Approximate Bayesian Computation and MCMC

Vincent Plagnol and Simon Tavaré

Program in Molecular and Computational Biology, University of Southern California, Los Angeles, CA 90089-1340, USA.
vincent.plagnol@ens.fr,stavare@usc.edu

To appear in *Proceedings of MCQMC2002*, edited by H. Niederreiter. Springer Verlag, 2003.

Summary. For many complex probability models, computation of likelihoods is either impossible or very time consuming. In this article, we discuss methods for simulating observations from posterior distributions without the use of likelihoods. A rejection approach is illustrated using an example concerning inference in the fossil record. A novel Markov chain Monte Carlo approach is also described, and illustrated with an example from population genetics.

1 Introduction

This paper concerns methods for simulating observations from a posterior distribution when likelihoods are hard or impossible to compute. This problem arises frequently in the context of complex probability models. Discrete data $\mathcal{D}$ are generated from a model $\mathcal{M}$ determined by parameters θ, whose prior we denote by $\pi(\theta)$. The posterior distribution of interest is given by

$$f(\theta \mid \mathcal{D}) = \mathbb{P}(\mathcal{D} \mid \theta)\pi(\theta)/\mathbb{P}(\mathcal{D}),$$

where $\mathbb{P}(\mathcal{D}) = \int \mathbb{P}(\mathcal{D} \mid \theta)\pi(\theta)\,\mathrm{d}\theta$ is the normalizing constant. The simplest method for generating observations from $f(\theta \mid \mathcal{D})$ uses rejection:

1. Generate θ from $\pi(\cdot)$
2. Accept θ with probability $h = \mathbb{P}(\mathcal{D} \mid \theta)$, and return to 1.

Accepted observations do indeed have the distribution $f(\theta \mid \mathcal{D})$, as shown for example in (11).

There are many variations on this theme. Of particular relevance here is the case in which the likelihood $\mathbb{P}(\mathcal{D} \mid \theta)$ cannot be computed explicitly. One variation is the following:

1. Generate θ from $\pi(\cdot)$
2. Simulate $\mathcal{D}'$ from model $\mathcal{M}$ with parameter θ
3. Accept θ if $\mathcal{D}' = \mathcal{D}$, and return to 1.

Notice that no likelihoods are required in this case. The success of this approach depends on the fact that the underlying stochastic process $\mathcal{M}$ is easy to simulate for a given set of parameters. We note that this approach is also useful when computation of the likelihood is possible, but time consuming.

The practicality of algorithms like these depends crucially on the size of $\mathbb{P}(\mathcal{D})$, because the probability of accepting an observation is just $\mathbb{P}(\mathcal{D})$. In cases where the acceptance rate is too small, one might resort to approximate methods such as the following:

1. Generate θ from $\pi(\cdot)$
2. Simulate $\mathcal{D}'$ from model $\mathcal{M}$ with parameter θ
3. Calculate a measure of distance $\rho(\mathcal{D}, \mathcal{D}')$ between $\mathcal{D}'$ and $\mathcal{D}$
4. Accept θ if $\rho \leq \epsilon$, and return to 1.

This approach requires selection of a suitable metric ρ as well as a choice of ϵ. As $\epsilon \to \infty$, accepted observations come from the prior, and as $\epsilon \to 0$ accepted observations come from the required density $f(\theta \mid \mathcal{D})$. The choice of ϵ reflects the interplay between computability and accuracy. For a given ρ and ϵ accepted observations are independent and identically distributed from $f(\theta \mid \rho(\mathcal{D}, \mathcal{D}') \leq \epsilon)$.

2 Inference in the Fossil Record

To illustrate the last approach, we describe a Bayesian approach to inference in the primate fossil record. The problem concerns estimation of the time to the most recent common ancestor of a group of species, using data from a sampling of the fossil record. For more background to this problem, see (13). Estimates of the divergence time of primates (more accurately, the time of the haplorhine-strepsirrhine split) based on molecular sequence data give a time of about 90 million years. A literal interpretation of the fossil record suggests a divergence time of about 60 million years. One reason for the present studies is to reconcile these two estimates.

2.1 The Problem

In Table 1 the number of primate species found as fossils in a series of stratigraphic intervals is given. In (14), a statistical method for estimating the temporal gap between the base of the stratigraphic interval in which the oldest fossil was found and the initial point of divergence of the species in the sample is given. The bias in the estimators and approximate confidence intervals for the parameters were found by using a parametric bootstrap approach. Here we contrast these results with those from a new Bayesian approach.

Table 1. Data for the primate fossil record. References can be found in the supplemental material in (14).

Epoch	Bin k	Time T_k	Observed number of species (D_k)
Late Pleistocene	1	0.15	19
Middle Pleistocene	2	0.9	28
Early Pleistocene	3	1.8	22
Late Pliocene	4	3.6	47
Early Pliocene	5	5.3	11
Late Miocene	6	11.2	38
Middle Miocene	7	16.4	46
Early Miocene	8	23.8	36
Late Oligocene	9	28.5	4
Early Oligocene	10	33.7	20
Late Eocene	11	37.0	32
Middle Eocene	12	49.0	103
Early Eocene	13	54.8	68
Pre-Eocene	14		0

2.2 A Model for Speciation and Sampling

We adopt the same framework as in (14). We model speciation with a non-homogeneous Markov birth-and-death process. To model evolution from the last common ancestor of all living and fossil species included in the analysis, we start with two species at time 0. Species go extinct at rate λ, and so have exponential lifetimes with mean $1/\lambda$, time being measured in millions of years. A species that goes extinct at time u is replaced by an average of $m(u)$ new species. We denote by Z_t the number of species alive at time t. The expected number of species extant at time t is given by

$$\mathbb{E}Z_t = 2\exp\left\{\lambda\int_0^t (m(u)-1)du\right\}; \tag{1}$$

cf. (4), Chapter 5.

We divide time into k stratigraphic intervals, as shown in Table 1 and Figure 1. The base of the first (youngest) stratigraphic interval is at T_1 million years ago (mya) and the base of the k^{th} is at T_k mya. The earliest known fossil is found in this interval. The founding species originate at time $T := T_k + \tau$ mya, and we define a $(k+1)^{st}$ stratigraphic interval that has its base at $T = T_{k+1}$ mya and ends T_k mya. Note that no fossils have been found in this interval. The aim is to approximate the posterior distribution of the time τ and other parameters of the model, using as data the number of different species found in the fossil record in the first, second, ..., k^{th} intervals. We model the number of species alive u mya by the value Z_{T-u} of the Markov branching process described earlier.

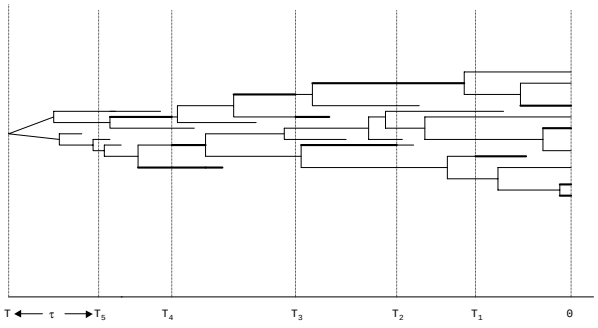

Fig. 1. An illustration of the stochastic model of fossil finds. Bases of 5 stratigraphic intervals at $T_1, , \ldots, T_5$ mya are shown along the x-axis. The temporal gap between the base of the final interval and the point at which the two founding species originate is denoted by τ. Thick lines indicate species found in the fossil record. Time 0 is the present day.

We assume that, conditional on the number of distinct species N_j that lived in the jth stratigraphic interval, the number of species D_j actually found in the fossil record in this interval is a binomial random variable with parameters N_j and α_j, $\;j = 1, 2, \ldots, k+1$. Furthermore, the D_j are assumed to be conditionally independent given the N_j. The parameter α_j gives the probability of sampling a species in the jth stratigraphic interval.

2.3 A Rejection Approach

We write $\mathcal{D} = (D_1, \ldots, D_{k+1})$ for the counts observed in the $k+1$ stratigraphic intervals, and we write θ for the vector of parameters of the process, one of which is τ, the temporal gap. While the likelihood is difficult to compute, the stochastic process itself can be simulated easily. Once the N_j are simulated, the D'_j are binomial samples with parameters N_j and α_j.

The counts $D_1, \ldots, D_{k+1}$ can be represented as the total number of fossils found,

$$D_+ = D_1 + \cdots + D_{k+1},$$

and a vector of proportions

$$(Q_1, \ldots, Q_{k+1}) := \left(\frac{D_1}{D_+}, \ldots, \frac{D_{k+1}}{D_+} \right).$$

We can therefore measure the distance between $\mathcal{D}$ and a simulated data set $\mathcal{D}'$ by

$$\rho(\mathcal{D}, \mathcal{D}') = \left| \frac{D'_+}{D_+} - 1 \right| + \frac{1}{2} \sum_{j=1}^{k+1} |Q_j - Q'_j|. \tag{2}$$

The first term measures the relative error in the total number of fossils found in a simulated data set and the actual number, while the second term is the total variation distance between the two vectors of proportions.

2.4 Results

Original Approach

In (14), the mean diversification was modelled via the logistic function, for which

$$\mathbb{E}Z_t = 2/\{\gamma + (1-\gamma)e^{-\rho t}\}. \tag{3}$$

This form is quite flexible; for example, $\gamma = 0$ corresponds to exponential growth. They equated the expected number of species known at the present time with the observed number (235), and also specified the time at which the mean diversification reached 90% of its current value. These two equations serve to determine the form of the speciation curve, and give $\rho = 0.2995, \gamma = 0.0085$; we assume these values are known in what follows. They also assumed a mean species lifetime of 2.5 my, and they modelled the sampling fractions α_j in the form

$$\alpha_j = \alpha p_j, \quad j = 1, 2, \ldots, k+1, \tag{4}$$

where the p_j are known proportions, and α is a scale parameter to be estimated from the data. The particular values of the p_j they used are given in Table 2. The average value is $\bar{p} = \sum_{j=1}^{k+1} p_j/(k+1) = 0.73$.

Table 2. Sampling proportions p_j

j	1	2	3	4	5	6	7	8	9	10	11	12	13	14
p_j	1.0	1.0	1.0	1.0	0.5	0.5	1.0	0.5	0.1	0.5	1.0	1.0	1.0	0.1

Using the data from Table 1, they estimated a temporal gap of 26.7 my. As the oldest known fossil primate is 54.8 my old, this is equivalent to an age of 81.5 my for the last common ancestor of living primates. The approximate 95% confidence interval is (17.2, 34.8) mya; this corresponds to an interval of (71.0, 89.6) mya for the time of divergence. The average sampling fraction $\bar{\alpha}$, defined as

$$\bar{\alpha} = \alpha \bar{p} \tag{5}$$

was estimated to be 5.7% with an upper 95% confidence limit of 7.4%.

Bayesian Approach

The parameter θ is given by $\theta = (\tau, \alpha)$, and uninformative prior distributions were chosen as $\tau \sim U(0, 100)$ and $\alpha \sim U(0, 0.3)$, the notation $U(a, b)$ denoting the uniform density on (a, b). From 2000 accepted observations with $\epsilon = 0.1$, we obtain the summaries in Figure 2 and Table 3. A median value of 26.3 my for the posterior value of the temporal gap τ is very close to the estimate found in (14) and is equivalent to an age of 81.1 mya for the last common ancestor of living primates. The 2.5% and 97.5% points of the posterior of τ are estimated to be 14.3 my and 61.0 my respectively, corresponding to a 95% credible interval of (69.1, 115.8) mya for the time of divergence. The 95% point of the posterior for $\bar{\alpha}$ is 17.8%. These values are all broadly consistent with the previous analysis, although the Bayesian credible intervals are wider.

The approximate Bayesian methods have a good deal of flexibility, in that they are often easier to extend to more complex models. See (17) for an analysis of different species diversity curves and sampling schemes.

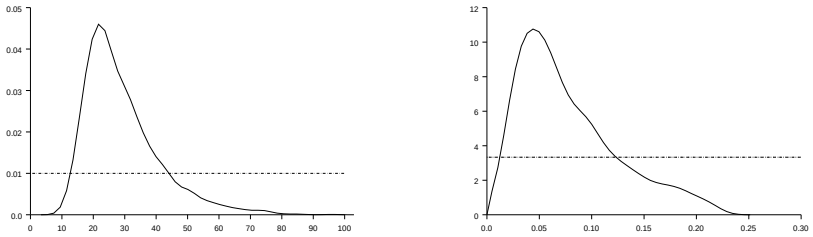

Fig. 2. Left panel: posterior for τ. Right panel: posterior for $\bar{\alpha}$. Horizontal lines show prior density.

Table 3. Summary statistics for τ, $\bar{\alpha}$.

	τ	$\bar{\alpha}(\%)$
1st quartile	20.8	4.1
mean	29.3	7.8
median	26.3	6.4
3rd quartile	34.6	10.3

2.5 A Markov Chain Monte Carlo Approach

The results for the model described above can, in this case, be compared to a full Markov chain Monte Carlo approach that generates observations from the true posterior. The Metropolis-Hastings algorithm (9; 5) generates observations from a Markov chain of θ-values as follows.

1. Suppose the chain is now at θ. Propose a move to θ' according to a transition kernel $q(\theta \to \theta')$
2. Calculate
$$h = \min\left(1, \frac{\mathbb{P}(\mathcal{D} \mid \theta')\pi(\theta')q(\theta' \to \theta)}{\mathbb{P}(\mathcal{D} \mid \theta)\pi(\theta)q(\theta \to \theta')}\right)$$
3. Move to θ' with probability h, else stay at θ; go to 1.

If q is carefully chosen, the limiting and stationary distribution of the chain is $f(\theta \mid \mathcal{D})$; see for example (3). In the next section, we implement this approach to find the posterior distribution of the temporal gap.

2.6 Updating the Birth-Death Process

Our algorithm proposes changes to trajectories of the underlying birth-and-death process illustrated in Figure 1. In what follows, ω denotes a realization of the process, including the parameters α and τ.

The Death/Birth Move

The idea of this move is very simple. We pick a node at random in the tree, kill the branch generated from this node and generate a new one. In this case the Hastings ratio becomes:

$$h(\omega, \omega') = \frac{\mathbb{P}(\mathcal{D} \mid \omega')N(\omega)}{\mathbb{P}(\mathcal{D} \mid \omega)N(\omega')}$$

where $N(\omega)$ is the number of nodes of ω.

In practice this move is too small because most of the nodes are at the bottom of the tree and modifying them does not influence what is happening in the vicinity of the root. A way to deal with this is to choose the node uniformly among the nodes whose times are smaller than a given time. By doing so, we pick only the ones which will actually influence the shape of the tree. In the likelihood ratio computation the total number of nodes must then be replaced by the number of nodes before the chosen time.

The Strip Move

We pick at random two times, a starting point and an ending point. These define a region, a strip, of the branching process. Then we erase this part, and we regenerate it. To do this, we have to simulate the law of the branching

process given the number of species living at the beginning and at the end of the process. This is done by letting the process run a large number of times, from the starting to the ending point, with the right number of individuals at the start, until we reach the ending point with the correct number of individuals. This move is relatively time consuming. We note that this step could exploit analytical results for the distribution of a non-homogeneous birth-death process (4), we have not attempted that here.

This description leaves a lot of freedom in the way one chooses the starting and ending point of the strip. This choice allows us to modulate the impact of the move. A small strip will lead to small changes and a Hastings ratio close to 1. A larger strip will lead to major changes, and a Hastings ratio that is much smaller. The ratio becomes in that case

$$h(\omega, \omega') = \frac{\mathbb{P}(\mathcal{D} \mid \omega')}{\mathbb{P}(\mathcal{D} \mid \omega)}.$$

In our algorithm we used a $U(2, 8)$ distribution for the width of the strip. The starting point of the strip is chosen uniformly between 0 and the total duration of the process.

The Scale Move

We look at the time intervals between two consecutive birth/death events. These times are exponentially distributed, with mean $1/(k\lambda)$, where λ is the death rate and k is the number of existing species at that time. We update these times, according to the prior distribution for example or by adding a Gaussian random variable. If the birth and death process is homogenous and if we use an exponential random variable with correct mean to update the process, the Hastings ratio would be

$$h(\omega, \omega') = \frac{\mathbb{P}(\mathcal{D} \mid \omega')}{\mathbb{P}(\mathcal{D} \mid \omega)}.$$

However for an inhomogeneous birth and death process the situation is slightly more complicated. The new Hastings ratio becomes in that case

$$h(\omega, \omega') = \frac{\mathbb{P}(\mathcal{D} \mid \omega') \prod_i \mathbb{P}_t(k'_i)}{\mathbb{P}(\mathcal{D} \mid \omega) \prod_i \mathbb{P}_t(k_i)},$$

where the product is taken over all the nodes of the process. k_i is the number of offspring at node i and $\mathbb{P}_t(k_i)$ is the probability that the process has such a number of offspring at time t. Of course, if this probability does not depend on t, both terms cancel and give the previous formula for the ratio. It is interesting to notice that this move is the only one which modifies the parameter τ whose law has to be described.

In practice, to make the Hastings ratio not too small it is good to update only a fraction (chosen at random) of the times between each pair of consecutive events. The key parameter is then the probability for each time to be

updated. That parameter will determine how big the change will be in the tree. It is clear that this move updates the total duration of the process which is important because this is what we try to estimate.

Which Move to Choose?

Now that we have three moves for the tree, an important decision is which move to choose at each step. This is chosen at random, and the three probabilities are chosen empirically; the values $.4, .3, .3$ seemed to work well in practice.

Updating the Other Parameters

We also have some other parameters to estimate, for example the sampling probability α. For example one can add to each of those parameters a random variable, uniform or Gaussian. In the case of the sampling probability α here is a useful trick: we update this parameter after each death/birth move. We look at the variation of the total number of species at the end. If this number increases, the random variable we add to α has a positive mean, if it decreases the mean is negative and if it is stable the mean is 0. Then the big moves are much more likely to be accepted and it dramatically speeds up the process.

Results

In Table 4 the summary statistics for 1200 observations (taken every 250 steps from 300,000 output values) are given. The 95% credible interval for τ is (12.3, 44.3) my, corresponding to a divergence time interval of (68.1, 99.1) my. We note that the plausible intervals are broader in the Bayesian analysis than those found in (14), and we would argue that they better reflect the variation in the estimates than the earlier approach. The upper 95% plausible value for $\bar{\alpha} = 11.6\%$, again somewhat larger than found in the original analysis.

Table 4. Summary statistics for τ, $\bar{\alpha}$ from MCMC

	τ	$\bar{\alpha}(\%)$
25th percentile	20.8	4.2
median	25.6	5.3
mean	26.6	6.1
75th percentile	31.3	7.2

In Figure 3 the estimated posterior densities for the temporal gap τ and the average sampling fraction $\bar{\alpha}$ are compared with the corresponding results for the rejection method. The main distinction is that the MCMC results have shorter right tails than the rejection results. The major differences occur in the estimated posteriors for $\bar{\alpha}$. As expected, the values of τ and $\bar{\alpha}$ are

negatively correlated (the shorter the tree the higher must be the average sampling fraction to match the data). This correlation is much stronger in the MCMC output ($\rho = -0.50$) than in the rejection method ($\rho = -0.17$). This disparity suggests that alternative metrics be tried in (2), perhaps ensuring that both terms on the right of (2) are small, as opposed to just their sum.

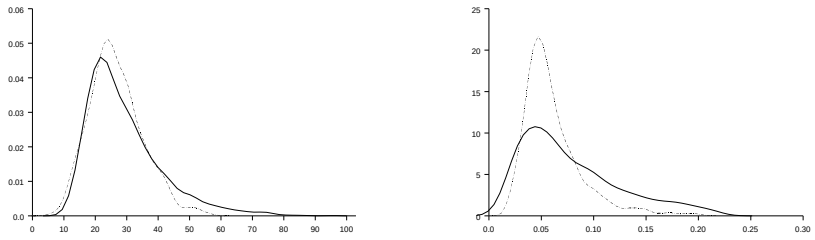

Fig. 3. Posterior densities of τ (left panel) and $\bar{\alpha}$ (right panel). Dashed line is from MCMC, solid line from rejection method.

3 Using Summary Statistics

In cases where the data $\mathcal{D}$ are complicated, it is not usually feasible to use the approach we adopted for the fossil data. Instead we can capture the information in $\mathcal{D}$ using a number of marginal statistics, denoted here by $S = (S_1, \ldots, S_p)$ and compare data sets via a comparison of these summary statistics. This approach also lends itself naturally to application in cases where $\mathcal{D}$ has continuous components.

3.1 Approximate Rejection

The rejection algorithm becomes:

1. Generate θ from $\pi(\cdot)$
2. Simulate $\mathcal{D}'$ from model $\mathcal{M}$ with parameter θ, and compute the corresponding statistics S'
3. Calculate the distance $\rho(S, S')$ between S and S'
4. Accept θ if $\rho \leq \epsilon$, and return to 1.

For examples of this approach in the population genetics literature, see (2; 16; 10; 15). In (1) a novel approach based on local non-linear regression is described. The idea is to weight all the observations generated in steps 1 and 2 above using the corresponding values of ρ.

For complex probability models sampling from the prior does not make good use of accepted observations, so the rejection methods can be prohibitively slow. In the next section we describe an MCMC approach to problems in which the likelihood cannot be readily computed.

3.2 Approximate MCMC

In (6) we present an MCMC approach that is the natural analog of the approach in the last section, in that no likelihoods are used in its implementation. It is based on the following steps:

1. If at θ propose a move to θ' according to a transition kernel $q(\theta \to \theta')$
2. Generate $\mathcal{D}'$ using model $\mathcal{M}$ with parameters θ'
3. If $\rho(S', S) \leq \epsilon$, go to 4, and otherwise stay at θ and return to 1,
4. Calculate
$$h = h(\theta, \theta') = \min\left(1, \frac{\pi(\theta')q(\theta' \to \theta)}{\pi(\theta)q(\theta \to \theta')}\right)$$
5. Move to θ' with probability h, else stay at θ; go to 1.

The stationary distribution of the chain is $f(\theta \mid \rho(S', S) \leq \epsilon)$. In the following section we illustrate this algorithm with an example from population genetics

4 Inference in Population Genetics

To illustrate these ideas, we have chosen an example of ancestral inference. We use a data set of $n = 42$ Yakima sequences obtained in (12); see also (7), p. 404. The observed base frequencies in the sequences are

$$(\pi_A, \pi_G, \pi_C, \pi_T) = (0.328, 0.113, 0.342, 0.217),$$

there are $H = 20$ distinct sequences, and $V = 31$ of the 360 base positions showed variation in the sample. These data have been discussed in a coalescent framework in (7; 8), where the posterior distribution of the (rescaled) mutation parameter θ and the height T of the coalescent tree of the sample (i.e. the time to the most recent common ancestor (MRCA) of the sample) were found by MCMC methods using the full sequence data. Further details of the coalescent model, the mutation model and its parameters may be found there.

4.1 Results

Using the Rejection Method

First we discuss inference about θ and T using the summary statistic $S = V$. This case can be analyzed using the rejection algorithm in Section 3.1, in which we take

$$\rho(V, V') = |V - V'|, \quad \epsilon = 0.$$

We chose wide uniform priors for θ, and used a constant population size coalescent to model T. The method involves simulation of the coalescent tree (and therefore its height) followed by superposition of mutations on that tree according to the given mutation model. Once the sequence data are simulated, the value of V' is computed, and the pair (θ, T) is accepted if $V' = V = 31$. We use this approach to check the behavior of the MCMC algorithm.

Using the MCMC Method

To use the MCMC approach in Section 3.2, we use the following updating strategy. For reasons of space the description is kept brief; further details may be found in (6). We keep track of the topology of the coalescent tree, the times between coalescence events in the tree, the number of mutations between these events, and the location of those mutations on the genome. The topology is updated according to the scheme described in (7). The new mutation rate is equal to the previous one plus a uniform random variable. The times between two levels are equal to the previous ones plus a Gaussian random variable. The new time and the new mutation rate produces a new intensity for the Poisson random variable which determines the number of mutations in each interval on the tree. That number is updated using basic properties of a Poisson random variable.

Results

In Table 5 we compare the output of the rejection method and the MCMC method that does not use likelihoods. As is to be expected, the algorithms produce very similar summary statistics, although the approaches differ somewhat in their acceptance rates.

Adding Further Summary Statistics

As noted in (1), the effects of using summary statistics instead of the full data can be hard to guess. In the present case, however, a full analysis is available; see Tables 1 and 2 of (7). In (7) the estimated mean of the posterior for θ was 0.044, and for T was 0.72. The effect of using only the number of variable sites is to leave the posterior height of the tree much closer to its prior, which has mean 1.95. This suggests adding another summary statistic. We present results for $S = (V, H)$, adding the number of distinct sequences H to the analysis.

A simulated data set $\mathcal{D}'$ is accepted if the number of variable sites and the number of distinct sequences match the observed numbers (31 and 20 respectively) exactly. In Table 6, a comparison of the rejection and MCMC approaches is given. The addition of the extra summary statistics has moved the approximate posterior closer to the posterior obtained using the full sequence data; the addition of H has (for example) reduced the posterior mean

Table 5. Comparison of approaches using $S = V, \epsilon = 0$

	Rej.[1]	No like. [2]
Acceptance rate	0.62%	3.9%
TMRCA T		
1st quartile	1.07	1.09
mean	1.71	1.72
median	1.49	1.49
3rd quartile	2.11	2.15
Mutation rate θ		
1st quartile	0.020	0.020
mean	0.026	0.025
median	0.024	0.024
3rd quartile	0.030	0.030

[1] Rejection method. Based on 2000 observations
[2] No likelihoods. Based on 1000 observations sampled every 30,000 steps.

of T from about 1.7 to 1.1, while the posterior mean for θ has increased from 0.026 to 0.030. This illustrates the importance of the choice of S.

Table 6. Comparison of approaches using $S = (V, H), \epsilon = 0$

	Rejection[1]	No likelihood [2]
Acceptance rate	0.03%	0.46%
TMRCA T		
1st quartile	0.73	0.74
mean	1.01	1.03
median	0.93	0.94
3rd quartile	1.22	1.24
Mutation rate θ		
1st quartile	0.025	0.025
mean	0.031	0.031
median	0.030	0.030
3rd quartile	0.036	0.036

[1] Based on 2000 accepted observations.
[2] Based on 1000 observations sampled every 50,000 steps.

5 Conclusions

This article has focussed on methods for approximating posterior distributions when likelihoods are hard or impossible to calculate. We gave examples concerning inference in the fossil record and population genetics. In both cases, the true posterior was known, so we were able to assess the adequacy of the approximate methods. In the population genetics example, in which the full data are complicated, we also summarized the data using readily computable statistics from the full data. Once more, we were able to compare these summaries with the true posterior. Typically we are not able to make such comparisons, and it is then hard to know how well the approximate methods perform. As a result, a systematic method for finding approximately sufficient statistics to summarize a given probability model is urgently required. This done, there is a chance of extending stochastic computation techniques to much more complicated problems than can now be attacked.

References

[1] M. A. Beaumont, W. Zhang, and D. J. Balding. Approximate Bayesian computation in population genetics. *Genetics*, 162:2025–2035, 2002.

[2] Y.-X. Fu and W.-H. Li. Estimating the age of the common ancestor of a sample of DNA sequences. *Mol. Biol. Evol.*, 14:195–199, 1997.

[3] W. R. Gilks, S. Richardson, and D. J. Spiegelhalter. *Markov Chain Monte Carlo in Practice*. Chapman and Hall, 1996.

[4] T. E. Harris. *The Theory of Branching Processes*. Springer Verlag, Berlin, 1963.

[5] W. K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57:97–109, 1970.

[6] P. Marjoram, J. Molitor, V. Plagnol, and S. Tavaré. Markov chain Monte Carlo without likelihoods. Submitted, 2003.

[7] L. Markovtsova, P. Marjoram, and S. Tavaré. The age of a unique event polymorphism. *Genetics*, 156:401–409, 2000a.

[8] L. Markovtsova, P. Marjoram, and S. Tavaré. The effects of rate variation on ancestral inference in the coalescent. *Genetics*, 156:1427–1436, 2000b.

[9] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A.H. Teller, and E. Teller. Equations of state calculations by fast computing machines. *J. Chem. Phys.*, 21:1087–1092, 1953.

[10] J. K. Pritchard, M. T. Seielstad, A. Perez-Lezaun, and M. W. Feldman. Population growth of human Y chromosomes: A study of Y chromosome microsatellites. *Mol. Biol. Evol.*, 16:1791–1798, 1999.

[11] B. D. Ripley. *Stochastic Simulation*. Wiley, New York, 1987.

[12] G. F. Shields, A. M. Schmeichen, B. L. Frazier, A. Redd, M. I. Vovoeda, J. K. Reed, and R. H. Ward. mtDNA sequences suggest a recent evolutionary divergence for Beringian and Northern North American populations. *Am. J. Hum. Genet.*, 53:549–562, 1993.

[13] C. Soligo, O. Will, S. Tavaré, C. R. Marshall, and R. D. Martin. New light on the dates of primate origins and divergence. In M. J. Ravosa and M. Dagosto, editors, *Primate Origins and Adaptations*. Kluwer Academic/Plenum Publishers, New York, 2003.

[14] S. Tavaré, C. R. Marshall, O. Will, C. Soligo, and R. D. Martin. Using the fossil record to estimate the age of the last common ancestor of extant primates. *Nature*, 416:726–729, 2002.

[15] J. D. Wall. A comparison of estimators of the population recombination rate. *Mol. Biol. Evol.*, 17:156–163, 2000.

[16] G. Weiss and A. von Haeseler. Inference of population history using a likelihood approach. *Genetics*, 149:1539–1546, 1998.

[17] O. Will, V. Plagnol, C. Soligo, R. D. Martin, and S. Tavaré. Statistical inference in the primate fossil record: a Bayesian approach. In preparation, 2003.