Monte Carlo Evaluation of the Likelihood for N_e From Temporally Spaced Samples

Eric C. Anderson,* Ellen G. Williamson† and Elizabeth A. Thompson‡

**Interdisciplinary Program in Quantitative Ecology and Resource Management, University of Washington, Seattle, Washington 98195,*
†*Department of Integrative Biology, University of California, Berkeley, California 94720-3141 and*
‡*Department of Statistics, University of Washington, Seattle, Washington 98195*

Manuscript received December 27, 1999
Accepted for publication August 7, 2000

ABSTRACT

A population's effective size is an important quantity for conservation and management. The effective size may be estimated from the change of allele frequencies observed in temporally spaced genetic samples taken from the population. Though moment-based estimators exist, recently Williamson and Slatkin demonstrated the advantages of a maximum-likelihood approach that they applied to data on diallelic genetic markers. Their computational methods, however, do not extend to data on multiallelic markers, because in such cases exact evaluation of the likelihood is impossible, requiring an intractable sum over latent variables. We present a Monte Carlo approach to compute the likelihood with data on multiallelic markers. So as to be computationally efficient, our approach relies on an importance-sampling distribution constructed by a forward-backward method. We describe the Monte Carlo formulation and the importance-sampling function and then demonstrate their use on both simulated and real datasets.

REDUCTIONS in population size can lead to inbreeding, which increases the probability of population extinction in typically outbreeding species (Frankham 1995). Reductions in population size also lead to a loss of genetic diversity, which may restrict a population's ability to adapt to changing conditions (Soulé 1986). To predict the risk to a population from these types of genetic factors, biologists are often interested in knowing the effective population size, N_e. An effective size is defined by comparison to an ideal population model, the Wright-Fisher model. The Wright-Fisher model assumes discrete, nonoverlapping generations of constant size, and it assumes that the gametes that unite to form adults in one generation are randomly sampled with replacement from the previous generation. The variance effective size of a natural population is the size of a Wright-Fisher population that would experience a comparable increase in variance of gene frequency over time. The inbreeding effective size is defined similarly, but is based on the increase in gene identity by descent over time.

It is possible to estimate the variance effective size from observed changes in allele frequencies in a population over time. Moment-based estimators using *F*-statistics have been developed for this purpose (Krimbas and Tsakas 1971; Nei and Tajima 1981; Pollak 1983; Waples 1989; Jorde and Ryman 1995). Recently, Williamson and Slatkin (1999) described a method to estimate N_e by the method of maximum likelihood. To find the maximum-likelihood estimate $\widehat{N}_e$ of N_e, given allele frequencies observed in samples taken from a population at different times, one models the population underlying the samples as a Wright-Fisher population. $\widehat{N}_e$ is then the size of that underlying, ideal population for which the observed data are most probable. In simulation studies Williamson and Slatkin (1999) showed that the maximum-likelihood estimator outperformed the moment-based estimators, and they also demonstrated how a likelihood approach may be extended to estimate parameters in more complex population models.

This likelihood method has been restricted to data on diallelic loci, because, with data on multiallelic loci, evaluating the likelihood for N_e exactly is computationally intractable. Here we describe the problem as one of inference from a hidden Markov chain (Baum *et al.* 1970). We develop an algorithm for importance sampling, which makes it possible to compute the likelihood by Monte Carlo.

FORMULATION OF THE MODEL AND MONTE CARLO

The model: The data are genetic samples collected at different generations. The first sample is collected at generation 0 and the last sample at generation T. Any samples drawn at intervening generations may be evenly or irregularly spaced in time. For notational simplicity, we assume for now that individuals are genotyped at a single locus, though we describe later the extension to multiple, independently segregating loci. The data

Corresponding author: Eric C. Anderson, Department of Statistics, University of Washington, Box 354322, Seattle, WA 98195.
E-mail: eriq@cqs.washington.edu

include K different allelic types, indexed by $k = 1, \ldots, K$. The allele frequencies observed in samples taken from different generations will differ due to genetic drift and sampling variation.

Let $\boldsymbol{Y}_t = (Y_{t,1}, \ldots, Y_{t,K})$ be the counts of the K different allelic types in the sample at generation t, and let S_t denote the number of diploid individuals in the sample. We assume that the samples were taken from a Wright-Fisher population of size N_e and denote the unobserved population allele counts at generation t by $\boldsymbol{X}_t = (X_{t,1}, \ldots, X_{t,K})$, with $\Sigma_{k=1}^{K} X_{t,k} = 2N_e$. By the formulation of the Wright-Fisher model, the $\boldsymbol{X}_t$ form a Markov chain in time, with transitions defined by multinomial probabilities depending on N_e,

$$P_{N_e}(\boldsymbol{X}_t|\boldsymbol{X}_0, \ldots, \boldsymbol{X}_{t-1}) = P_{N_e}(\boldsymbol{X}_t|\boldsymbol{X}_{t-1}) = (2N_e)! \prod_{k=1}^{K} \frac{[X_{t-1,k}/(2N_e)]^{X_{t,k}}}{X_{t,k}!}, \quad t = 1, 2, \ldots. \tag{1}$$

The genetic sample at a time t is assumed to be drawn with replacement from the copies of alleles present in the population at time t. This is equivalent to drawing the sample $\boldsymbol{Y}_t$ from a very large gamete pool produced by the population at time t: sampling plan II of Waples (1989). This type of sampling applies to many organisms, especially those species with high fecundity that may be sampled as juveniles, or those that may be sampled (preferably noninvasively) as adults in populations having census sizes considerably larger than their effective sizes (Waples 1989). The sample allele counts $\boldsymbol{Y}_t$, given the latent variable $\boldsymbol{X}_t$, are conditionally independent of all the other variables and follow the multinomial distribution depending on the parameter N_e, the sample size S_t, and $\boldsymbol{X}_t$,

$$P_{N_e}(\boldsymbol{Y}_t|\boldsymbol{X}_t) = (2S_t)! \prod_{k=1}^{K} \frac{[X_{t,k}/(2N_e)]^{Y_{t,k}}}{Y_{t,k}!}, \tag{2}$$

when $S_t > 0$. If there is no sample taken from the population at generation t, then $S_t \equiv 0$, and we define $P_{N_e}(\boldsymbol{Y}_t|\boldsymbol{X}_t) \equiv 1$.

Such a system forms a hidden Markov chain with the dependence structure shown in the directed graph of Figure 1. The allele counts in the population when the first sample is drawn, $\boldsymbol{X}_0$, are nuisance parameters. To avoid estimating $\boldsymbol{X}_0$ and the associated consistency prob-

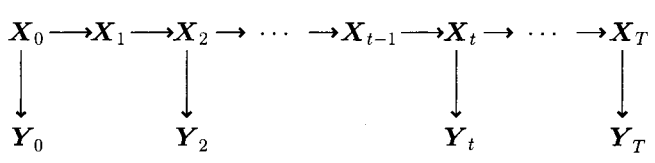


Figure 1.—A directed graph showing the dependence structure of the components of $\mathbf{X}$ and $\mathbf{Y}$. The $\boldsymbol{Y}_t$'s are observations of a hidden Markov chain. The graph shown represents a situation where samples were taken at generations 0, 2, t, and T, and no samples were taken at generations 1 and $t - 1$.

lems with the maximum-likelihood estimator (Neyman and Scott 1948), we consider the integrated likelihood, assuming a uniform prior distribution, $\pi(\boldsymbol{X}_0)$, on the population allele counts at time 0. The likelihood for N_e is the probability of the data $\mathbf{Y} = (\boldsymbol{Y}_0, \ldots, \boldsymbol{Y}_T)$ given the parameter N_e. The probability of $\mathbf{Y}$ is the sum of the joint probability of $\mathbf{Y}$ and the latent variables $\mathbf{X} = (\boldsymbol{X}_0, \ldots, \boldsymbol{X}_T)$ over the space of all $\mathbf{X}$:

$$P_{N_e}(\mathbf{Y}) = \sum_{\mathbf{X}} P_{N_e}(\mathbf{Y}, \mathbf{X}) = \sum_{\boldsymbol{X}_0, \ldots, \boldsymbol{X}_T} \left(\pi(\boldsymbol{X}_0) P_{N_e}(\boldsymbol{Y}_0|\boldsymbol{X}_0) \prod_{t=1}^{T} P_{N_e}(\boldsymbol{X}_t|\boldsymbol{X}_{t-1}) P_{N_e}(\boldsymbol{Y}_t|\boldsymbol{X}_t) \right). \tag{3}$$

For the case of $K = 2$ and N_e small, the likelihood in (3) may be computed exactly. Williamson and Slatkin (1999) effected the summation in (3) in terms of multiplication of transition probability matrices. The dimension of the square matrices is $(N_e - 1)!/[(N_e - K)!(K - 1)!]$, which increases rapidly with N_e and K. We note that the hidden Markov form of the system allows a more efficient computation of the likelihood using the algorithm of Baum (1972). Nonetheless, exact evaluation for multiple alleles would still require prohibitively large amounts of computation and storage. An alternative is to estimate $P_{N_e}(\mathbf{Y})$ by Monte Carlo.

Monte Carlo evaluation: For likelihood inference, we must evaluate $P_{N_e}(\mathbf{Y})$ for a number of different values of N_e. Expressing this probability as an expectation with respect to the distribution of $\mathbf{X}$ gives

$$P_{N_e}(\mathbf{Y}) = \sum_{\mathbf{X}} P_{N_e}(\mathbf{Y}, \mathbf{X}) = \sum_{\mathbf{X}} P_{N_e}(\mathbf{Y}|\mathbf{X}) P_{N_e}(\mathbf{X}) = \mathbb{E}_{N_e}(P_{N_e}(\mathbf{Y}|\mathbf{X})). \tag{4}$$

In this form the expectation would be taken over the marginal probabilities of $\mathbf{X}$, and it could be estimated by Monte Carlo as

$$P_{N_e}(\mathbf{Y}) \approx \frac{1}{m} \sum_{i=1}^{m} P_{N_e}(\mathbf{Y}|\mathbf{X}^{(i)}) \tag{5}$$

for large m, with $\mathbf{X}^{(i)}$ being the ith realization from the marginal distribution of $\mathbf{X}$. Such a naive scheme fails, however, because $P_{N_e}(\mathbf{Y}|\mathbf{X}^{(i)})$ varies greatly over the values of $\mathbf{X}$ realized from their marginal distribution, resulting in enormous Monte Carlo variance.

Instead, we pursue a more efficient Monte Carlo approximation by using importance sampling (Hammersley and Handscomb 1964). We express $P_{N_e}(\mathbf{Y})$ as an expectation with respect to a different distribution $P^{*}_{N_e}(\mathbf{X})$ having the property that $P^{*}_{N_e}(\mathbf{X}) > 0$ for all $\mathbf{X}$ such that $P_{N_e}(\mathbf{Y}, \mathbf{X}) > 0$. Thus, we have

$$P_{N_e}(\boldsymbol{Y}) = \sum_{\mathbf{X}} \frac{P_{N_e}(\mathbf{Y}, \mathbf{X})}{P^{*}_{N_e}(\mathbf{X})} P^{*}_{N_e}(\mathbf{X}) = \mathbb{E}^{*}_{N_e}\left(\frac{P_{N_e}(\mathbf{Y}, \mathbf{X})}{P^{*}_{N_e}(\mathbf{X})} \right), \tag{6}$$

where $\mathbb{E}^{*}_{N_e}$ indicates that the expectation is over the space

of $\mathbf{X}$ weighted by the distribution $P^*_{N_e}(\mathbf{X})$. The expectation (6) may be estimated by Monte Carlo, giving

$$P_{N_e}(\mathbf{Y}) \approx \tilde{P}_{N_e}(\mathbf{Y}) = \frac{1}{m}\sum_{i=1}^{m}\frac{P_{N_e}(\mathbf{Y}, \mathbf{X}^{(i)})}{P^*_{N_e}(\mathbf{X}^{(i)})} \tag{7}$$

for large m, where $\mathbf{X}^{(i)}$ is the ith realization of $\mathbf{X}$ drawn from $P^*_{N_e}(\mathbf{X})$. The Monte Carlo variance of $\tilde{P}_{N_e}(\mathbf{Y})$ is made small when $P_{N_e}(\mathbf{Y}, \mathbf{X})/P^*_{N_e}(\mathbf{X})$ varies little across the possible values of $\mathbf{X}$ and would be minimized if $P^*_{N_e}(\mathbf{X})$ were exactly proportional to $P_{N_e}(\mathbf{Y}, \mathbf{X})$. Such a distribution of $\mathbf{X}$ would, by definition, be the conditional distribution $P_{N_e}(\mathbf{X}|\mathbf{Y})$. Unfortunately, for the same reasons that $P_{N_e}(\mathbf{Y})$ cannot be computed exactly, it is infeasible to compute $P_{N_e}(\mathbf{X}|\mathbf{Y})$. Nonetheless, the Monte Carlo variance of $\tilde{P}_{N_e}(\mathbf{Y})$ will be reduced to the extent that $P^*_{N_e}(\mathbf{X})$ resembles $P_{N_e}(\mathbf{X}|\mathbf{Y})$. We now describe a method for rapid simulation of $\mathbf{X}^{(i)}$'s from a distribution $P^*_{N_e}(\mathbf{X})$ that is close to $P_{N_e}(\mathbf{X}|\mathbf{Y})$. As is required for the importance sampling, it is also possible to compute $P^*_{N_e}(\mathbf{X}^{(i)})$ quickly for each $\mathbf{X}^{(i)}$ generated.

Sampling from $P^*_{N_e}(\mathbf{X})$ by a forward-backward method: Baum *et al.* (1970) describe a method applicable to general, hidden Markov chains for realizing latent variables, such as $\mathbf{X} = (\boldsymbol{X}_0, \ldots, \boldsymbol{X}_T)$, from their exact conditional distribution given the observed variables, such as $\mathbf{Y} = (\boldsymbol{Y}_0, \ldots, \boldsymbol{Y}_T)$. Their algorithm first employs a "forward step" in which the conditional probability distributions of each $\boldsymbol{X}_t$, given the observed variables up to and including $\boldsymbol{Y}_t$, are recursively computed and stored using the relation

$$P(\boldsymbol{X}_t|\boldsymbol{Y}_0, \ldots, \boldsymbol{Y}_t) \propto \sum_{\mathbf{X}_{t-1}} P(\boldsymbol{X}_{t-1}|\boldsymbol{Y}_0, \ldots, \boldsymbol{Y}_{t-1})P(\boldsymbol{X}_t|\boldsymbol{X}_{t-1})P(\boldsymbol{Y}_t|\boldsymbol{X}_t), \tag{8}$$

which is normalized by the sum of that quantity over all the values of $\mathbf{X}_t$. The last such conditional distribution computed is $P(\boldsymbol{X}_T|\boldsymbol{Y}_0, \ldots, \boldsymbol{Y}_T)$. The "backward step" begins with simulating a value $\boldsymbol{X}_T^{(i)}$ from this distribution (where, as before, the superscript (i) indicates a realized value of a random variable). One then proceeds backward, realizing $\boldsymbol{X}_{T-1}^{(i)}$ from its conditional distribution given all of the observed variables, $\mathbf{Y}$, and $\boldsymbol{X}_T^{(i)}$. In similar fashion, one realizes $\boldsymbol{X}_{T-2}^{(i)}$ and so forth back to $\boldsymbol{X}_0^{(i)}$. In this backward phase, each $\boldsymbol{X}_t^{(i)}$ is simulated from its conditional distribution given all the data $\mathbf{Y}$ and all of the components of $\mathbf{X}$ that have been realized so far. That is, $\boldsymbol{X}_t^{(i)}$ is drawn from

$$P(\boldsymbol{X}_t|\boldsymbol{Y}_0, \ldots, \boldsymbol{Y}_T, \boldsymbol{X}_{t+1}^{(i)}, \ldots, \boldsymbol{X}_T^{(i)}). \tag{9}$$

Because of the conditional independence structure in a hidden Markov chain, (9) reduces to $P(\boldsymbol{X}_t|\boldsymbol{Y}_0, \ldots, \boldsymbol{Y}_t, \boldsymbol{X}_{t+1}^{(i)})$, which may be computed using the distributions stored during the forward step by the relation

$$P(\boldsymbol{X}_t|\boldsymbol{Y}_0, \ldots, \boldsymbol{Y}_t, \boldsymbol{X}_{t+1}^{(i)}) \propto P(\boldsymbol{X}_t|\boldsymbol{Y}_0, \ldots, \boldsymbol{Y}_t)\ P(\boldsymbol{X}_{t+1}^{(i)}|\boldsymbol{X}_t). \tag{10}$$

At the end of the backward step, it is thus clear that the resulting realization $(\boldsymbol{X}_0^{(i)}, \ldots, \boldsymbol{X}_T^{(i)})$ is from the conditional distribution of $\mathbf{X}$ given $\mathbf{Y}$.

An approximation for multiple alleles: In our application, with multiple alleles at a locus, since there are so many possible states that each $\boldsymbol{X}_t$ may take, the above procedure is computationally infeasible. However, we make use of the Baum *et al.* (1970) algorithm in spirit, employing two alterations to make it feasible to simulate from $P^*_{N_e}(\mathbf{X})$ and to compute its value. We emphasize that although the method described below involves a series of "approximations" by which $P^*_{N_e}(\mathbf{X})$, differs from $P_{N_e}(\mathbf{X}|\mathbf{Y})$, the final sampling and computation of $P^*_{N_e}(\mathbf{X})$ is exactly from the $P^*_{N_e}(\mathbf{X})$ as constructed, so its use in (7) gives a true Monte Carlo estimate.

The first approximation is to perform the forward-backward cycle separately for each allele. To describe this, we introduce some more notation. Denote by $\boldsymbol{X}_{(k)}$ the vector $(X_{0,k}, \ldots, X_{T,k})$ of latent counts of the kth allele from time $t = 0$ to $t = T$. Similarly we define $\boldsymbol{Y}_{(k)} = (Y_{0,k}, \ldots, Y_{T,k})$. To do the forward-backward cycle separately over alleles we first focus on allele 1, simulating $\boldsymbol{X}_{(1)}^{(i)}$ by the forward-backward mechanism as if the data were on a diallelic locus with observed counts $\boldsymbol{Y}_{(1)}$ from samples of size $S_0, \ldots, S_T$ through time. Once we have realized $\boldsymbol{X}_{(1)}^{(i)}$ we update the sizes of the population and the sample. Thus we define the updated population size vector $2\boldsymbol{N}_{(2)}^* = (2N_e - X_{0,1}^{(i)}, \ldots, 2N_e - X_{T,1}^{(i)})$ and an updated sample size vector $2\boldsymbol{S}_{(2)}^* = (2S_{0,2}^*, \ldots, 2S_{T,2}^*) = (2S_0 - Y_{0,1}, \ldots, 2S_T - Y_{T,1})$, in effect removing the first allelic type from the remainder of the data and the population. We then use the forward-backward mechanism again to simulate $\boldsymbol{X}_{(2)}^{(i)}$ as though the data were counts $\boldsymbol{Y}_{(2)}$ from a diallelic locus drawn from a population with sizes that change over time $\boldsymbol{N}_{(2)}^*$ and sample sizes $\boldsymbol{S}_{(2)}^*$. This continues sequentially over alleles, updating population sizes and sample sizes as above: $2\boldsymbol{N}_{(k)}^* \leftarrow (2\boldsymbol{N}_{(k-1)}^* - \boldsymbol{X}_{(k-1)}^{(i)})$ and $2\boldsymbol{S}_{(k)}^* \leftarrow (2\boldsymbol{S}_{(k-1)}^* - \boldsymbol{Y}_{(k-1)}^{(i)})$, until $\boldsymbol{X}_{(K-1)}$ has been realized, which also determines that $\boldsymbol{X}_{(K)} \leftarrow (2\boldsymbol{N}_{(K-1)}^* - \boldsymbol{X}_{(K-1)}^{(i)})$. (Here and later we use the notation $A \leftarrow B$ to mean "the value B is assigned to the variable A.") At the end one has obtained a realized value $\mathbf{X}^{(i)}$, which may be used in (7).

$P^*_{N_e}(\mathbf{X})$ using a continuous approximation: Although realizing alleles sequentially, as above, greatly reduces the number of terms required to use (8) and (10), the method would still involve a prohibitive amount of summation over binomial probabilities. Thus we construct $P^*_{N_e}(\mathbf{X})$, employing a normal approximation to binomial probabilities, which replaces all such sums by analytically tractable integrals. Recall that if $W \sim \text{Binomial}(n, p)$, then the transformed variable $\sin^{-1}(W/n)^{1/2}$ is approximately normally distributed with variance $1/(4n)$. Note that this quantity does not depend on p. Hence we use this transformation to define the quantities $\phi_{t,k} = \sin^{-1}[Y_{t,k}/(2S_{t,k}^*)]^{1/2}$ when $S_{t,k}^* > 0$, and $\theta_{t,k} = \sin^{-1}[X_{t,k}/2N_{t,k}^*)]^{1/2}$. By realizing the continuous values $\theta_{t,k}^{(i)}$ in a

forward-backward framework within a continuous setting, the computational demands are greatly reduced. And then, by transforming each $\theta_{t,k}^{(i)}$ back into the appropriate discrete $X_{t,k}^{(i)}$ we have a way to realize $\mathbf{X}^{(i)}$ from $P_{N_e}^{*}(\mathbf{X})$ and to compute the probability $P_{N_e}^{*}(\mathbf{X}^{(i)})$. The details of this procedure are given in the APPENDIX. We use it to compute the Monte Carlo estimate $\tilde{P}_{N_e}(\mathbf{Y})$ using (7).

Monte Carlo variance and multiple loci: The quantity $\tilde{P}_{N_e}(\mathbf{Y})$ is only an estimate of the true value $P_{N_e}(\mathbf{Y})$. By the central limit theorem, for large m, $P_{N_e}(\mathbf{Y})$ will be approximately normally distributed (HAMMERSLEY and HANDSCOMB 1964) with mean $\tilde{P}_{N_e}(\mathbf{Y})$ and a variance that may be approximated without bias by the quantity

$$\widehat{\mathrm{Var}}(\tilde{P}_{N_e}(\mathbf{Y})) = \frac{1}{m(m-1)}\sum_{i=1}^{m}\left(\frac{P_{N_e}(\mathbf{Y}, \mathbf{X}^{(i)})}{P_{N_e}^{*}(\mathbf{X}^{(i)})} - \tilde{P}_{N_e}(\mathbf{Y})\right)^2. \tag{11}$$

These facts may be used to obtain a confidence interval estimate around each $\tilde{P}_{N_e}(\mathbf{Y})$.

The ability to estimate N_e typically requires data from many loci. The extension to data on J independently segregating loci, indexed by $j = 1, \ldots, J$, is straightforward—each locus is treated separately, and the estimated likelihoods from each locus are multiplied together. Thus, let $\tilde{P}_{N_e,j}(\mathbf{Y})$ be the Monte Carlo likelihood estimate from the data on the jth locus. The Monte Carlo likelihood estimate using all the loci is then

$$\tilde{P}_{N_e}^{l}(\mathbf{Y}) = \prod_{j=1}^{J}\tilde{P}_{N_e,j}(\mathbf{Y}). \tag{12}$$

This requires that the initial allele counts have independent prior distributions, $\pi(\mathbf{X}_0)$. An implicit assumption of (12) is consequently that the loci used are in linkage equilibrium at $t = 0$. $\tilde{P}_{N_e}^{l}(\mathbf{Y})$ will also have an approximately normal distribution. An unbiased estimator for its Monte Carlo variance (derived in Equation A9 in the APPENDIX) is

$$\widehat{\mathrm{Var}}(\tilde{P}_{N_e}^{l}(\mathbf{Y})) = \prod_{j=1}^{J}\left(\tilde{P}_{N_e,j}(\mathbf{Y})\right)^2 - \prod_{j=1}^{J}\left([\tilde{P}_{N_e,j}(\mathbf{Y})]^2 - \widehat{\mathrm{Var}}(\tilde{P}_{N_e,j}(\mathbf{Y}))\right). \tag{13}$$

This can be used to compute a confidence interval estimate around $\tilde{P}_{N_e}^{l}(\mathbf{Y})$.

When displaying the Monte Carlo likelihood curve it is preferable to plot the log-likelihood values, $\log \tilde{P}_{N_e}^{l}(\mathbf{Y})$, for different values of N_e. In this case, the endpoints of the confidence intervals may be similarly log-transformed.

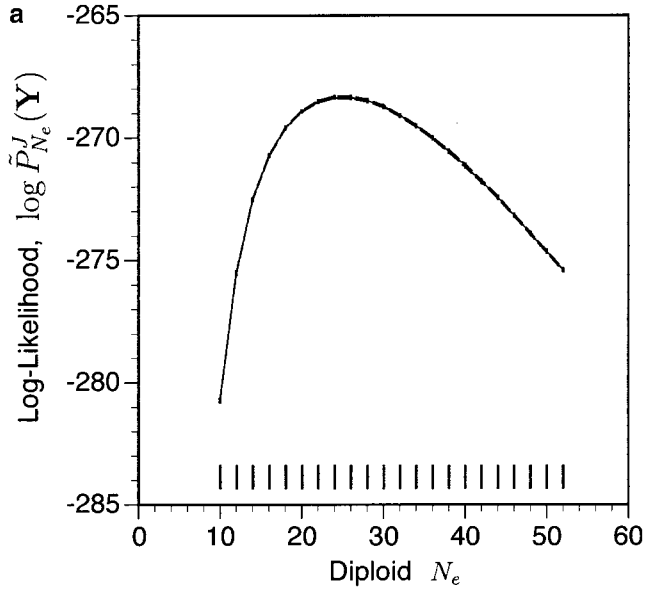
SIMULATED AND REAL DATASETS

We demonstrate the method by computing log-likelihood curves for N_e from three different datasets. First, to verify that the method gives correct results we apply it to a simple simulated dataset (dataset 1) for which it is possible to compute the likelihood exactly. We simulated samples of 20 diallelic loci from 100 diploid individuals at generations 0, 6, and 12 from a Wright-Fisher population of 25 diploid individuals. For each locus, the initial allele frequency in the population at time zero was an independently drawn uniform real number between 0 and 1. The log-likelihood for N_e given these data was estimated for values between 10 and 52, in steps of 2, using $m = 20{,}000$ realizations of $\mathbf{X}$ from $P_{N_e}^{*}(\mathbf{X})$ for each locus and each N_e.

We simulated a second dataset (dataset 2) to see how the method performed with multiallelic markers taken from a Wright-Fisher population. The dataset included three samples of 100 diploids for 12 five-allele loci at generations 0, 4, and 8 from a population of 50 diploids. The allele frequencies at each locus in generation 0 for these simulations were independently drawn from a uniform Dirichlet density with five components. For this dataset, the log-likelihood was computed for values of N_e between 20 and 100 in increments of 4 using $m = 50{,}000$ realizations of $\mathbf{X}$ for each locus and each value of N_e.

Finally, we computed a log-likelihood curve for N_e given data on a population of Drosophila in BEGON *et al.* (1980). These data were analyzed using *F*-statistics by BEGON *et al.* (1980) as well as by POLLAK (1983). They observed allele frequencies in three samples at each of nine enzyme loci. The first two samples were taken a little more than 1 yr apart, and the third sample was taken some 8 mo later. Though the natural populations do not have discrete generations, they have been modeled previously by Begon *et al.* and Pollak as populations with discrete generations. Because of the different growth rates of flies during different seasons, seven generations separate the first two samples, while only two generations separate the second two samples (BEGON *et al.* 1980). The sample sizes for all loci were the same, with larger sample sizes taken in the latter sampling periods. The sample sizes were $S_0 = 190$, $S_7 = 250$, and $S_9 = 335$ flies. POLLAK (1983) notes that since BEGON *et al.* (1980) sampled adult flies, their sampling scheme is closer to what is known in the literature as sampling scheme I than it is to sampling scheme II. However, as discussed by WAPLES (1989), the probability models underlying the two different sampling schemes are very similar when the actual size of the population is much larger than the effective size of the population. This is the case with these Drosophila. BEGON *et al.* (1980) report census sizes in the tens of thousands of flies, while the estimated N_e is orders of magnitude smaller. Because of this, it is still reasonable to analyze the data using the likelihood method we have developed here.

The data appear as allele frequencies in Table 1 of BEGON *et al.* (1980). Unfortunately the allele frequencies at the *Pgm* locus are misreported there and fail to sum



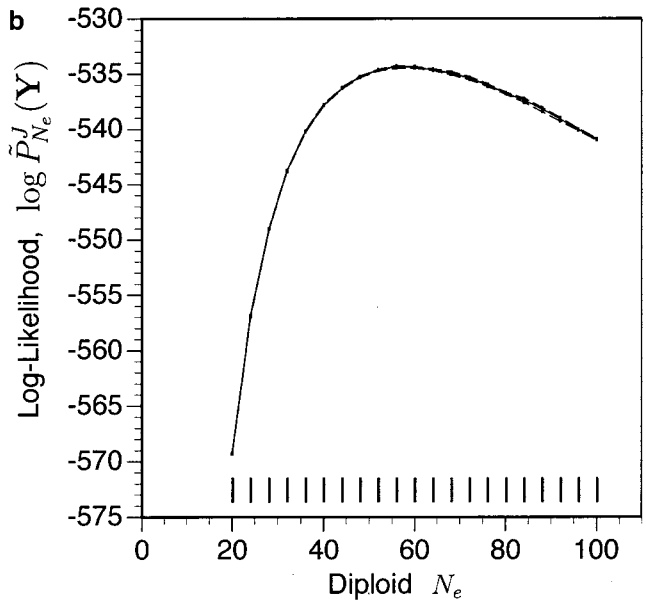


FIGURE 2.—Log-likelihood curves estimated by Monte Carlo from datasets 1 and 2. The values of N_e at which the likelihood was computed are indicated by vertical lines above the horizontal axis in each figure. The log-likelihood values are connected by a solid line. Vertical bars intersecting the solid line indicate 90% confidence intervals around log $\tilde{P}^I_{N_e}(\mathbf{Y})$ computed using the Monte Carlo variance estimate (13). The endpoints of the confidence intervals are connected by dashed lines. (a) Dataset 1 is simulated data from 20 diallelic loci. (b) Dataset 2 is simulated data from 12 loci with five alleles each.

to one. We thus used only the remaining eight loci. Of these eight, three had three alleles, two had four alleles, two had five alleles, and one had six alleles. We evaluated $\tilde{P}^I_{N_e}(\mathbf{Y})$ at values of N_e between 200 and 1200 in increments of 50, with two more points (N_e = 425 and N_e = 475) included near the peak of the likelihood curve. For each point we used m = 500,000 realizations of $\mathbf{X}$.

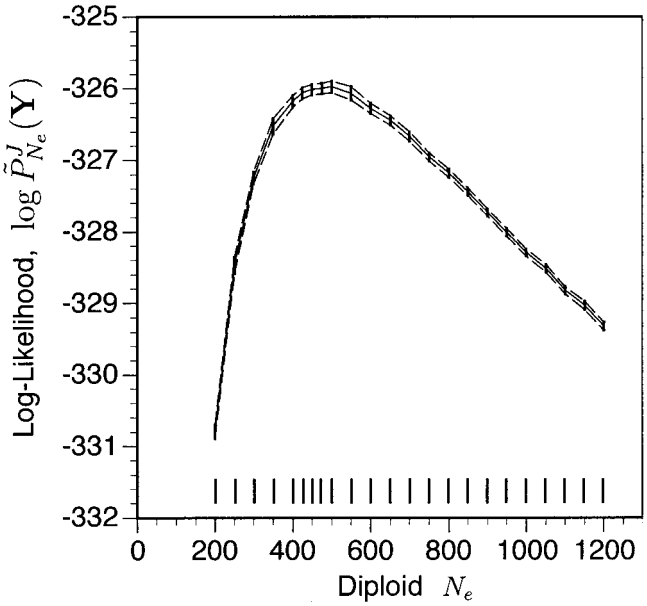


FIGURE 3.—Log-likelihood curves from the data of BEGON *et al.* (1980) estimated by Monte Carlo. The format of the plot is as for Figure 2.

RESULTS

For each of the datasets, we were able to use our importance-sampling method to compute a log-likelihood curve. Using a program we wrote in C, the runs for datasets 1 and 2 each took ~10 hr on a 266-Mhz laptop computer. The log-likelihood curves from datasets 1 and 2 appear as solid lines in Figure 2. The estimated 90% confidence intervals around each value of log $\tilde{P}^I_{N_e}(\mathbf{Y})$ appear as two dashed lines bordering the log-likelihood curve. Despite the fact that few Monte Carlo replicates (m = 20,000 and 50,000) were used, the Monte Carlo variance is minimal, as indicated by the fact that the dotted lines practically lie on top of the estimated log-likelihood curve. In both cases, the true values of N_e (25 and 50, respectively) are well within 2 units of log-likelihood from the maximum-likelihood estimates, which may be read from the graph as 24 and 56. Since dataset 1 consists only of diallelic loci, it is possible to compute the exact log-likelihood curve. This exact curve has been plotted as a dotted line in Figure 2a. It is impossible to distinguish the exact curve because the Monte Carlo estimate is very accurate in this case.

The log-likelihood curve computed for the data of BEGON *et al.* (1980) is shown in Figure 3. It took ~54 hr on a 450-Mhz desktop computer to produce the re-

sults. As before, the 90% confidence intervals around the Monte Carlo estimates appear as dotted lines. With this dataset, even with $m = 500{,}000$ realizations of $\mathbf{X}$, the Monte Carlo variance is not negligible. It is, however, small enough that reliable inferences may be made from the log-likelihood curve. The maximum-likelihood estimate of N_e is 500. Using the values of N_e at which the log-likelihood has decreased 2 units from its maximum gives an estimate of a 95% confidence interval for the true N_e. These points are 250 and 975. By contrast, Pollak (1983), using an *F*-statistic method, estimated N_e to be 251 with a standard error of 115. We discuss the discrepancy between the two estimates in the next section. Our results are not comparable to the N_e estimated by Begon *et al.* (1980) because, at the time, those authors were unable to make a single estimate of N_e using the samples at all three time points.

DISCUSSION

As discussed in Williamson and Slatkin (1999), estimating N_e by maximum likelihood has advantages over estimating N_e using *F*-statistics. Until now, it was impractical to compute the likelihood for N_e using all the data when loci with more than two alleles were available. While it has been suggested that one may bin low-frequency alleles together to turn multiallelic loci into apparently diallelic loci and then apply exact likelihood calculation methods to such reduced data, this invariably throws away some information. Furthermore, different binning strategies lead to different results. Allowing full use of the data, the Monte Carlo likelihood procedure described here is a preferable way to analyze temporal data on multiallelic loci. The method is suitable for multiallelic loci such as the microsatellite markers becoming available in a wide variety of species.

Monte Carlo methods use realizations of random variables to estimate an expectation by a sample average. There are a number of ways one can express the likelihood of N_e as an expectation, and then estimate it by Monte Carlo, but few of those schemes will be successful, because most will have high Monte Carlo variance. We attempted several different schemes before settling on the importance-sampling method presented here. Although these less sophisticated Monte Carlo estimators produced reasonable estimates in very small problems, when applied to data involving loci with many alleles these methods failed to converge reliably, even after many days of computation (E. C. Anderson and E. G. Williamson, unpublished data).

The importance-sampling method we use is successful because our importance sampling function, $P^*_{N_e}(\mathbf{X})$, closely resembles $P_{N_e}(\mathbf{X}|\mathbf{Y})$, the conditional probability of $\mathbf{X}$ given $\mathbf{Y}$. This is achieved by recognizing the hidden Markov chain structure of the problem and using the forward-backward algorithm of Baum *et al.* (1970). In doing so, we have developed a Monte Carlo estimator with demonstrably small Monte Carlo variance. Though the computational demands of this procedure are substantial, the reduction in Monte Carlo variance obtained makes it worthwhile. Nonetheless, it may be possible to improve the estimates by making additional changes to $P^*_{N_e}(\mathbf{X})$ so that it more closely resembles $P_{N_e}(\mathbf{X}, \mathbf{Y})$, especially in the tails of the distribution. This would further reduce the Monte Carlo variance.

It should be pointed out that while many Monte Carlo problems involving high-dimensional random variables like $\mathbf{X}$ make use of Markov chain Monte Carlo (MCMC) methods, our method is not an MCMC method. In MCMC, successive realizations are correlated. In our method each $\mathbf{X}^{(i)}$ realized from the distribution $P^*_{N_e}(\mathbf{X})$ is independent of the other realized values. As a result, our method does not have the same problems of convergence assessment as does Markov chain simulation (Gelman 1996).

It is interesting that our maximum-likelihood estimate differs so much from the estimate given by Pollak (1983) for the same data. Though perhaps some of this is attributable to the fact that we chose not to use the incorrectly reported data at the *Pgm* locus, there are differences between the two estimation methods that could also account for some of the discrepancy. The most notable differences occur when combining information from multiple samples in time. Consider the fact that a better estimate of N_e may be made from two samples taken many generations apart than from two samples separated by fewer generations. Likewise two large samples will yield a better estimate than two small samples. When there are many samples, the relative information content in different intersample intervals will depend on the relative sample sizes and the number of generations between the samples. By its nature, the maximum-likelihood approach will appropriately weight information from different intervals. In contrast, Pollak's *F*-statistic, F_{K_t}, neither includes terms for sample size nor interval length between samples, and, $\tilde{N}_{K_t}$, his estimate of N_e based on F_{K_t}, includes a term for only the number of generations between the first and the last sample and is invariant to permutations of the sample sizes at different times. Since the data from Begon *et al.* (1980) span sampling intervals of different lengths and include different sample sizes at different times, differences between our results and those in Pollak (1983) should be expected.

The Monte Carlo variance of our estimate of the likelihood given the data from a natural population of Drosophila was higher than the variance associated with our estimates from simulated data. Although a good estimate was achieved after sufficient computation, it may still be that data generated under a model that differs from the Wright-Fisher model present difficulty for the Monte Carlo likelihood method. For example, it may be that the effective size of the natural Drosophila population was different during the two different sam-

pling intervals. We note that one could extend the likelihood framework to allow for N_e changing over time. For example, if the estimated census size of the population were available and was known to change over time it would be more sensible to estimate directly the ratio, λ, of the effective size of the population to the census size of the population. This ratio would be more useful for the purposes of modeling genetic change in the population than a single estimate of N_e over the entire time period between the first and last samples. The forward-backward approach implemented here could easily be modified to estimate this parameter, λ.

It would also be possible to extend the present approach to compute likelihoods from different stochastic models of allele frequency change. We suspect there would seldom be enough information in the data to estimate accurately models in which allele frequency change is due jointly to genetic drift and some other genetic mechanism such as mutation or selection. However, our methods could be modified to handle different demographic models. For example, consider an alteration of the Wright-Fisher model such that the current generation is formed by sampling genes *without replacement* from a gamete pool in which each parental allele is represented by M gametes. In such a case, allele frequency changes occur due to multivariate hypergeometric sampling determined by the parameters N and M. For values of M that are not very small, the normal distribution is still a reasonable approximation to the hypergeometric distribution, and importance sampling could proceed as before. The importance-sampling distribution $P^*_{N,M}(\mathbf{X})$ should then account for the decrease in variance (by a factor of $2N[M-1]/[2NM-1]$) of the hypergeometric distribution relative to the binomial distribution. The same sort of sampling model and adjustment in the importance-sampling distribution could be used to accommodate scenarios in which one's genetic samples (the data $\mathbf{Y}$) were assumed drawn without replacement from the population.

Another possible extension would be to stochastic models involving populations of organisms with more complex life histories, for example, overlapping generations or age-structured populations. As it becomes easier to gather extensive genetic data on populations, and as understanding of the structure within those populations improves, it will be possible to specify much richer probability models for temporal population genetic data. The forward-backward method here, and variations of it, should be useful in estimating population parameters from such models using Monte Carlo likelihood.

A software package, MCLEEPS, implementing the algorithm described in this article is available for free download from http://www.stat.washington.edu/thompson/Genepi/Mcleeps.shtml.

We thank an anonymous referee for helpful comments on the manuscript and for suggesting the extension to the population model with hypergeometric sampling. This study was supported by National Science Foundation grant BIR-9807747 to E.T. Additionally, E.A. was supported by National Science Foundation grant BIR-9256537, the University of Washington QERM program, and a Burroughs Wellcome Fund PMMB training fellowship. E.W. was supported by National Institutes of Health grant GM40282 to M. Slatkin.

LITERATURE CITED

Baum, L. E., 1972 An inequality and associated maximization technique in statistical estimation for probabilistic functions of Markov processes, pp. 1–8 in *Inequalities—III: Proceedings of the Third Symposium on Inequalities Held at the University of California, Los Angeles, September 1–9, 1969*, edited by O. Shisha. Academic Press, New York.

Baum, L. E., T. Petrie, G. Soules and N. Weiss, 1970 A maximization technique occurring in the statistical analysis of probabilistic functions on Markov chains. Ann. Math. Stat. **41:** 164–171.

Begon, M., C. B. Krimbas and M. Loukas, 1980 The genetics of *Drosophila subobscura* populations. XV. Effective size of a natural population estimated by three independent methods. Heredity **45:** 335–350.

Cavalli-Sforza, L. L., and A. W. F. Edwards, 1967 Phylogenetic analysis: models and estimation procedures. Evolution **21:** 550–570.

Frankham, R., 1995 Inbreeding and extinction: a threshold effect. Conserv. Biol. **9:** 792–799.

Gelman, A., 1996 Inference and monitoring convergence. pp. 131–143 in *Markov Chain Monte Carlo in Practice*, edited by W. R. Gilks, S. Richardson and D. J. Spiegelhalter. Chapman & Hall, New York.

Gelman, A., J. B. Carlin, H. S. Stern and D. B. Rubin, 1996 *Bayesian Data Analysis.* Chapman and Hall, New York.

Hammersley, J. M., and D. C. Handscomb, 1964 *Monte Carlo Methods.* Methuen & Co. Ltd., London.

Jorde, P. E., and N. Ryman, 1995 Temporal allele frequency change and estimation of effective size in populations with overlapping generations. Genetics **139:** 1077–1090.

Krimbas, C. B., and S. Tsakas, 1971 The genetics of *Dacus oleae.* V. Changes of esterase polymorphism in a natural population following insecticide control—selection or drift? Evolution **25:** 454–460.

Nei, M., and F. Tajima, 1981 Genetic drift and estimation of effective population size. Genetics **98:** 625–640.

Neyman, J., and E. L. Scott, 1948 Consistent estimates based on partially consistent observations. Econometrica **16:** 1–32.

Pollak, E., 1983 A new method for estimating the effective population size from allele frequency changes. Genetics **104:** 531–548.

Soulé, M. E. (Editor), 1986 *Conservation Biology: The Science of Scarcity and Diversity.* Sinauer and Associates, Sunderland, MA.

Waples, R. S., 1989 A generalized approach for estimating effective population size from temporal changes in allele frequency. Genetics **121:** 379–391.

Williamson, E. G., and M. Slatkin, 1999 Using maximum likelihood to estimate population size from temporal changes in allele frequencies. Genetics **152:** 755–761.

Communicating editor: G. A. Churchill

APPENDIX

Using $\theta_{t,k}$ and $\phi_{t,k}$ in a continuous setting: We define the random variables $\phi_{t,k} = \sin^{-1}[Y_{t,k}/(2S^*_{t,k})]^{1/2}$ when $S^*_{t,k} > 0$, and $\theta_{t,k} = \sin^{-1}[X_{t,k}/(2N^*_{t,k})^{1/2}$ when $N^*_{t,k} > 0$. These quantities have an approximate normal distribution, which is independent of their means. We use them in our construction of the importance-sampling function $P_{N_e}(\mathbf{X})$. Below, we concentrate on their use for realizing $\boldsymbol{X}^{(i)}_{(k)}$, keeping in mind that if $k > 1$ then we will have already realized $\boldsymbol{X}^{(i)}_{(k-1)}$, and we use the updated population and sample sizes $\boldsymbol{N}^*_{(k)}$ and $\boldsymbol{S}^*_{(k)}$. If $k = 1$ then

$\boldsymbol{N}^*_{(1)} = (N_e, \ldots, N_e)$ and $\boldsymbol{S}^*_{(1)} = (S_0, \ldots, S_T)$, respectively.

The forward step: Following Cavalli-Sforza and Edwards (1967), if $\theta_{t-1,k}$ is normally ($\mathcal{N}$) distributed with mean μ_{t-1} and variance σ^2_{t-1}, then, after a generation of genetic drift in a population of $N^*_{t,k}$ diploids, $\theta_{t,k}$ has an approximate normal distributions with mean μ_{t-1} and variance $\sigma^2_t = \sigma^2_{t-1} + 1/(8N^*_{t,k})$. If there are data $\boldsymbol{Y}_{t,k}$ from a sample of size $S_{t,k}$ at time t, then $\phi_{t,k}$ has an approximate normal distribution with mean $\theta_{t,k}$ and variance $1/(8S^*_{t,k})$, so, given that $\theta_{t,k} \sim \mathcal{N}(\mu_t, \sigma^2_t)$, the conditional distribution of $\theta_{t,k}$ given $\phi_{t,k}$ is also normal. These relations form the basis of a continuous approximation for doing the forward step. For the purpose of realizing $\mathbf{X}$ we assume that the uniform prior on $\boldsymbol{X}_0$ is equivalent to a diffuse prior on $\theta_{0,k}$. Therefore $\theta_{0,k} \,|\phi_{0,k} \sim \mathcal{N}(\mu_0, \sigma^2_0)$ with $\mu_0 = \phi_{0,k}$ and $\sigma^2_0 = 1/(8S^*_{0,k})$. With that as a starting point, we work iteratively forward in time, assigning values

$$\mu_t \leftarrow \mu_{t-1} \tag{A1}$$

$$\sigma^2_t \leftarrow \sigma^2_{t-1} + 1/(8N^*_{t,k}) \tag{A2}$$

if $S^*_{t,k} = 0$. If $S^*_{t,k} > 0$, however, then one first computes μ_t and σ^2 as in (A1) and (A2), but then further updates the values to reflect the information in the sample at time t:

$$\mu_t \leftarrow \frac{\mu_t/(8S^*_{t,k}) + \sigma^2_t\phi_{t,k}}{1/(8S^*_{t,k}) + \sigma^2_t} \tag{A3}$$

$$\sigma^2_t \leftarrow \frac{\sigma^2_t/(8S^*_{t,k})}{1/(8S^*_{t,k}) + \sigma^2_t}. \tag{A4}$$

This is analogous to computing a posterior distribution from a normal prior and normal data (see, for example, Gelman *et al.* 1996, p. 43).

Carrying this out until $t = T$ gives values for the mean and variance of $\theta_{T,k}$ given $\phi_{0,k}, \ldots, \phi_{T,k}$, assuming they follow a normal distribution. In fact, for each t, it gives us the parameters for the normal distribution of $\theta_{t,k}$ conditional on $\phi_{r,k}$ for all $r \le t$. We are thus in a position to realize $\theta^{(i)}_{t,k}$'s in the backward step and transform those $\theta^{(i)}_{t,k}$'s back into the $X^{(i)}_{t,k}$'s that we need.

The backward step: The backward step is more complicated than the forward step, because after realizing each value of $\theta^{(i)}_{t,k}$ we must transform it into the discrete value $X^{(i)}_{t,k}$ that we require. This transformation process requires some extra bookkeeping to ensure that we do not waste time realizing $\mathbf{X}^{(i)}$'s that are incompatible with the data. This is described in the next section of the appendix. We first realize the value $\theta^{(i)}_{T,k}$ from a $\mathcal{N}(\mu_T, \sigma^2_T)$ distribution. Then we transform that to the realization $X^{(i)}_{T,k}$ by a many-to-one map $\mathcal{M}$, which has two effects: the first is that of folding and translating the distribution of $\theta_{T,k}$ so that it is bounded between 0 and $\pi/2$, mapping $\theta^{(i)}_{T,k} \in (-\infty, \infty)$ to a value $\theta^*_{T,k} \in [0, \pi/2]$. The second is transforming that $\theta^*_{T,k}$ into the appropriate value $X^{(i)}_{T,k}$ (see the next section).

Working backward, each $\theta^{(i)}_{T,k}$ for $t = T - 1$ down to $t = 0$, is realized from a $\mathcal{N}(\mu_t, \sigma^2_t)$ distribution and then transformed into the corresponding $\theta^*_{T,k}$ and $X^{(i)}_{t,k}$ by $\mathcal{M}$. In keeping with (10), before $\theta^{(i)}_{t,k}$ is realized, μ_t and σ^2_t must be appropriately updated, on the basis of the values of μ_t and σ^2_t stored during the forward step and the realized value $\theta^{(i)}_{t+1,k}$. This involves making the assignments

$$\mu_t \leftarrow \frac{\mu_t/(8N^*_{t+1}) + \sigma^2_t\theta^*_{t+1,k}}{1/(8N^*_{t+1}) + \sigma^2_t} \tag{A5}$$

$$\sigma^2_t \leftarrow \frac{\sigma^2_t/(8N^*_{t+1})}{1/(8N^*_{t+1}) + \sigma^2_t} \tag{A6}$$

in the order as written.

Computing the probability $P^*_{N_e}(\boldsymbol{X}^{(i)})$*:* By carrying out the forward-backward steps above on the first allele, the realization $\boldsymbol{X}^{(i)}_{(1)}$ is obtained. Then, $\boldsymbol{N}^*_{(2)}$ and $\boldsymbol{S}^*_{(2)}$ are computed and used in the forward-backward steps to obtain $\boldsymbol{X}^{(i)}_{(2)}$. Executing these steps for all the alleles yields the realization $\mathbf{X}^{(i)}$, which is used in (7). $P_{N_e}(\mathbf{Y}, \mathbf{X}^{(i)})$ in (7) is easily computed using the expansion shown between the large parentheses in (3).

It remains only to compute $P^*_{N_e}(\mathbf{X}^{(i)})$, which can be done by recording the probability of realizing each component $\boldsymbol{X}^{(i)}_{t,k}$. Although this probability depends on the values of μ_t, σ^2_t, $N^*_{(k)}$, and several bookkeeping variables, we denote it here simply by $\mathcal{P}(X^{(i)}_{t,k})$. (The actual function $\mathcal{P}$ is described later in this appendix.) As long as the realization of $\boldsymbol{X}^{(i)}_{(k)}$ over alleles occurs in the same order over k ($k = 1, 2, \ldots, K$) for each i, then

$$P^*_{N_e}(\mathbf{X}^{(i)}) = \prod_{k=1}^{K}\prod_{t=1}^{T}\mathcal{P}(X^{(i)}_{t,k}). \tag{A7}$$

Details of $\mathcal{M}$: The fact that we are realizing $\boldsymbol{X}^{(i)}_{(k)}$'s one allele at a time requires that we do some extra bookkeeping to keep our importance-sampling scheme efficient. Primarily, we must avoid realizing $\boldsymbol{X}^{(i)}$'s for which $P_{N_e}(\mathbf{Y}, \mathbf{X}^{(i)}) = 0$. Potential problems arise because by the method we use to realize values from $P^*_{N_e}(\mathbf{X})$, $X_{t,k}$ may only take values between 0 and $2N^*_{t,k}$, inclusive. If $2N^*_{t,k} = 0$ at any value of t, then for any $s > t$, $X^{(i)}_{s,k}$ must also be 0. To avoid situations in which this leads to $P_{N_e}(\mathbf{Y}, \mathbf{X}^{(i)})$ being 0 (as when $X^{(i)}_{t,k} = 0$ and $Y_{t,k} > 0$) we introduce the following scheme and additional notation:

$$\delta_{t,k} = \begin{cases} 1 & \text{if } X^{(i)}_{t,k} = 0 \text{ implies } P_{N_e}(\mathbf{Y}, \mathbf{X}^{(i)}) = 0 \\ 0 & \text{otherwise} \end{cases}$$

$$\gamma_{t,k} = \min_{r<t} 2N^*_{r,k}$$

$$\kappa_{t,k} = \text{the number of allelic subscripts } \ell\text{: } k < \ell \le K \text{ such that } Y_{r,\ell} > 0 \text{ for at least one } r \ge t. \tag{A8}$$

Knowing the above quantities, we can define the func-

tion $\mathcal{M}$. In the remainder of this section and in the following one we drop the t and k subscripts for clarity.

With N^* and γ positive integers, $\delta \in \{0, 1\}$, and $\kappa \in \{0, 1, \ldots, \min(2N^* - \delta, \gamma - \delta)\}$, let $\mathcal{M}(\theta; N^*, \delta, \gamma, \kappa)$: $\mathbb{R}^1 \to \{\delta, \ldots, 2N^* - \kappa\} \times [0, \pi/2]$ be the many-to-one map that takes a realization of $\theta \in (-\infty, \infty)$ to the ordered pair (X, θ^*), where X is an integer such that $\delta \le X \le 2N^* - \kappa$, and θ^* is a real number between 0 and $\pi/2$, inclusive. $\mathcal{M}$ may be described by the following pseudocode. We first define the quantities $L = \sin^{-1}(0.5/(2N^*))^{1/2}$ and

$$H = \begin{cases} \sin^{-1}[(2N^* - \kappa + 0.5)/(2N^*)]^{1/2}, & \kappa \ge 1 \\ \sin^{-1}[(2N^* - 0.5)/(2N^*)]^{1/2}, & \kappa = 0. \end{cases}$$

Then,

if ($\delta = 2N^* - \kappa$ or $\delta = \gamma - \kappa = 0$) then $\theta^* \leftarrow 0$

else if ($L < \theta < H$) then $\theta^* \leftarrow \theta$

else if ($\theta < L$)

and if ($\delta = 0$) then $\theta^* \leftarrow \theta$

else if ($\delta = 1$) then $\theta_{[L]} \leftarrow 2L - \theta$ (*this is reflection around* $\theta = L$), and then

if ($L \le \theta_{[L]} < H$) then $\theta^* \leftarrow \theta_{[L]}$

else we know $\theta_{[L]} \ge H$, and we consider the sequence $\theta_{[i]} = i(L - H) + \theta_{[L]}$, $i = 1, 2, \ldots$, and we assign $\theta^* \leftarrow \theta_{[i^*]}$, where i^* is the least i such that $L < \theta_{[i]} < H$. (*The sequence* $\theta_{[i]}$ *represents successive translation leftward.*)

else if ($\theta > H$)

and if ($\kappa = 0$) then $\theta^* \leftarrow \pi/2$

else if ($\kappa > 1$) then $\theta_{[H]} \leftarrow 2H - \theta$ (*this is reflection around* $\theta = H$), and then

if ($L < \theta_{[H]} < H$) then $\theta^* \leftarrow \theta_{[H]}$

else we know $\theta_{[H]} < L$ and we consider the sequence $\theta_{[j]} = j(H - L) + \theta_{[H]}$, $j = 1, 2, \ldots$, and we

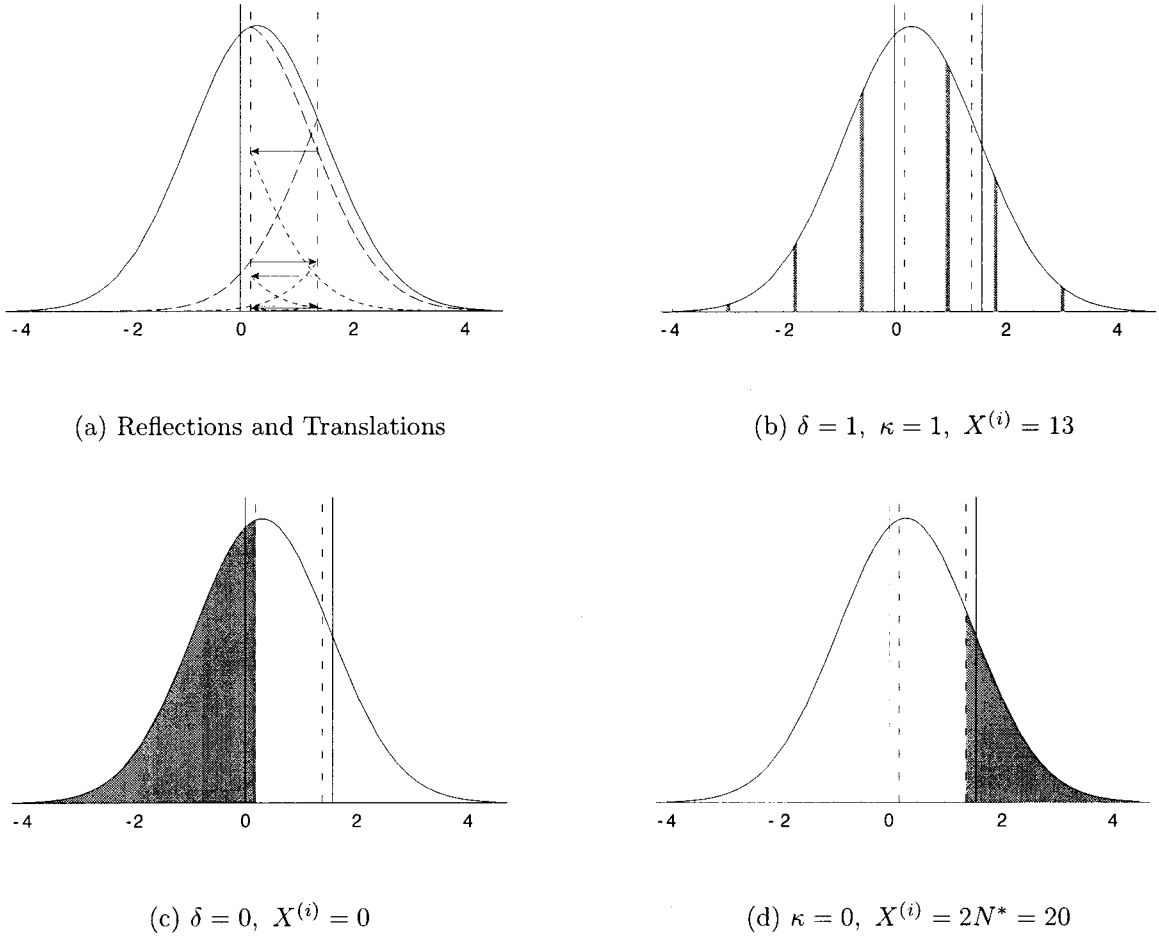


FIGURE A1.—Figures representing $\mathcal{M}$ and $\mathcal{P}$ for $2N^* = 20$. The normal curve is the density for θ. (a) Reflections and translations as described in the APPENDIX. Long-dashed lines represent the curve after reflection through L or H, while the short-dashed lines represent the reflected curve after one or more successive translations. (b) If $\delta = 1$ and $\kappa = 1$, then $X^{(i)}$ is constrained to be in $\{1 \ldots, 2N^* - 1\}$. The shaded regions correspond to those values of θ for which $X^{(i)} = 13$ by $\mathcal{M}$. The total shaded area is equal to $\mathcal{P}_{\mu,\sigma^2}(X = 13; 10, 1, \gamma, 1)$. (c) If $\delta = 0$ then $X^{(i)}$ may take the value 0. The shaded area shows $\mathcal{P}_{\mu,\sigma^2}(X = 0; 10, 0, \gamma, \kappa)$. (d) If $\kappa = 0$ then X may take the value $2N^*$. The shaded area shows $\mathcal{P}_{\mu,\sigma^2}(X = 2N^*; 10, \delta, \gamma, 0)$.

assign $\theta^* \leftarrow \theta_{[j^*]}$, where j^* is the least j such that $L \le \theta_{[j]} < H$. (*The sequence* $\theta_{[j]}$ *represents successive translation rightward.*)

finally we use θ^*, making the assignment $X \leftarrow \lfloor 2N^*\sin^2\theta^* + 0.5 \rfloor$,

where $\lfloor x \rfloor$ denotes the largest integer $\le x$. The reflections and translations are depicted graphically in Figure A1a.

The probability $\mathcal{P}_{\mu,\sigma^2}(X = X^{(i)}; N^*, \delta, \gamma, \kappa)$ of realizing $X = X^{(i)}$: If $\theta \sim \mathcal{N}(\mu, \sigma^2)$, and $(X, \theta^*) = \mathcal{M}(\theta; N^*, \delta, \gamma, \kappa)$, then we denote by $\mathcal{P}_{\mu,\sigma^2}(X = X^{(i)}; N^*, \delta, \gamma, \kappa)$ the marginal probability that $X = X^{(i)}$. The value of $\mathcal{P}_{\mu,\sigma^2}(X = X^{(i)}; N^*, \delta, \gamma, \kappa)$ can be expressed using the notation from the above section. First, $\mathcal{P}_{\mu,\sigma^2}(X = X^{(i)}; N^*, \delta, \gamma, \kappa) = 0$ if $X^{(i)} < \delta$ or $X^{(i)} > 2N^* - \kappa$, though such values of $X^{(i)}$ should never occur from $\mathcal{M}$ anyway. Second, there are cases when $\mathcal{M}$ constrains $X^{(i)}$ to be either 0 or 1 with probability 1. Hence if $2N^* - \kappa = \delta$ or $\gamma - \kappa = \delta = 0$ then $\mathcal{P}_{\mu,\sigma^2}(X = \delta; N^*, \delta, \gamma, \kappa) = 1$.

If, on the other hand, $2N^* - \kappa > \delta$ and $\gamma - \kappa > 0$, then for $X^{(i)} = 0$ and $X^{(i)} = 2N^*$ we have

$$\mathcal{P}_{\mu,\sigma^2}(X = 0; N^*, 0, \gamma, \kappa) = P(-\infty < \theta < L)$$

$$\mathcal{P}_{\mu,\sigma^2}(X = 2N^*; N^*, \delta, \gamma, 0) = P(H \le \theta < \infty),$$

while for $0 < X^{(i)} < 2N^* - \kappa$ we define $a = \sin^{-1}[(X^{(i)} - 0.5)/2N^*)]^{1/2}$ and $b = \sin^{-1}[(X^{(i)} + 0.5)/(2N^*)]^{1/2}$ and have

$$\begin{aligned}\mathcal{P}_{\mu,\sigma^2}(X = X^{(i)}; N^*, \delta, \gamma, \kappa) = {} & P(a \le \theta < b) \\ & + I\{\delta = 1\}P(a \le \theta_{[L]} < b) \\ & + I\{\kappa > 0\}P(a \le \theta_{[H]} < b) \\ & + I\{\delta = 1\}\sum_{i=1}^{\infty} P(a \le \theta_{[i]} < b) \\ & + I\{\kappa > 0\}\sum_{j=1}^{\infty} P(a \le \theta_{[j]} < b),\end{aligned}$$

where $I\{\cdot\}$ is the indicator function and $P(a \le \theta < b)$ is the probability that a $\mathcal{N}(\mu, \sigma^2)$ random variable is between a and b, namely $\int_a^b (2\pi\sigma^2)^{-1/2} \exp\{[-(\theta - \mu)^2]/(2\sigma^2)\}d\theta$. We compute this probability by numerical integration in our programs. In practice, the infinite sums are approximated by summing the first several terms of the series, until the contribution of the next term is very small (*e.g.*, $<10^{-7}$). Values of $\mathcal{P}$ for different values of δ and κ appear as shaded regions in Figure A1, b–d.

This folding and translating might seem to be a very involved process, but it is computationally much faster than realizing θ from a truncated normal distribution and computing the probability of $X^{(i)}$ when θ is from such a distribution.

Multilocus Monte Carlo variance calculation: We derive an expression for the variance of a product of J independent random variables, W_j, $j = 1, \ldots J$:

$$\begin{aligned}\mathrm{Var}(\textstyle\prod W_j) &= \mathbb{E}\,([\prod W_j]^2) - [\mathbb{E}(\prod W_j)]^2 && \text{(definition of variance)} \\ &= \mathbb{E}([\textstyle\prod W_j^2]) - [\mathbb{E}(\prod W_j)]^2 && \text{(powers distribute over products)} \\ &= \textstyle\prod \mathbb{E}(W_j^2) - \prod[\mathbb{E}(W_j)]^2 && \text{(independence of the } W_j\text{)} \\ &= \textstyle\prod \mathbb{E}(W_j^2) - \prod[\mathbb{E}(W_j^2) - \mathrm{Var}(W_j)] && \text{(definition of variance).}\end{aligned}$$

Denoting $\tilde{P}_{N_e,J}(\mathbf{Y})$ in (13) by W_j and taking the expectation gives the same result, verifying that the expression in (13) is unbiased for $\mathrm{Var}(\tilde{P}^J_{N_e}(\mathbf{Y}))$.