

Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics

Jody Hey^{†*} and Rasmus Nielsen[‡]

[†]Department of Genetics, Rutgers, The State University of New Jersey, Piscataway, NJ 08846; and [‡]Center for Comparative Genomics, Department of Biology, University of Copenhagen, DK-1017 Copenhagen, Denmark

Communicated by Daniel L. Hartl, Harvard University, Cambridge, MA, December 15, 2006 (received for review October 17, 2006)

In 1988, Felsenstein described a framework for assessing the likelihood of a genetic data set in which all of the possible genealogical histories of the data are considered, each in proportion to their probability. Although not analytically solvable, several approaches, including Markov chain Monte Carlo methods, have been developed to find approximate solutions. Here, we describe an approach in which Markov chain Monte Carlo simulations are used to integrate over the space of genealogies, whereas other parameters are integrated out analytically. The result is an approximation to the full joint posterior density of the model parameters. For many purposes, this function can be treated as a likelihood, thereby permitting likelihood-based analyses, including likelihood ratio tests of nested models. Several examples, including an application to the divergence of chimpanzee subspecies, are provided.

speciation | population structure | divergence

Population genetic and phylogenetic models that take a genealogical (i.e., gene tree) approach suffer two nested levels of ambiguity. First, the uncertainty of an estimate of a genealogy can be large and difficult to quantify, and second, it can be difficult to interpret a genealogy estimate explicitly in terms of an evolutionary or population genetics model. In his 1988 review, Felsenstein (1) conceptualized a way thru these uncertainties by positioning the genealogy as a nuisance variable in the definition of the likelihood of the parameters given the data (proportional to the sampling probability of the data):

$$L(\Theta|X) \propto \Pr(X|\Theta) = \sum_{G \in \psi} \Pr(X|G)p(G|\Theta), \quad [1]$$

where X is the data, G is a genealogy, ψ is the set of all possible genealogies, and Θ is the vector of model parameters to be estimated. The basic idea of considering all of the possible genealogies in proportion to their probability is also contained explicitly in the recursion approach of Griffiths (2) and is suggested in much other work on genealogical models. Although Felsenstein used the notation for summation, integral forms have often since been used, reflecting the fact that genealogies are complex entities with both discrete components (branching topology) and continuous components (branch lengths).

Felsenstein's equation does not have a general closed-form solution, and numerical evaluation is difficult because of the very large number of possible tree topologies for even small data sets. It is possible to approximate Eq. 1 by simulating k independent genealogies from $p(G|\Theta)$, G_1, \dots, G_k , in which case, a simulation-consistent estimator of the likelihood can be obtained as

$$L(\Theta|X) \approx \frac{1}{k} \sum_{i=1}^k \Pr(X|G_i). \quad [2]$$

However, this is usually far too inefficient, because the variance in $\Pr(X|G)$ will be very large for randomly generated genealogies. Efficient stochastic evaluation of Eq. 1 requires the availability of methods for sampling G with some consideration of the data. For a given parameter value, Θ , the distribution of G that minimizes the simulation variance is

$$p(G|X, \Theta) = \Pr(X|G)p(G|\Theta)/\Pr(X|\Theta). \quad [3]$$

However, direct sampling from this distribution is not possible because it requires that the likelihood function can be calculated analytically (3).

Kuhner, Yamato, and Felsenstein (1995). A solution to the question of how to sample genealogies was described by Kuhner *et al.* (4), who devised a Markov chain Monte Carlo (MCMC) simulation approach. In the simulation, updates of G are accepted with probability given by the Metropolis–Hastings (5, 6) criterion

$$\min \left\{ 1, \frac{\Pr(X|G^*)p(G^*|\Theta)q(G \rightarrow G^*)}{\Pr(X|G)p(G|\Theta)q(G \rightarrow G^*)} \right\}, \quad [4]$$

where $q(G \rightarrow G^*)$ is the probability that G^* is proposed as an update from G . At stationarity, the residence time in the Markov chain will be proportional to the posterior density of that genealogy (i.e., as given by Eq. 3), and trees sampled successively from the Markov chain are correlated draws from the posterior density of genealogies.

The approach devised by Kuhner *et al.* and used thereafter for a variety of models (7–10) is to use the genealogies that have been sampled by using one parameter value, Θ_0 , to estimate the relative likelihood for other values. The likelihood surface for Θ is obtained by running a Markov chain at a fixed value Θ_0 close to the mode of the likelihood function while evaluating the likelihood for multiple values of Θ by using importance sampling (11). Let $p(G|\Theta_0)$ be $p(G|\Theta)$ evaluated at the point $\Theta = \Theta_0$, and assume $p(G|\Theta_0) > 0$ if $p(G|\Theta) > 0$ for all Θ and G . Then from Eqs. 1 and 3, we see that

$$\begin{aligned} \frac{\Pr(X|\Theta)}{\Pr(X|\Theta_0)} &= \int_{\psi} \frac{\Pr(X|G)w(\Theta, \Theta_0, G)p(G|\Theta_0)}{\Pr(X|\Theta_0)} dG \\ &= E_{G|X, \Theta_0}[w(\Theta, \Theta_0, G)], \end{aligned} \quad [5]$$

Author contributions: J.H. and R.N. designed research; J.H. performed research; J.H. and R.N. contributed new reagents/analytic tools; J.H. analyzed data; and J.H. and R.N. wrote the paper.

The authors declare no conflict of interest.

Abbreviation: MCMC, Markov chain Monte Carlo.

*To whom correspondence should be addressed at: Department of Genetics, Rutgers, the State University of New Jersey, 604 Allison Road, Piscataway, NJ 08846. E-mail: hey@biology.rutgers.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0611164104/DC1.

© 2007 by The National Academy of Sciences of the USA

where $w_i(\Theta, \Theta_0, G) = p(G|\Theta)/p(G|\Theta_0)$. A single set of values of G are drawn from $p(G|X, \Theta_0)$ (i.e., by sampling from the Markov chain) and used to estimate the relative likelihoods for other values of Θ :

$$\frac{L(\Theta|X)}{L(\Theta_0|X)} \approx \frac{1}{k} \sum_{i=1}^k w_i(\Theta, \Theta_0, G^i), \quad [6]$$

where $w_i(\Theta, \Theta_0, G)$ and $G^{(i)}$ are the value of $w_i(\Theta, \Theta_0, G)$ and G , respectively, in the i th sampled step of the chain.

The method of Kuhner *et al.* (4) was the first true MCMC method in population genetics, and it showed that likelihood inference of population genetics parameters is possible for complex mutational models. It led to the development of related methods (12–14), and it preceded the use of closely related MCMC methods for phylogenetic inference (15–17). However, it suffers the significant shortcoming that the distribution of $w_i(\Theta, \Theta_0, G)$ will be very skewed when Θ differs from Θ_0 , causing the variance of the estimate of the likelihood to be very large and difficult to estimate when $|\Theta - \Theta_0|$ is large (14, 18). Because of the skewed distribution of $w_i(\Theta, \Theta_0, G)$, the method will tend to underestimate the likelihood when Θ differs from Θ_0 and thus bias the estimator toward values close to Θ_0 . Kuhner *et al.* (4) address the problem of large variance when $|\Theta - \Theta_0|$ is large by running multiple chains and updating Θ_0 each time the chain is restarted (19).

An alternative to MCMC sampling of genealogies is the sequential importance sampling method of Griffiths and Tavaré (20, 21). Stephens and Donnelly (22) suggested a modification of the approach of Griffiths and Tavaré that samples more efficiently from an approximation to Eq. 3.

Bayesian MCMC. One way to extend the MCMC approach to generating likelihood surfaces is to explicitly consider a prior distribution of Θ , $p(\Theta)$, and to simulate a Markov chain with stationary measure given by the joint posterior density of G and Θ ,

$$p(G, \Theta|X) = \Pr(X|G)p(G|\Theta)p(\Theta)/\Pr(X) \quad [7]$$

(12, 14, 23). This approach, of running a Markov chain over a state space of genealogies and model parameters, has been extended to multilocus applications for a variety of models (24–27). Apart from having a large state space and associated MCMC mixing challenges, the main shortcomings stem from the essential form of the result, which is not a function estimate but merely a record of parameter values. Density estimates can be obtained by binning or by kernel estimators, but the nature of the results effectively precludes estimates of the joint posterior density for models with more than a small number of parameters. In such cases the volume of the parameter space is so large that the number of recorded values that will fall in any portion of it may be low, even for very long runs and even for portions of the parameter space associated with high posterior densities. Because of this “curse of dimensionality,” the number of samples needed increases exponentially with the dimension (28). This means that applications have mostly been limited to the generation of estimates of the marginal posterior densities for each of the model parameters. It also means that it has been difficult to estimate likelihood ratios for models involving several parameters. Here, we propose a method that eliminates the need for a driving value (Θ_0) and that generates estimates of the entire posterior probability density function, suitable for optimization and likelihood-ratio tests of nested models.

Theory. This approach relies on the analytical calculation of the prior probability of G by integration of $p(G|\Theta)$ over the prior distribution of Θ . This makes it possible to draw samples by MCMC directly from the marginal posterior probability of genealogies,

$p(G|X)$. Then, using a sample of these genealogies, one can construct an estimate of the posterior density function, $p(\Theta|X)$.

As we will show, the calculation of the marginal prior density of genealogies,

$$p(G) = \int p(G|\Theta)p(\Theta)d\Theta, \quad [8]$$

can be done analytically and easily when $p(\Theta)$ has a uniform distribution. Access to the prior for G permits an MCMC simulation that has a marginal posterior density for G given by

$$p(G|X) = \Pr(X|G)p(G)/p(X) \quad [9]$$

(contrast with Eq. 7). Then, the posterior density for Θ is given by

$$p(\Theta|X) = \int_{\psi} p(\Theta|G)p(G|X)dG \quad [10]$$

(contrast with Eq. 1). This can be proved by noting that under the proper scaling, $p(X|G, \Theta) = \Pr(X|G)$ (4). Then $p(\Theta|G, X) = p(\Theta|G)p(G|X)/\Pr(X|G) = p(\Theta|G)$, and Eq. 10 follows by the law of total probability.

The posterior density of Θ can then be consistently estimated as

$$p(\Theta|X) \approx \frac{1}{k} \sum_{i=1}^k p(\Theta|G_i), \quad [11]$$

where $G_i, i = 1, 2, \dots, k$, are the samples from $p(G|X)$ that are generated by the MCMC simulation. Inferences can then be based on $p(\Theta|X)$ or on the likelihood function deduced as $L(\Theta) \propto p(\Theta|X)p(\Theta)$. If $p(\Theta)$ is a constant, then the posterior probability is directly proportional to the likelihood over the prior range of Θ . In effect, a Bayesian sampling strategy is being used to generate an estimate of the relative likelihood, which can be used in turn to find a maximum likelihood estimate of Θ and to conduct other likelihood-based analyses. It is also useful to note that $\Pr(X|G)$ is not part of the final calculation in Eq. 11. As in the method of Kuhner *et al.*, (4) the data are used to determine the probability density from which the genealogies are sampled and thereafter are not required (29).

A Single-Population Model. Consider a model in which Θ includes just one parameter, $\theta = 4Nu$, and a sample of n gene copies, for a locus with neutral mutation rate u , drawn from a population with effective chromosomal population size $2N$ and evolving according to Kingman’s coalescent (30). Letting the coalescent times in the genealogy be $\tau = \{\tau_2, \dots, \tau_n\}$, where τ_i is the time interval in G in which there are i ancestors of the sample, then

$$p(G|\theta) = \left(\frac{2}{\theta}\right)^{n-1} \exp(-2f_n/\theta), \quad [12]$$

where

$$f_n = \sum_{i=2}^n \tau_i(i-1)$$

is the total coalescent rate measured over the genealogy (30, 31). If we consider a uniform prior distribution for θ over the interval $\{0, \theta_{max}\}$, then placing Eq. 12 into Eq. 8 yields

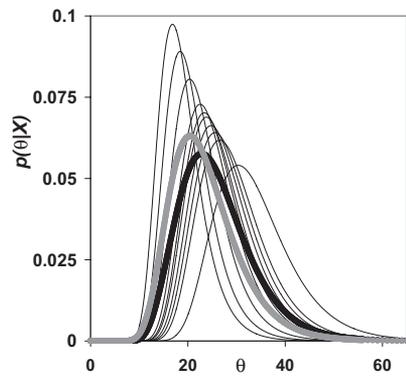


Fig. 1. A single population model was simulated with a sample of 20 gene copies drawn from a population with a true population mutation rate of $\theta = 20$. Functions based on 10 individual sampled genealogies are shown along with their mean (gray line) as well as the mean for 100 sampled genealogies (black line).

mate of the marginal posterior density for t . Also described in *SI Text* is a method for considering data from multiple loci that vary in their neutral mutation rates.

Implementation and Examples. A computer program was written that implements a Markov chain simulation for generating samples from $p(G|X)$ for models with one or two populations, as well as for a two-population isolation with migration model (i.e., with a population-splitting time parameter, t). The state space of the Markov chain includes the prior distribution of G (and t if population splitting is in the model), with a general Metropolis–Hastings update criterion

$$\min \left\{ 1, \frac{\Pr(X|G^*, t^*)p(G^*, t^*)q(G^*, t^* \rightarrow G, t)}{\Pr(X|G, t)p(G)q(G, t \rightarrow G^*, t^*)} \right\}. \quad [21]$$

The update of G to G^* is done by using branch sliding (14) in which a randomly selected branch is moved a random distance in the tree. The migration events originally on the branch are removed, and a random number of new migration events is drawn from a Poisson distribution, conditioned on there being an even or odd number of migration events (depending on whether the starting and ending populations of the branch are the same). The Poisson parameter is taken to be the expected number of migrations over the span of the new branch length, given the current number of migration events that occur over the total length of the tree.

If the model includes t , then it is also necessary to do joint updates of G and t . For these updates, we follow the method of Rannala and Yang (34), in which the new value, t^* , is drawn from a uniform distribution over the interval $\{0, t_{\max}\}$, and the times of all migration and coalescent events in G before t are multiplied by t^*/t , and the times of events after t are summed with $(t^* - t)$. At evenly spaced intervals, records are made of $t, p(G, t)$, and of those quantities from G that are needed to calculate $p(G|\Theta, t)$. For the case of multiple loci, the updates of the mutation rate scalars are done as in Hey and Nielsen (25).

In general, it is expected that each genealogy will make its greatest contribution to the overall probability over some limited range of Θ . By including a large number of genealogies, sampled from a long-running, well-mixing Markov chain that has reached stationarity, it should be possible to obtain good estimates of $p(\Theta|X)$ for any value of Θ . Optimization of the estimate function, under full or nested models, requires some care because the surface may be multimodal over broad and fine scales, either because of the data or because of the particular genealogies that happened to end up

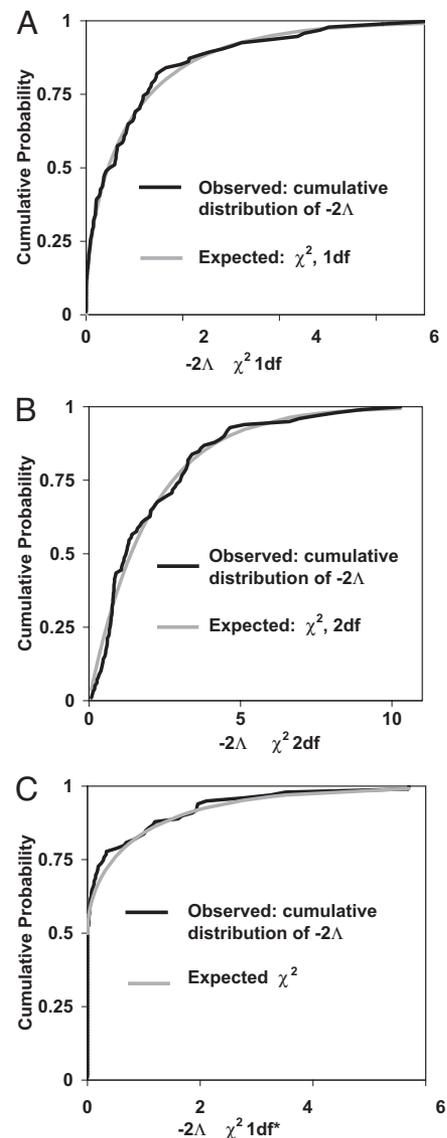


Fig. 2. Cumulative distributions of $-2\hat{\lambda}$ calculated from Markov chain simulations run on 100 simulated data sets. (A) The true model has two populations of identical size, with simulated data sets sampled following: $\theta_1 = \theta_2 = 10$, $m_1 = 0.5$, $m_2 = 0.01$, six loci, and 15 gene copies sampled for each population per locus. The MCMC simulation generated 5,000 genealogies with prior maxima: $\theta_1 = \theta_2 = 100$, $m_1 = m_2 = 5$. The observed cumulative distribution of $-2\hat{\lambda}$ is shown with that expected from a χ^2 distribution with 1 degree of freedom. The Kolmogorov–Smirnov test statistic (the greatest departure between two cumulative distributions), is 0.0851, which does not approach statistical significance. (B) The true model has two populations with identical migration rates and population sizes with simulated data sets sampled following: $\theta_1 = \theta_2 = 20$, $m_1 = m_2 = 0.1$, 25 loci, and six gene copies sampled for each population per locus. The Markov chain simulations generated samples of 8,000 genealogies with prior maximum values as follows: $\theta_1 = \theta_2 = 500$, $m_1 = m_2 = 1$. The Kolmogorov–Smirnov test statistic is 0.0791, which does not approach statistical significance. (C) The true model is an isolation with migration model with unidirectional gene flow. Simulated data sets were sampled following: $\theta_1 = 20$; $\theta_2 = 40$, $\theta_A = 30$, $m_1 = 0.15$, $m_2 = 0$, $t = 10$, 25 loci, and six gene copies sampled for each population per locus. The estimated posterior density has four dimensions [for $\theta_1, \theta_2, \theta_A$, and m_1 (note that $m_2 = 0$)]. Given one parameter fixed at the boundary of the parameter space, the likelihood ratio statistic, contrasting the true model and the full five-parameter model, should be asymptotically distributed as a random variable that takes the value 0 with probability 0.5 and takes on a value from a χ^2_1 distribution with probability 0.5 (23, 37). The Kolmogorov–Smirnov test statistic is 0.0590, which does not approach statistical significance.

Table 1. Tests of nested models for *P. t. troglodytes* and *P. t. verus*

Model (Θ)	$\log(p'(\hat{\Theta} X))$	$-2\hat{\Lambda}$	P	df
$\theta_1 \theta_2 \theta_A m_1 m_2$	7.734	—	—	—
$\theta_1 \theta_2 \theta_A m_1 = 0 m_2$	5.976	3.516	0.03039*	1†
$\theta_1 \theta_2 \theta_A m_1 m_2 = 0$	7.131	1.206	0.13606	1†
$\theta_1 \theta_2 \theta_A m_1 = m_2$	6.403	2.663	0.10271	1
$\theta_1 = \theta_2 \theta_A m_1 m_2$	-24.925	65.319	6.4×10^{-16} *	1
$\theta_1 = \theta_2 = \theta_A m_1 m_2$	-38.951	93.37	5.3×10^{-21} *	2
$\theta_1 = \theta_2 \theta_A m_1 = m_2$	-31.34	78.148	1.1×10^{-17} *	2
$\theta_1 = \theta_2 = \theta_A m_1 = m_2$	-60.058	135.584	3.4×10^{-29} *	3
$\theta_2 \theta_1 = \theta_A m_1 m_2$	1.548	12.374	0.00043*	1
$\theta_2 \theta_1 = \theta_A m_1 = m_2$	0.097	15.274	0.00048*	2
$\theta_1 \theta_2 = \theta_A m_1 m_2$	4.73	6.01	0.01423*	1
$\theta_1 \theta_2 = \theta_A m_1 = m_2$	3.131	9.207	0.01002*	2

*The probability of achieving the test statistic by chance under the null model is <0.05 .

†When the null model is true and has a parameter fixed at the boundary of the parameter space, the expected distribution is a mixture. In the case of a single fixed parameter, $-2\hat{\Lambda}$ should be asymptotically distributed as a random variable that takes the value 0 with probability 0.5 and takes on a value from a χ^2_1 distribution with probability 0.5 (23, 37).

simulated under a particular model, Θ_r . For each data set, a Markov chain simulation was run to generate an estimate of the posterior density function under the full model, $p'(\Theta|X)$. This function was maximized over all parameters to generate $p'(\hat{\Theta}|X)$ and then maximized over just those parameters that were free to vary in Θ_r to generate $p'(\hat{\Theta}_r|X)$, and $-2\hat{\Lambda}$ was calculated. Fig. 2 shows the resulting cumulative distributions for three different models, each of which is consistent with the corresponding χ^2 distribution, showing that the asymptotic result holds approximately for these moderately sized simulated data sets and that the added simulation variance introduced by the method does not invalidate the use of the classical likelihood-ratio tests. Additionally, the good fit of the χ^2 distribution suggests that the estimation and optimization of the likelihood surface is reasonable accurate. Other simulations with small data sets do show that, as expected with less data, that the distribution of $-2\hat{\Lambda}$ will have a variance larger than that for the corresponding χ^2 distribution.

Chimpanzee case study. To demonstrate the approach for a model in which an ancestral populations splits into two, we considered the case of two chimpanzee subspecies, *Pan troglodytes troglodytes* (the Central African Chimpanzee) and *Pan troglodytes verus* (the Western African Chimpanzee). This divergence has previously been

studied by using a Markov chain simulation in which the state space includes both genealogies and model parameters for a data set of 48 genes drawn from the literature (36).

Fig. 3 shows the marginal posterior density estimates from the original method (36), which generates histogram-based estimates, and the new method. As expected, both sets of marginal density estimates are very similar. Fig. 4 shows examples of contour plots of marginal posterior density estimates for pairs of parameters.

Table 1 shows the likelihood ratio statistic for a series of nested models applied to the chimpanzee data. All of the ratio statistics were calculated as the difference between the highest posterior probability for the full model and the highest posterior probability for the nested model. Only two models were not rejected: the model in which the two migration rates are equal to each other and the one in which m_2 is equal to 0. If we were to correct for multiple tests, then other models would also not be rejected.

Discussion

Felsenstein's equation has become a centerpiece of modern population genetics and phylogenetic analysis as computational approaches have been developed for faster and improved approximate solutions. Here, we describe an approach that provides greatly improved access to a broad family of population genetics models, i.e., those that can be described with one or more population size and migration parameters. Relying on a Markov chain simulation, the state space is limited to just the posterior density of genealogies, thereby avoiding those MCMC mixing problems that arise because of correlations between G and Θ , when both are part of the state space (25). In addition, the method provides a convenient approach for estimating likelihood ratios.

The finding that the estimate of the likelihood ratio, from nested models, closely approximates the χ^2 distribution that is expected under asymptotic assumptions is strong affirmation of the validity of the approach, and it means that the method can be used for many questions that involve a contrast of different demographic models. Model selection and testing of demographic hypotheses based on the full-likelihood function have often been neglected in the fields of molecular ecology and population genetics because appropriate tools for calculating likelihood ratios have not been available. The methods described here should greatly alleviate this problem by providing a powerful computational framework for estimating likelihood functions and likelihood ratios.

We thank Yong Wang, David Ruppert, and Naomi Altman for helpful discussions. This work was supported in part by a National Science Foundation grant (to J.H.) and by grants from Danmarks Grundforskningsfond and the Danish Forskningsrådet for Natur og Univers (to R.N.).

- Felsenstein J (1988) *Annu Rev Genet* 22:521–565.
- Griffiths RC (1989) *J Math Biol* 27:667–680.
- Stephens M (2001) in *Handbook of Statistical Genetics*, eds Balding DJ, Bishop M, Cannings C (Wiley, West Sussex, UK).
- Kuhner MK, Yamato J, Felsenstein J (1995) *Genetics* 140:1421–1430.
- Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E (1953) *J Chem Phys* 21:1087–1091.
- Hastings WK (1970) *Biometrika* 57:97–109.
- Kuhner MK, Yamato J, Felsenstein J (1998) *Genetics* 149:429–434.
- Beerli P, Felsenstein J (1999) *Genetics* 152:763–773.
- Kuhner MK, Yamato J, Felsenstein J (2000) *Genetics* 156:1393–1401.
- Beerli P, Felsenstein J (2001) *Proc Natl Acad Sci USA* 98:4563–4568.
- Thompson EA, Guo SW (1991) *IMA J Math App Med Biol* 8:149–169.
- Wilson IJ, Balding DJ (1998) *Genetics* 150:499–510.
- Beaumont MA (1999) *Genetics* 153:2013–2029.
- Nielsen R (2000) *Genetics* 154:931–942.
- Yang Z, Rannala B (1997) *Mol Biol Evol* 14:717–724.
- Larget B, Simon DL (1999) *Mol Biol Evol* 16:750–759.
- Mau B, Newton MA, Larget B (1999) *Biometrics* 55:1–12.
- Stephens M (1999) *Bulletin of the 52nd Session of the International Statistical Institute Book* 1:273–276.
- Geyer CJ, Thompson EA (1992) *J R Stat Soc B* 54:567–699.
- Griffiths RC, Tavaré, S (1994) *Theor Popul Biol* 46:131–159.
- Griffiths RC, Tavaré, S (1994) *Stat Sci* 9:307–319.
- Stephens M, Donnelly P (2000) *J R Stat Soc B* 62:605–655.
- Nielsen R, Wakeley J (2001) *Genetics* 158:885–896.
- Hey J, Won Y-J, Sivasundar A, Nielsen R, Markert JA (2004) *Mol Ecol* 13:909–919.
- Hey J, Nielsen R (2004) *Genetics* 167:747–760.
- Palsbøll PJ, Berube M, Aguilar A, Notarbartolo-Di-Sciara G, Nielsen R (2004) *Evol Int J Org Evol* 58:670–675.
- Hey J (2005) *PLoS Biol* 3:0965–0975.
- Scott DW (1992) *Multivariate Density Estimation: Theory, Practice, and Visualization* (Wiley, New York).
- Felsenstein J, Kuhner MK, Yamato J, Beerli P (1999) in *Statistics in Genetics and Molecular Biology*, ed Seillier-Moiseiwitsch F (Inst Math Stat and Am Math Soc, Hayward, CA), Vol 33, pp 163–185.
- Kingman JFC (1982) *Stochastic Processes App* 13:235–248.
- Felsenstein J (1992) *Genet Res (Cambridge, UK)* 60:209–220.
- Maruyama T (1970) *Theor Popul Biol* 1:273–306.
- Wright S (1931) *Genetics* 16:97–159.
- Rannala B, Yang Z (2003) *Genetics* 164:1645–1656.
- Press WH (1992) *Numerical Recipes in C: The Art of Scientific Computing* (Cambridge Univ Press, Cambridge, UK).
- Won YJ, Hey J (2005) *Mol Biol Evol* 22:297–307.
- Chernoff H (1954) *Ann Math Stat* 25:573–578.