Monte Carlo samplers. *J. R. Stat. Soc. Ser. B* 68:411–36

Diggle PJ, Gratton RJ. 1984. Monte-Carlo methods of inference for implicit statistical models. *J. R. Stat. Soc. Ser. B* 46:193–227

Drummond AJ, Nicholls GK, Rodrigo AG, Solomon W. 2002. Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. *Genetics* 161:1307–20

Estoup A, Baird SJE, Ray N, Currat M, Cornuet J-M, et al. 2010. Combining genetic, historical and geographic data to reconstruct the dynamics of bioinvasions: application to the cane toad *Bufo marinus*. *Mol. Ecol. Resour.* 10:886–901

Estoup A, Beaumont M, Sennedot F, Moritz C, Cornuet J-M. 2004. Genetic analysis of complex demographic scenarios: spatially expanding populations of the cane toad, *Bufo marinus*. *Evolution* 58:2021–36

Estoup A, Clegg SM. 2003. Bayesian inferences on the recent island colonization history by the bird *Zosterops lateralis lateralis*. *Mol. Ecol.* 12:657–74

Estoup A, Wilson IJ, Sullivan C, Cornuet J-M, Moritz C. 2001. Inferring population history from microsatellite and enzyme data in serially introduced cane toads, *Bufo marinus*. *Genetics* 159:1671–87

Ewens WJ. 1972. The sampling theory of selectively neutral alleles. *Theor. Popul. Biol.* 3:87–112

Excoffier L, Estoup A, Cornuet J-M. 2005. Bayesian analysis of an admixture model with mutations and arbitrarily linked markers. *Genetics* 169:1727–38

Fagundes NJR, Ray N, Beaumont M, Neuenschwander S, Salzano FM, et al. 2007. Statistical evaluation of alternative models of human evolution. *Proc. Natl. Acad. Sci. USA* 104:17614–19

Foll M, Beaumont MA, Gaggiotti O. 2008. An approximate Bayesian computation approach to overcome biases that arise when using amplified fragment length polymorphism markers to study population structure. *Genetics* 179:927–39

Foll M, Gaggiotti O. 2008. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics* 180:977–93

François O, Blum MGB, Jakobsson M, Rosenberg NA. 2008. Demographic history of European populations of *Arabidopsis thaliana*. *PLoS Genet.* 4:e1000075

Fu Y-X, Li W-H. 1997. Estimating the age of the common ancestor of a sample of DNA sequences. *Mol. Biol. Evol.* 14:195–99

Gelman A, Carlin JB, Stern HS, Rubin DB. 2003. *Bayesian Data Analysis*. London: Chapman & Hall/CRC Press

Gerbault P, Moret C, Currat M, Sanchez-Mazas A. 2009. Impact of selection and demography on the diffusion of lactase persistence. *PLoS One* 4:e6369

Green PJ. 1995. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* 82:711–32

Grelaud A, Robert CP, Marin JM, Rodolphe F, Taly JF. 2009. ABC likelihood-free methods for model choice in Gibbs random fields. *Bayesian Anal.* 4:317–35

Griffith R, Tavaré S. 1994. Simulating probability distributions in the coalescent. *Theor. Popul. Biol.* 46:131–59

Guillemaud T, Beaumont MA, Ciosi M, Cornuet J-M, Estoup A. 2009. Inferring introduction routes of invasive species using approximate Bayesian computation on microsatellite data. *Heredity* 104:88–99

Hamilton G, Currat M, Ray N, Heckel G, Beaumont M, Excoffier L. 2005a. Bayesian estimation of recent migration rates after a spatial expansion. *Genetics* 170:409–17

Hamilton G, Stoneking M, Excoffier L. 2005b. Molecular analysis reveals tighter social regulation of immigration in patrilocal populations than in matrilocal populations. *Proc. Natl. Acad. Sci. USA* 102:7476–80

Hellenthal G, Auton A, Falush D. 2008. Inferring human colonization history using a copying model. *PLoS Genet.* 4:11

Hey J. 2005. On the number of New World founders: a population genetic portrait of the peopling of the Americas. *PLoS Biol.* 3:965–75

Hey J, Nielsen R. 2004. Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. *Genetics* 167:747–60

Hickerson MJ, Carstens BC, Cavender-Bares J, Crandall KA, Graham CH, et al. 2010. Phylogeography's past, present, and future: 10 years after Avise, 2000. *Mol. Phylogenet. Evol.* 54:291–301

Hickerson MJ, Meyer CP. 2008. Testing comparative phylogeographic models of marine vicariance and dispersal using a hierarchical Bayesian approach. *BMC Evol. Biol.* 8:322

Hickerson MJ, Stahl EA, Lessios HA. 2006. Test for simultaneous divergence using approximate Bayesian computation. *Evol. Int. J. Org. Evol.* 60:2435–53

Hudson RR. 2002. Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics* 18:337–38

Itan Y, Powell A, Beaumont MA, Burger J, Thomas MG. 2009. The origins of lactase persistence in Europe. *PLoS Comput. Biol.* 5:e1000491

Jabot F, Chave J. 2009. Inferring the parameters of the neutral theory of biodiversity using phylogenetic information and implications for tropical forests. *Ecol. Lett.* 12:239–48

Jakobsson M, Hagenblad J, Tavaré S, Säll T, Halldén C, et al. 2006. A unique recent origin of the allotetraploid species *Arabidopsis suecica*: evidence from nuclear DNA markers. *Mol. Biol. Evol.* 23:1217–31

Jensen JD, Thornton KR, Andolfatto P. 2008. An approximate Bayesian estimator suggests strong, recurrent selective sweeps in *Drosophila*. *PLoS Genet.* 4:e1000198

Jobin MJ, Mountain JL. 2008. REJECTOR: software for population history inference from genetic data via a rejection algorithm. *Bioinformatics* 24:2936–37

Joyce P, Marjoram P. 2008. Approximately sufficient statistics and Bayesian computation. *Stat. Appl. Genet. Mol. Biol.* 7(1):Article 26

Kolmogorov AN. 1942. Determination of the centre of dispersion and degree of accuracy for a limited number of observation. *Izv. Akad. Nauk, USSR Ser. Mat.* 6:3–32

Kuhner MK, Yamato J, Felsenstein J. 1995. Estimating effective population size and mutation rate from sequence data using Metropolis-Hastings sampling. *Genetics* 140:1421–30

Leache AD, Crews SC, Hickerson MJ. 2007. Two waves of diversification in mammals and reptiles of Baja California revealed by hierarchical Bayesian analysis. *Biol. Lett.* 3:646–50

Legrand D, Tenaillon MI, Matyot P, Gerlach J, Lachaise D, Cariou ML. 2009. Species-wide genetic variation and demographic history of *Drosophila sechellia*, a species lacking population structure. *Genetics* 182:1197–206

Leuenberger C, Wegmann D. 2010. Bayesian computation and model selection without likelihoods. *Genetics* 184:243–52

Li N, Stephens M. 2003. Modeling linkage disequilibrium and identifying recombination hotspots using single-nucleotide polymorphism data. *Genetics* 165:2213–33

Lopes JS, Balding D, Beaumont MA. 2009. PopABC: a program to infer historical demographic parameters. *Bioinformatics* 25:2747–49

Lopes JS, Beaumont M. 2010. ABC: a useful Bayesian tool for the analysis of population data. *Infect. Genet. Evol.* 10:826–33

Luciani F, Sisson SA, Jiang HL, Francis AR, Tanaka MM. 2009. The epidemiological fitness cost of drug resistance in *Mycobacterium tuberculosis*. *Proc. Natl. Acad. Sci. USA* 106:14711–15

MacKay DJC. 2003. *Information Theory, Inference, and Learning Algorithms*. Cambridge, UK: Cambridge Univ. Press

Mardulyn P, Mikhailov YE, Pasteels JM. 2009. Testing phylogeographic hypotheses in a Euro-Siberian cold-adapted leaf beetle with coalescent simulations. *Evolution* 63:2717–29

Marjoram P, Molitor J, Plagnol V, Tavaré S. 2003. Markov chain Monte Carlo without likelihoods. *Proc. Natl. Acad. Sci. USA* 100:15324–28

Marjoram P, Tavaré S. 2006. Modern computational approaches for analysing molecular genetic variation data. *Nat. Rev. Genet.* 7:759–70

McKinley T, Cook AR, Deardon R. 2009. Inference in epidemic models without likelihoods. *Int. J. Biostat.* 5(1):Article 24

Miller N, Estoup A, Toepfer S, Bourguet D, Lapchin L, et al. 2005. Multiple transatlantic introductions of the western corn rootworm. *Science* 310:992

Neuenschwander S. 2006. AQUASPLATCHE: a program to simulate genetic diversity in populations living in linear habitats. *Mol. Ecol. Notes* 6:583–85

Neuenschwander S, Largiadr CR, Ray N, Currat M, Vonlanthen P, Excoffier L. 2008. Colonization history of the Swiss Rhine basin by the bullhead (*Cottus gobio*): inference under a Bayesian spatially explicit framework. *Mol. Ecol.* 17:757–72

Nielsen R. 2000. Estimation of population parameters and recombination rates from single nucleotide polymorphisms. *Genetics* 154:931–42
Nielsen R, Beaumont MA. 2009. Statistical inferences in phylogeography. *Mol. Ecol.* 18:1034–47
Nielsen R, Williamson S, Kim Y, Hubisz MJ, Clark AG, Bustamante C. 2005. Genomic scans for selective sweeps using SNP data. *Genome Res.* 15:1566–75
Palero F, Lopes J, Abelló P, Macpherson E, Pascual M, Beaumont MA. 2009. Rapid radiation in spiny lobsters (*Palinurus* spp) as revealed by classic and ABC methods using mtDNA and microsatellite data. *BMC Evol. Biol.* 9:263
Pascual M, Chapuis MP, Mestres F, Balany J, Huey RB, et al. 2007. Introduction history of *Drosophila subobscura* in the New World: a microsatellite-based survey using ABC methods. *Mol. Ecol.* 16:3069–83
Patin E, Laval G, Barreiro LB, Salas A, Semino O, et al. 2009. Inferring the demographic history of African farmers and Pygmy hunter-gatherers using a multilocus resequencing data set. *PLoS Genet.* 5(4):e1000448
Piou C, Berger U, Grimm V. 2009. Proposing an information criterion for individual-based models developed in a pattern-oriented modelling framework. *Ecol. Model.* 220:1957–67
Pluzhnikov A, Di Rienzo A, Hudson RR. 2002. Inferences about human demography based on multilocus analyses of Noncoding sequences. *Genetics* 161:1209–18
Pritchard JK, Seielstad MT, Perez-Lezaun A, Feldman MW. 1999. Population growth of human Y chromosomes: a study of Y chromosome microsatellites. *Mol. Biol. Evol.* 16:1791–98
Przeworski M. 2003. Estimating the time since the fixation of a beneficial allele. *Genetics* 164:1667–76
Putnam AS, Scriber JM, Andolfatto P. 2007. Discordant divergence times among Z-chromosome regions between two ecologically distinct swallowtail butterfly species. *Evolution* 61:912–27
Rabosky DL. 2009. Heritability of extinction rates links diversification patterns in molecular phylogenies and fossils. *Syst. Biol.* 58:629–40
Ramakrishnan U, Hadly EA. 2009. Using phylochronology to reveal cryptic population histories: review and synthesis of 29 ancient DNA studies. *Mol. Ecol.* 18:1310–30
Ramakrishnan U, Storz JF, Taylor BL, Lande R. 2004. Estimation of genetically effective breeding numbers using a rejection algorithm approach. *Mol. Ecol.* 13:3283–92
Ratmann O. 2010. *Approximate Bayesian Computation under model uncertainty, with an application to stochastic processes of network evolution*. PhD thesis. Imperial Coll., London
Ratmann O, Andrieu C, Wiuf C, Richardson S. 2009. Model criticism based on likelihood-free inference, with an application to protein network evolution. *Proc. Natl. Acad. Sci. USA* 106:10576–81
Ratmann O, Jorgensen O, Hinkley T, Stumpf M, Richardson S, Wiuf C. 2007. Using likelihood-free inference to compare evolutionary dynamics of the protein networks of *H. pylori* and *P. falciparum*. *PLoS Comput. Biol.* 3:2266–78
Robert CP, Casella G. 2004. *Monte Carlo Statisical Methods*. New York: Springer
Rosenblum EB, Hickerson MJ, Moritz C. 2007. A multilocus perspective on colonization accompanied by selection and gene flow. *Evolution* 61:2971–85
Rubin DB. 1984. Bayesianly justifiable and relevant frequency calculations for the applied statistician. *Ann. Stat.* 12:1151–72
Ruppert D, Wand M. 1994. Multivariate locally weighted least squares regression. *Ann. Stat.* 22:1346–70
Secrier M, Toni T, Stumpf MPH. 2009. The ABC of reverse engineering biological signalling systems. *Mol. Biosyst.* 5:1925–35
Shriner D, Liu Y, Nickle DC, Mullins JI. 2006. Evolution of intrahost HIV-1 genetic diversity during chronic infection. *Evolution* 60:1165–76
Sisson SA, Fan Y, Tanaka MM. 2007. Sequential Monte Carlo without likelihoods. *Proc. Natl. Acad. Sci. USA* 104:1760–65
Sisson SA, Fan Y, Tanaka MM. 2009. Sequential Monte Carlo without likelihoods. Erratum 1041760. *Proc. Natl. Acad. Sci. USA* 106:16889–89
Solow AR, Smith WK. 2009. Estimating species number under an inconvenient abundance model. *J. Agric. Biol. Environ. Stat.* 14:242–52
Sottoriva A, Tavaré S. 2010. Integrating approximate Bayesian computation with complex agent-based models for cancer research. In *COMPSTAT 2010–Proceedings in Computational Statistics*, ed. G Saporta, Y Lechevallie, p. 57–67. Heidelberg: Springer, Physica Verlag

Sousa VC, Fritz M, Beaumont MA, Chikhi L. 2009. Approximate Bayesian computation without summary statistics: the case of admixture. *Genetics* 181:1507–19

Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123:585–95

Tallmon DA, Koyuk A, Luikart G, Beaumont MA. 2008. ONeSAMP: a program to estimate effective population size using approximate Bayesian computation. *Mol. Ecol. Resour.* 8:299–301

Tallmon DA, Luikart G, Beaumont MA. 2004. Comparative evaluation of a new effective population size estimator based on approximate Bayesian computation. *Genetics* 167:977–88

Tanaka MM, Francis AR, Luciani F, Sisson SA. 2006. Using approximate Bayesian computation to estimate tuberculosis transmission parameters from genotype data. *Genetics* 173:1511–20

Tavaré S. 2005. Ancestral inference for branching processes. In *Branching Processes in Biology: Variation, Growth, Extinction*, ed. P Haccou, P Jagers, V Vatutin, pp. 208–17. Cambridge, UK: Cambridge Univ. Press

Tavaré S, Balding DJ, Griffiths RC, Donnelly P. 1997. Inferring coalescence times from DNA sequence data. *Genetics* 145:505–18

Thornton K, Andolfatto P. 2006. Approximate Bayesian inference reveals evidence for a recent, severe bottleneck in a Netherlands population of *Drosophila melanogaster*. *Genetics* 172:1607–19

Thornton KR. 2009. Automating approximate Bayesian computation by local linear regression. *BMC Genet.* 10:35

Tiemann-Boege I, Calabrese P, Cochran DM, Sokol R, Arnheim N. 2006. High-resolution recombination patterns in a region of human chromosome 21 measured by sperm typing. *PLoS Genet.* 2:682–92

Tishkoff SA, Varkonyi R, Cahinhinan N, Abbes S, Argyropoulos G, et al. 2001. Haplotype diversity and linkage disequilibrium at human G6PD: recent origin of alleles that confer malarial resistance. *Science* 293:455–62

Toni T, Stumpf MPH. 2010. Simulation-based model selection for dynamical systems in systems and population biology. *Bioinformatics* 26:104–10

Toni T, Welch D, Strelkowa N, Ipsen A, Stumpf MPH. 2009. Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. *J. R. Soc. Interface* 6:187–202

Wakeley J. 2004. Recent trends in population genetics: More data! More math! Simple models? *J. Hered.* 95:397–405

Walker DM, Allingham D, Lee HWJ, Small M. 2010. Parameter inference in small world network disease models with approximate Bayesian computational methods. *Phys. A* 389:540–48

Wall JD. 1999. Recombination and the power of statistical tests of neutrality. *Genet. Res.* 74:65–69

Wang JL. 2005. Estimation of effective population sizes from data on genetic markers. *Philos. Trans. R. Soc. B* 360:1395–409

Wegmann D, Leuenberger C, Excoffier L. 2009. Efficient approximate Bayesian computation coupled with Markov chain Monte Carlo without likelihood. *Genetics* 182:1207–18

Weiss G, von Haeseler A. 1998. Inference of population history using a likelihood approach. *Genetics* 149:1539–46

Wilkinson RD. 2007. *Bayesian inference of primate divergence times*. PhD thesis. Univ. Cambridge, UK

Wilkinson RD, Steiper ME, Soligo C, Martin RD, Yang Z, Tavaré S. 2010. Dating primate divergences through an integrated analysis of palaeontological and molecular data. *Syst. Biol.* In press

Wilkinson RD, Tavaré S. 2009. Estimating primate divergence times by using conditioned birth-and-death processes. *Theor. Popul. Biol.* 75:278–85

Wilson DJ, Gabriel E, Leatherbarrow AJ, Cheesbrough J, Gee S, et al. 2009. Rapid evolution and the importance of recombination to the gastroenteric pathogen *Campylobacter jejuni*. *Mol. Biol. Evol.* 26:385–97

Wiuf C, Brameier M, Hagberg O, Stumpf MPH. 2006. A likelihood approach to analysis of network data. *Proc. Natl. Acad. Sci. USA* 103:7566–70

Wiuf C, Ratmann O. 2009. Evolutionary analysis of protein interaction networks. In *Statistical and Evolutionary Analysis of Biological Networks*, pp. 17–49. London: Imperial Coll. Press

RELATED RESOURCES

Packages and Toolboxes. A number of packages and scripts have been devised for inference based on ABC, and some of those known to the author are briefly outlined below.

ABC_distrib Contains basic R scripts for ABC based on local linear regression (Beaumont et al. 2002). Available at **http://www.rubic.rdg.ac.uk/~mab/stuff/ABC_distrib.zip**

abc_nnet Contains R scripts for ABC based on nonlinear regression, and improved modeling of variance (Blum & François 2010). Available at **http://membres-timc.imag.fr/Michael.Blum/publications/R_function/abc_nnetv2.R**

ABCreg Contains code for local linear regression ABC (Thornton 2009). Available at **http://www.molpopgen.org/software/abcreg**

ABC-SysBio Contains code for SMC-based ABC in the context of problems in systems biology (Toni et al. 2009), allowing also for model-choice. Available at **http://abc-sysbio.sourceforge.net/**

ABCtoolbox Contains an extensive collection of programs for ABC inference using MCMC and conditional density estimation techniques, and also for validating and checking the methods (Wegmann et al. 2009). Available at **http://www.cmpg.iee.unibe.ch/content/softwares_services/computer_programs/abctoolbox/index_eng.html**

2BAD Software to infer parameters in models of genetic admixture (Bray et al. 2010). Available at **http://compbio.igc.gulbenkian.pt/pcg/pcg_software.html**

Bayesian Serial SimCoal (BayeSSC) Contains software based on Simcoal (see **http://cmpg.unibe.ch/software/simcoal/**), which is a program for genealogical simulation (much like *ms*). BayeSSC allows ABC-based analysis to be performed, and also allows serially sampled data (that is, ancient DNA) to be used (Anderson et al. 2005). Available at **http://www.stanford.edu/group/hadlylab/ssc/index.html**

DIYABC This program has a 'point and click' interface, allowing the user to build up relatively complex demographic models, for genealogical simulation, which are then analyzed using regression-based ABC, allowing also for model choice (Cornuet et al. 2008). Available at **http://www1.montpellier.inra.fr/CBGP/diyabc/**

MIMAR This program enables demographic parameters relating to a population splitting model to be inferred. It is useful when the data contain recombining sequences. Available at **http://przeworski.uchicago.edu/cbecquet/download.html** (Becquet & Przeworski 2007)

ms The *ms* program is commonly used for genealogical simulation, and, although not specifically oriented to the ABC method, contains a number of useful scripts that enable it to be used for ABC-based inference (Hudson 2002). Available at **http://home.uchicago.edu/~rhudson1/source/mksamples.html**

msABC This program provides a more ABC-friendly interface to *ms*. Available at **http://bio.lmu.de/~pavlidis/msabc**

msBayes This program uses *ms* for ABC inference and is aimed toward applications in statistical phylogeography (Hickerson et al. 2006) Available at **http://msbayes.sourceforge.net/**

PopABC This program gives a pipeline for genealogical simulation with ABC, using regression, and allowing for model choice (Lopes et al. 2009). Available at **http://www.personal.rdg.ac.uk/~sar05sal/software.htm**

REJECTOR2 This program provides an ABC-friendly interface to *msHOT* (variant of *ms*; see website above) (Jobin & Mountain 2008). Available at **http://www.stanford.edu/~mjobin/rejector/Rejector/Home.html**

Annual Review of Ecology, Evolution, and Systematics

Volume 41, 2010

Contents

Indexes

Errata

An online log of corrections to *Annual Review of Ecology, Evolution, and Systematics* articles may be found at http://ecolsys.annualreviews.org/errata.shtml