**Mark A. Beaumont** (*University of Reading*)

One of the many points raised by this important paper is that the computational techniques involved in solving these problems are complex, and independent corroboration is useful. I present some results that are relevant to one of the applications that the authors discuss.

The Markov chain Monte Carlo (MCMC) approach described in Beaumont (1999) is similar to that of Wilson and Balding (1998) but draws samples from $\pi(\theta, \mathcal{H}|A_n)$. Using a uniform (improper) prior, relative likelihoods for $\theta$ can be obtained by standard density estimation on the sampled values. Stephens and Donnelly note that the results depend on the smoothing method used (Fig. 4) and suggest (Section 3.3) that more efficient approaches are available.

A feature of using an MCMC approach that samples $\mathcal{H}$ is that it is possible to use Rao–Blackwellization (Gelfand and Smith, 1990) to estimate the marginal posterior distribution of $\theta$. It is straightforward to calculate

$$\pi(\theta|\mathcal{H}^i, A_n) = \frac{T^i}{2} \frac{(T^i\theta/2)^{m^i} \exp(-T^i\theta/2)}{m^i!},$$

where $m^i$ and $T^i$ are respectively the total number of mutations and the total branch length in the $i$th sampled genealogical history $\mathcal{H}^i$. We then estimate

$$\hat{\pi}(\theta|A_n) = \frac{1}{n} \sum^{n} \pi(\theta|\mathcal{H}^i, A_n).$$

The aim of this contribution is to obtain posterior densities for $\theta$ by using Rao–Blackwellization, and to compare the relative likelihoods with those obtained by Stephens and Donnelly (Figs 3 and 4) using the test data of Wilson and Balding (1998). In addition, likelihoods are estimated using rejection sampling as described by Beaumont (1999). The advantage of the latter approach is that sampling is independent with known error.

Five independent simulations of $10^7$ iterations were run. Trial values of $\theta$ were proposed with probability 0.05 ($\mathcal{H}$ was updated otherwise). Every 500 iterations, values of $\mathcal{H}$ and $\theta$ were sampled. Each simulation took about 5 min on a 500 MHz Pentium computer.

The estimated posterior distributions are scaled to enclose the same volume over the range $\theta = 0$–20 as the estimates from rejection sampling. There is good concordance between all the methods (Fig. 8), although there is appreciable variability between independent runs of the MCMC simulation. Estimates of the density using the program `Locfit` showed a similar degree of variability between independent runs but there was substantial oversmoothing of the lower tail. A better fit would be obtained by modifying the estimation procedure, but the Rao–Blackwellization avoids an *ad hoc* treatment. In more complex models, the conditional distributions for the parameters will be more difficult to estimate and approaches such as that suggested by Chen (1994) will be required.

**Mary K. Kuhner and Peter Beerli** (*University of Washington, Seattle*)

In their Fig. 2 Stephens and Donnelly show our Markov chain Monte Carlo (MCMC) algorithm `Fluctuate` producing curves which vary greatly between runs and are narrower than their importance sampling (IS) curve. We have independently repeated these simulations and confirm that on these data our MCMC algorithm produces unstable results even in lengthy searches. We shall discuss a possible reason for this behaviour and suggest a way to improve MCMC performance.

Stephens and Donnelly mention, but do not emphasize, a fundamental difference between current MCMC and IS approaches. (Actually both sets of methods use IS, so the term 'IS' is somewhat misleading). In MCMC sampling, the missing data of the genealogy is represented as a topology with branch lengths. The size of the search space depends on the number of branches. Short sequences lead to many possible combinations of branch lengths and make a stable estimate difficult.

In contrast, IS represents the missing data as a topology with mutations. The size of the search space is determined by the number of mutations, which depends mainly on the number of sequence positions. For short sequences the program needs to assign only a few mutations.

The data set used in Fig. 2 contains 50 individuals and only 20 sites, making it well suited to the IS approach. As Stephens and Donnelly note, MCMC sampling and IS have distinct and complementary strengths. Additional data would stabilize the MCMC estimate at little cost in speed.

Stephens and Donnelly suggest that MCMC sampling with a fixed 'driving value' may produce a likelihood curve that is too narrow. Our simulations suggest that this does not happen with estimation

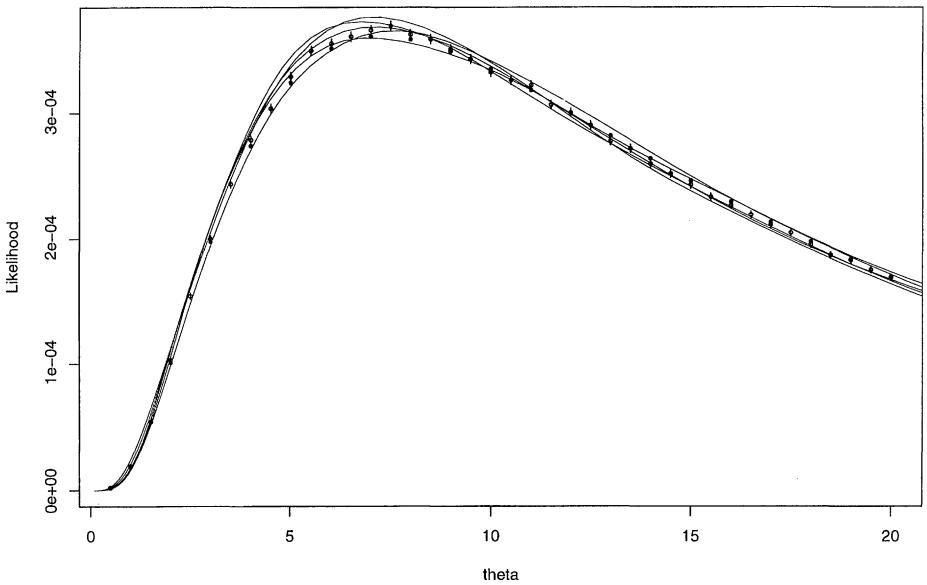


**Fig. 8.** Plot of the likelihood for different values of $\theta$: ———, scaled posterior density estimates from the MCMC simulations; ●, likelihood estimates from the method of Stephens and Donnelly; ⌽, estimates from rejection sampling, with approximate 95% confidence intervals

of $\theta$ on larger data sets (Kuhner *et al.*, 1995) but may be a problem for the co-estimation of $\theta$ and the growth rate (Kuhner *et al.*, 1998). A possible solution is to make several runs with different driving values and to combine the resulting samples by using the method of Geyer (1991). Shown in Fig. 9 is an MCMC curve (generated by the MIGRATE program of Beerli and Felsenstein (1999), without migration) for the data of Fig. 2 combining 10 independent runs at different driving values of $\theta$. The curve is now similar to the result of Stephens and Donnelly.

The following contributions were received in writing after the meeting.

**Stephen Brooks** (*University of Surrey, Guildford*) **and Andrew Gelman** (*Columbia University, New York*)
First, we congratulate the authors on a stimulating paper. Our attention was drawn in particular to Section 6.4 where we see some overlap with our own work.

One approach to detecting a lack of convergence is to estimate, using simulation, quantities that have known values under the target distribution. If $\theta$ denotes the parameter vector sampled via iterative simulation, then we can use simulation draws to estimate $E\{U(\theta)\}$ for any computable function $U$. Many diagnostic techniques are based on monitoring functions that converge to some specific value. However, in general this value is not known and so the resulting diagnostic is rather

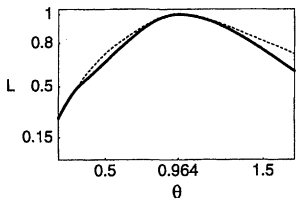


**Fig. 9.** Results from MIGRATE (———) superimposed on the Fig. 2 results of Stephens and Donnelly (·····): the MIGRATE estimate combines samples from the final long chains (100000 steps each) of 10 independent runs