A pseudo-likelihood method for estimating effective population size from temporally spaced samples

JINLIANG WANG*

Institute of Zoology, Zoological Society of London, Regent's Park, London NW1 4RY, UK

(*Received 30 January 2001 and in revised form 8 May 2001*)

Summary

A pseudo maximum likelihood method is proposed to estimate effective population size (N_e) using temporal changes in allele frequencies at multi-allelic loci. The computation is simplified dramatically by (1) approximating the multi-dimensional joint probabilities of all the data by the product of marginal probabilities (hence the name pseudo-likelihood), (2) exploiting the special properties of transition matrix and (3) using a hidden Markov chain algorithm. Simulations show that the pseudo-likelihood method has a similar performance but needs much less computing time and storage compared with the full likelihood method in the case of 3 alleles per locus. Due to computational developments, I was able to assess the performance of the pseudo-likelihood method against the F-statistic method over a wide range of parameters by extensive simulations. It is shown that the pseudo-likelihood method gives more accurate and precise estimates of N_e than the F-statistic method, and the performance difference is mainly due to the presence of rare alleles in the samples. The pseudo-likelihood method is also flexible and can use three or more temporal samples simultaneously to estimate satisfactorily the N_es of each period, or the growth parameters of the population. The accuracy and precision of both methods depend on the ratio of the product of sample size and the number of generations involved to N_e, and the number of independent alleles used. In an application of the pseudo-likelihood method to a large data set of an olive fly population, more precise estimates of N_e are obtained than those from the F-statistic method.

1. Introduction

Effective population size (N_e) determines the genetic stochasticity of a population (Wright, 1931). Many genetic processes and their consequences in a finite population (e.g. inbreeding and inbreeding depression, random drift and loss of variation, accumulation of deleterious mutations and decline of fitness) are associated closely with this important parameter (Frankham, 1995). Until recently, however, we have known very little about the effective sizes of natural populations. Several approaches have been developed and used to estimate the current or short-term N_e of a population (Roff, 1997; Schwartz *et al.*, 1999). The ecological approach is based on predictive equations (reviewed by Caballero, 1994; Wang & Caballero, 1999) that require certain demographic parameters,

such as variance of family size, which are difficult to obtain from natural populations. The lethal allelism approach, first described by Dobzhansky & Wright (1941), is applicable only to species (e.g. *Drosophila*) with a system of balanced lethal chromosomes required for surveying lethal genes (Nei & Tajima, 1981). The third approach, within which several methods are subsumed, uses information from genetic markers.

The linkage disequilibrium method requires prior knowledge of linkage relationships among loci (Hill, 1981) which are largely unknown in most species. The heterozygosity excess method has a very low precision and is quite sensitive to nonrandom mating (Pudovkin *et al.*, 1996; Luikart & Cornuet, 1999). The temporal method is based on the empirical observations of temporal changes in marker allele frequency. The relationship between N_e and the standardized variance of change in gene frequency (F) led to the development of this temporal method (Krimbas & Tsakas, 1971;

*Tel: +44 (0)20 74496620. Fax: +44 (0)20 75862870. e-mail: jinliang.wang@ioz.ac.uk

Nei & Tajima, 1981; Pollak, 1983; Waples, 1989), which has been extended (e.g. Jorde & Ryman, 1995) and applied to data from a variety of species and populations (Funk *et al.*, 1999; Labate *et al.*, 1999; Fiumera *et al.*, 1999, 2000; Turner *et al.*, 1999; Kantanen *et al.*, 1999; see also many earlier references cited in Williamson & Slatkin, 1999).

The estimation of N_e from F utilizes only the first two moments of allele frequency distribution; higher-order moments are ignored. The information on the full distribution of allele frequency can be exploited by using the maximum likelihood method, developed by Williamson & Slatkin (1999, referred to as W&S99 hereafter). They showed that the likelihood estimator has a smaller bias and higher precision than the *F*-statistic estimator but is much more computationally intensive. As a consequence, they considered only very small populations with a quite limited subset of parameters. More work is necessary to assess the performance of the likelihood method against the *F*-statistic method over a wide range of parameters. Moreover, their method applies to bi-allelic loci only, while most markers in practical use nowadays are highly polymorphic. Anderson *et al.* (2000) recently extended the likelihood approach to include multi-allelic loci. The extension uses Monte Carlo technique to compute the likelihood and is therefore highly computation demanding. More efficient methods for using highly polymorphic markers in estimating N_e are obviously desirable.

In this study, I propose a pseudo-likelihood method for estimating N_e with data on multi-allelic loci. I also use several algorithms to reduce the computing time and memory requirements of the likelihood method. Simulations were run to justify these developments, and to compare the performances of the pseudo-likelihood and *F*-statistic methods under a wide range of parameters. The pseudo-likelihood method is also applied to the data from an olive fly population which have been analysed several times by the *F*-statistic method.

2. Method

(i) *The basic model*

Following previous analyses, I assume a diploid population with discrete generations. Selection, migration and mutation are assumed to be unimportant in changing allele frequencies compared with genetic drift. Samples of sizes $1/2n_0$ and $1/2n_t$ individuals for genetic analyses are drawn at random from the population at generations 0 and t, respectively, yielding allelic counts n_{01}, n_{02} ($n_{01}+n_{02}=n_0$) and n_{t1}, n_{t2} ($n_{t1}+n_{t2}=n_t$) for a bi-allelic locus. Each sampling does not affect the gene frequency of the population (sampling plan I in Nei & Tajima, 1981). This is achieved by sampling after reproduction or sampling with replacement before reproduction. If the actual population size is much larger than the sample size, then sampling without replacement before reproduction can also be treated similarly (Waples, 1989). The loci for genotyping are assumed to be neutral and unlinked.

The likelihood of the (harmonic) mean effective size (N_e) of the population during the sampling period (generations 0–t), given the data, is (W&S99)

$$P(n_{01}, n_{02}; n_{t1}, n_{t2} \mid t, N_e)$$
$$= \sum_{q_0 q_t} P(n_{01}, n_{02} \mid q_0) P(q_0 \mid N_e)$$
$$\times P(n_{t1}, n_{t2} \mid q_t) P(q_t \mid q_0, t, N_e), \quad (1)$$

where P denotes the probability of an event and q_i is the frequency of allele 1 in the population at time i ($i = 0, t$). The allelic counts n_{i1} and n_{i2}, with allelic sample size $n_i = n_{i1}+n_{i2}$, are binomially distributed:

$$P(n_{i1}, n_{i2} \mid q_i) = \frac{n_i!}{n_{i1}! n_{i2}!} q_i^{n_{i1}} (1-q_i)^{n_{i2}}. \quad (2)$$

The probability of an initial allele frequency, $P(q_0 \mid N_e)$, is generally unknown. Following W&S99, I assume q_0 is uniformly distributed. Because only segregating loci are used in the estimation, $P(q_0 \mid N_e) = 1/(C-2)$ for any possible value of q_0, where $C = 2N_e + 1$ is the number of possible configurations of the effective numbers of the two alleles in the population.

The probability of q_t given q_0 and N_e, $P(q_t \mid q_0, N_e)$, can be computed by using the transition matrix **M** (Ewens, 1979). The transition of a population's allelic configurations from generations $t-1$ to t can be described by **M**, with element m_{ij} ($i, j = 0, \ldots, 2N_e$) being the probability of the *j*th configuration of the parental population being converted to the *i*th configuration of the offspring population. m_{ij} can be calculated by (2), replacing the allelic counts by the *i*th configuration (i.e. effective numbers of alleles 1 and 2 being i and $2N_e - i$, respectively) and allelic frequency q by the *j*th configuration divided by $2N_e$. Given the initial distribution of allele frequency $\mathbf{P}(q_0)$ (a column vector with C elements), the distribution after t generations is determined completely by N_e and t, calculated by the recurrent equation

$$\mathbf{P}(q_t \mid q_{t-1}, N_e) = \mathbf{M} \mathbf{P}(q_{t-1} \mid q_{t-2}, N_e). \quad (3)$$

(ii) *Extensions to the basic model*

(a) *The pseudo-likelihood method*

For marker loci with more than two alleles, (2) can be replaced by a multinomial probability function. The number of possible configurations, $C = (2N_e + k - 1)!/(2N_e!(k-1)!)$ for a diploid locus with k

alleles (Feller, 1950; W&S99), increases rapidly with k, however. It is obvious that the evaluation of the exact likelihood for multi-allelic loci requires prohibitively large amounts of computation and storage. For a moderate value of N_e, even $k = 3$ poses a serious computational problem for the likelihood method. Anderson *et al.* (2000) proposed a Monte Carlo method to evaluate the likelihood function with data on multi-allelic markers. Their method does not need much storage but is highly computationally demanding and applies only to very small problems. To simplify the likelihood computation for multi-allelic loci, a k-allele locus can be transformed into k bi-allelic 'loci', each having one of the k alleles with all the other alleles pooled. The overall log-likelihood is approximated by the sum of the log-likelihoods across 'loci' multiplied by the factor of $(k-1)/k$. This method is called pseudo-likelihood herein because it actually approximates the joint probability of all the data by the product of marginal probabilities. Although such reconstructed 'loci' from a multi-allelic locus are not completely independent and therefore the total likelihood is not exactly the multiplication across these 'loci' in theory, this is a good approximation which works well as verified by extensive simulations. This treatment is actually similar to that of the *F*-statistic method (Nei & Tajima, 1981).

(b) *Genetic markers*

For bi-allelic dominant markers such as restriction fragment length polymorphisms (RFLPs), (2) can be replaced by

$$P(n_{iR}, n_{iD} \,|\, q_i) = \frac{n_i!}{n_{iR}! n_{iD}!} (q_i^2)^{n_{iR}} (1-q_i^2)^{n_{iD}}, \qquad (4)$$

if the markers are in Hardy–Weinberg equilibrium (W&S99). In (4), n_{iR} and n_{iD} are the numbers of individuals with recessive and dominant phenotypes, respectively, in the ith sample. The individual sample size is $n_i = n_{iR} + n_{iD}$, and the recessive allele frequency is q_i.

For haploid markers such as mtDNA polymorphisms, $C = N_e + 1$ and the transition matrix reduces to dimensions $(N_e + 1) \times (N_e + 1)$. The computation is simplified compared with the diploid case.

Different types of marker loci (such as dominant and co-dominant) can be combined to give an overall estimation of N_e. Given that the loci are independent, the total likelihood is just the multiplication of the likelihoods calculated separately for each locus. Care should be taken when considering diploid and haploid loci (such as microsatellite and mtDNA polymorphisms) jointly, because they refer to different effective sizes. Unless the relationship between haploid

and diploid effective sizes is known, these two types of markers cannot be combined in the likelihood estimation.

(c) *More than two samples*

For three samples obtained sequentially at generations 0, t_1 and t_2, joint estimates of N_{e1} and N_{e2} for periods 0–t_1 and t_1–t_2, respectively, can be obtained by extending the likelihood function (1). The function is now

$$P(n_{01}, n_{02}; n_{t_11}, n_{t_12}; n_{t_21}, n_{t_22} \,|\, t_1, t_2, N_{e1}, N_{e2})$$
$$= \sum_{q_0 q_{t_1} q_{t_2}} P(n_{01}, n_{02} \,|\, q_0) P(q_0 \,|\, N_{e1}) P(n_{t_11}, n_{t_12} \,|\, q_{t_1})$$
$$\times P(q_{t_1} \,|\, q_0, t_1, N_{e1}) P(n_{t_21}, n_{t_22} \,|\, q_{t_2})$$
$$\times P(q_{t_2} \,|\, q_0, t_1, t_2, N_{e1}, N_{e2}). \qquad (5)$$

Notice that here we are interested in both estimates of N_{e1} and N_{e2} (presumably different). If we assume that $N_{e1} = N_{e2}$ or we are interested only in the average N_e, (5) is greatly simplified in computation. Equation (5) can be extended to more than three samples. The computation intensity increases prohibitively, however, with increasing sampling points if the N_es for each sampling period are to be estimated jointly. A continuous model of growth in population size can be fitted instead.

For a population with effective size growing or shrinking at each generation, it is more desirable to fit a continuous growth model. A simple model is exponential growth, with effective size at generation t determined completely by growth rate (r) and initial size (N_{e0}): $N_{et} = r^t N_{e0}$. W&S99 have considered this model but not evaluated the performance of the likelihood estimation because of the computational intensity of their method. Using $m+1 (m > 1)$ samples taken at generations $t_i (i = 0, ..., m)$, the maximum likelihood function can be extended to

$$P(n_{t_01}, n_{t_02}; ...; n_{t_m1}, n_{t_m2} \,|\, t_0, ..., t_m, N_{e0}, r)$$
$$= \sum_{q_0 \cdots q_{t_m}} [P(n_{t_01}, n_{t_02} \,|\, q_{t_0}) P(q_{t_0} \,|\, N_{e0}, r, t_0)] \dots$$
$$[P(n_{t_m1}, n_{t_m2} \,|\, q_{t_m}) P(q_{t_m} \,|\, N_{e0}, r, t_m)]. \qquad (6)$$

To obtain the likelihood estimates of N_{e0} and r simultaneously, (6) is calculated with the transition matrix $\mathbf{M}$ changing in dimensions at each generation. At generation t_i, for example, the numbers of rows and columns of $\mathbf{M}$ are $C_i = 2N_{et_i} + 1$ and $C_{i-1} = 2N_{et_{i-1}} + 1$, respectively.

More sophisticated growth models with more parameters, such as logistic growth, can also be fitted. The computational intensity, however, increases rapidly with the number of parameters to be estimated simultaneously.

(iii) *Computation*

The computation of the likelihood function – equation (1), (5) or (6) – could be problematic when N_e is large. A large transition matrix not only takes a lot of memory but also incurs intensive computation. Here I extend the computational algorithm of the likelihood method of W&S99 in the following respects.

First, the number of elements in the transition matrix (**M**) actually calculated and used in the likelihood evaluation is reduced to a very small fraction of the original total. For simplicity, consider (1) as an example. In the likelihood method it is necessary to calculate and use **M** repeatedly in matrix multiplication with different values of N_e, and the value that gives the maximum likelihood is taken as the N_e estimate. With a 'golden section' search (Press *et al.*, 1992), it requires typically about 15 iterations to obtain the maximum likelihood estimate. This means that **M** is calculated 15 times and used in matrix multiplication by $15 \times$ total number of alleles over loci $\times t$ times. Reducing **M**, therefore, decreases both the memory requirement and the computing time.

In the present work, only elements of **M** that are large enough to contribute significantly to the transition of allele frequency distributions are evaluated, stored and utilized. For row i, these elements are m_{ii} and its neighbours. Because the diagonal element m_{ii} is generally (not always) the largest for row i (denoted as m_i), m_{ij} is calculated from $j = i$ in both directions (j decreasing and increasing). Once m_{ij} is smaller than a threshold value compared with m_i (e.g. $m_{ij}/m_i < 10^{-5}$), the calculation is terminated for row i for the direction. Essentially, the resulting **M** contains only diagonal and adjacent elements, a small fraction of the total original number of C^2. The extent of reduction increases with increasing values of N_e. When the threshold value is 10^{-5}, for example, the numbers of elements of **M** evaluated and utilized are reduced to about 4% and 1% of the original 2001^2 and 20001^2 for $N_e = 1000$ and 10000, respectively.

For more than two samples and N_es different among sampling periods (equation 5 or 6), **M** is even more important in determining the efficiency of likelihood calculation because it must be evaluated more times for each iteration. When the population size changes, say, from N_{e1} to N_{e2}, **M** has $C_2 = 2N_{e2}+1$ rows and $C_1 = 2N_{e1}+1$ columns. It can be calculated similarly to the square **M** matrix described above. The largest element for row i is first estimated as m_{iJ}, where J is the nearest integer of iC_1/C_2. Calculation of m_{ij} for row i starts from column J in both directions and the largest element is kept updated.

Second, the likelihood function – equation (1), (5) or (6) – is calculated more efficiently by using the hidden Markov chain algorithm (Baum, 1972; Anderson *et al.*, 2000). Let us denote the prior

distribution of initial allele frequency by the column vector **W**, and the probabilities $P(n_{i1}, n_{i2}|q_i)$ of the data observed at generation i ($0 \leqslant i \leqslant t$) given different values of q_i by the column vector $\mathbf{U}_i$. If the componentwise vector multiplication of two column vectors **A** and **B** of length r is denoted by the symbol $\langle \mathbf{A}, \mathbf{B} \rangle = (a_1b_1, a_2b_2, \ldots, a_rb_r)^\mathrm{T}$, then the probabilities of population allele frequencies conditional on the sample taken at time 0, denoted by the column vector $\mathbf{V}_0$, are calculated by

$$\mathbf{V}_0 = \frac{\langle \mathbf{W}, \mathbf{U}_0 \rangle}{\phi_0}, \tag{7}$$

where the normalizing constant $\phi_0 = \mathbf{W}^\mathrm{T}\mathbf{U}_0$. The conditional probabilities of the population allele frequency at generation i ($i = 1, \ldots, t$) given all the N_es and samples taken from the population at generations up to and including i, denoted by the column vector $\mathbf{V}_i$, are calculated by

$$\mathbf{V}_i = \mathbf{M}_{i-1}\mathbf{V}_{i-1} \tag{8}$$

if no sample is taken at generation i, or by

$$\mathbf{V}_i = \frac{\langle \mathbf{M}_{i-1}\mathbf{V}_{i-1}, \mathbf{U}_i \rangle}{\phi_i} \tag{9}$$

if a sample is taken at generation i, where $\phi_i = [\mathbf{M}_{i-1}\mathbf{V}_{i-1}]^\mathrm{T}\mathbf{U}_i$. When the recurrent calculations with (8–9) reach generation t (when the last sample was taken), the likelihood is calculated by

$$\prod_{i \in \Omega} \phi_i, \tag{10}$$

where Ω is the set of generations at which a sample was drawn from the population.

To further simplify the algorithm, ϕ_i ($0 \leqslant i < t$) is not calculated and is dropped from (7) and (9). Corresponding to this treatment, the likelihood is simply ϕ_t.

Notice that the above algorithm applies to the likelihood evaluation with N_e either changing (equation 5 or 6) or constant (equation 1). In the former case, **M** also changes over generations as indicated by the subscript. For calculating the distribution of allele frequencies at generation i, for example, $\mathbf{M}_{i-1}$ is used which has $2N_{e, i-1}+1$ columns and $2N_{e, i}+1$ rows.

Third, in calculating $\mathbf{V}_i$ and $\mathbf{U}_i$ ($i = 0, \ldots, t$) in the above hidden Markov chain algorithm, only elements of large values are evaluated, stored and used, in a way similar to determining **M**. For $\mathbf{U}_i$, these involve the Jth element (J: the nearest integer to $2N_e n_{i1}/n_i$), which is usually the largest, and its neighbours. The larger the allele sample size (n_i), the greater the reduction of the number of elements in these vectors actually evaluated and utilized in computation. In evaluating $\mathbf{V}_i$ in (8–9), I also start the calculation in

opposite directions from the largest element (which is also predictable) and terminate the calculation once the element is smaller than a threshold value relative to the largest one. The extent of reduction for $\mathbf{V}_i$ with $i > 0$ depends on both the sample sizes and N_es up to and including generation i.

Because of these improvements in computing the likelihood function, the memory requirement is greatly decreased, and the computing speed is increased dramatically. The program can cope with populations with N_e of the order of 10^5 and marker loci with any number of alleles.

3. Simulation methods

Simulations were run to compare the performance of the pseudo-likelihood method with that of the full likelihood method in the case of tri-allelic loci, and against the F-statistic method over a wide range of parameters. The changes in allele frequencies in a diploid population of effective size N_e were simulated by Monte Carlo method, and samples of alleles were taken binomially (sampled with replacement) at generations 0 and t. For each replicate, the likelihood estimate was obtained by using the golden section search (Press *et al.*, 1992) for the maximum likelihood, and the F-statistic estimate was calculated by

$$\hat{N}_e = \frac{t}{2(\hat{F} - 1/n_0 - 1/n_t)}, \tag{11}$$

where $\hat{F}$ was computed as

$$\hat{F} = \frac{(q_0 - q_t)^2}{(q_0 + q_t)/2 - [(q_0 + q_t)/2]^2}. \tag{12}$$

$\hat{F}$ was calculated for each allele and the average over alleles and loci was used in (11) (Nei & Tajima, 1981; Waples, 1989). Other F-statistic estimators (e.g. Nei & Tajima, 1981; Pollak, 1983) gave very similar results to (11–12) (Waples, 1989; W&S99) and are not used in the present study. When $\hat{N}_e < 0$, all the observed change in gene frequency and the increase in F can be explained by sampling alone and the effective size is regarded, therefore, as infinite in this case (Waples, 1989).

For each set of parameters, 10000 replicates were run. To reduce running time, a replicate was terminated once the likelihood estimate of N_e was determined to be larger than the threshold value of $10N_e$ in comparing the pseudo-likelihood and F-statistic methods, and of $4N_e$ in comparing the pseudo- and full likelihood methods. For a fair comparison, all replicates with either estimator larger than the threshold were discarded and only those remaining were used to calculate the means and standard deviations of the N_e estimates (W&S99). The numbers of replicates with $\hat{N}_e$ greater than the threshold value

from each estimator and either of the two estimators in comparison were recorded. By discarding the 2·5 % smallest and the 2·5 % largest estimates from the 10000 replicates, the range of the 95 % remaining estimates (denoted as 95 % R) was also obtained for each method.

4. Simulation results

(i) *Comparison between the pseudo-likelihood and full likelihood methods*

The dependence of the frequencies of different alleles at a locus increases with a decreasing number of alleles at the locus. For a bi-allelic locus, the allele frequencies are completely dependent. In this case, however, the pseudo-likelihood method reduces to the full likelihood method. The worst situation for the pseudo-likelihood method is therefore 3 alleles per locus, which was considered in simulations to compare the performance of the pseudo- and full likelihood methods. Even though the same algorithm and computing tricks are applied, the full likelihood method is still highly computationally demanding for tri-allelic loci. As a result, only very small populations were considered ($N_e = 10$) and a replicate was terminated once the estimated effective size was determined to be larger than $4N_e$. The simulation results for various combinations of bi-allelic and tri-allelic loci, resulting in the same total number of independent alleles (12, total number of alleles over loci – number of loci), are shown in Table 1.

As can be seen, the two methods yield similar results for any combination of loci. For each method, the results are also similar among combinations of loci. This indicates that the approximation made in the pseudo-likelihood method works well, and the performance of both methods depends mainly on the total number of independent alleles across loci (other parameters such as allele frequency being kept constant).

If the dependence among the frequencies of alleles at a locus is important and such dependence is neglected by the pseudo-likelihood method, then it would be expected to overestimate the amount of information in the data and underestimate the confidence interval, as indicated by an increase in the proportion of replicates in which the true value of N_e falls outside of the confidence interval defined by a drop of 2 units in log-likelihood. Table 1 shows that there is little difference in confidence interval between the two methods for any combination of loci.

For a larger actual N_e, it is difficult to compare the two methods directly because of the intensive computation and storage requirement for the full likelihood method with multi-allelic loci. Instead, the appropriateness of the pseudo-likelihood method can

Table 1. *Comparison between the pseudo-likelihood and full likelihood methods*

No. of loci/no. of alleles	Pseudo-likelihood estimate					Full likelihood estimate				
	μ_{PL}	σ_{PL}	d_{PL}	$95\%R_{PL}$	P_{PL}	μ_{FL}	σ_{FL}	d_{FL}	$95\%R_{FL}$	P_{FL}
12/2	9·9	4·1	50	4·9, 22·5	0·109	9·9	4·1	50	4·9, 22·5	0·109
1/3 + 10/2	9·9	4·1	51	4·9, 22·5	0·106	10·1	4·1	52	4·9, 23·4	0·102
2/3 + 8/2	9·9	4·0	63	4·9, 22·5	0·098	10·3	4·1	67	4·9, 23·4	0·097
4/3 + 4/2	10·1	4·0	59	4·9, 22·8	0·090	10·8	4·3	61	4·9, 24·3	0·086
6/3	10·1	4·1	67	4·9, 23·4	0·084	11·2	4·4	85	5·4, 25·9	0·088

The simulated populations have a true effective size of $N_e = 10$. Various combinations of the numbers of bi-allelic and tri-allelic loci, giving the same total number (12) of independent alleles, were used. The initial frequencies are 0·1 for $k-1$ alleles and $1-0·1(k-1)$ for the remaining allele at a k-allele locus ($k = 2$ or 3). Samples of 50 individuals were taken with replacement from the population at generations 0 and 2. For each population, 10 000 replicates were run and a replicate was terminated once the estimated N_e was determined to be larger than 40. d_{PL} (d_{FL}) is the number of replicates in which the pseudo (full) likelihood estimate was larger than 40. For each estimator, the mean, μ, and standard deviation, σ, of the estimates from the $10\,000-d$ retained replicates are listed. $95\%R_{PL}$ ($95\%R_{FL}$) is the range of the remaining estimates after discarding the 2·5% smallest and the 2·5% largest from the 10 000 replicate estimates for the pseudo (full) likelihood method. P_{PL} (P_{FL}) is the proportion of the 10 000 replicates in which the maximum log-likelihood is larger by at least 2 than the log-likelihood of $N_e = 10$ for the pseudo (full) likelihood method.

Table 2. *Comparison between the* F-*statistic and pseudo-likelihood estimators: initial allele frequencies*

q_0	F-statistic estimate				Maximum likelihood estimate				d
	μ_F	σ_F	d_F	$95\%R_F$	μ_{PL}	σ_{PL}	d_{PL}	$95\%R_{PL}$	
0·01	207	109	87	89, 571	116	47	4	62, 239	87
0·02	156	87	44	70, 422	116	58	12	55, 276	46
0·04	134	82	39	58, 373	123	71	25	55, 330	42
0·08	129	88	65	52, 414	126	83	62	52, 400	75
0·16	130	93	55	49, 429	128	90	53	49, 417	57
0·32	130	93	77	48, 442	128	91	72	48, 437	78
0·50	130	96	83	49, 464	129	94	82	49, 456	83

The simulated populations have a true effective size of $N_e = 100$. Twenty bi-allelic loci, each with the initial frequency of the uncommon allele being q_0, were used. Samples of 100 individuals were taken with replacement from the population at generations 0 and 4. For each population, 10 000 replicates were run, and the number of replicates discarded due to either estimator being larger than $10N_e$ (1000) is d. For each estimator, the mean, μ, and the standard deviation, σ, of the estimates from the $10\,000-d$ retained replicates are listed. d_F (d_{PL}) is the number of replicates in which the F-statistic (pseudo-likelihood) estimate was larger than $10N_e$. $95\%R_F$ ($95\%R_{PL}$) is the range of the remaining estimates after discarding the 2·5% smallest and 2·5% largest from the 10 000 replicate estimates for the F-statistic (pseudo-likelihood) method.

be checked indirectly by comparing the accuracy and precision of the pseudo-likelihood estimates using data on different combinations of bi-allelic and multi-allelic loci, as shown in Table 3 (see below).

The likelihood computation is greatly simplified by using the hidden Markov chain algorithm in which all matrices and vectors are reduced by considering elements of large values only. Elements of small values are not calculated, stored or used at all. The selected threshold value determines the range of elements that are calculated and used, and therefore determines the computing time and storage, and the precision. Fig. 1 shows the changes in the likelihood curve with different threshold values for a simulated data set. As can be seen, the log-likelihood curve does not change essentially once the threshold value is greater than

0·001 in this case. An appropriate threshold value for a particular data set depends partly on sample size and the actual effective size. The larger these sizes are, the larger the threshold value that could give satisfactory results. To be conservative, a small threshold value of 10^{-10} was used in all the simulations shown below.

(ii) *Comparison between the pseudo-likelihood and* F-*statistic methods*

Because of the computational ease of the pseudo-likelihood method, it is possible to compare it with the widely used F-statistic method in performance over broad ranges of different parameters by extensive simulations.

Table 3. *Comparison between the* F-*statistic and pseudo-likelihood estimators: numbers of marker loci and alleles per locus*

No. of loci/ no. of alleles (k)	F-statistic estimate				Pseudo-likelihood estimate					
	μ_F	σ_F	d_F	$95\%R_F$	μ_{PL}	σ_{PL}	d_{PL}	$95\%R_{PL}$	P_{PL}	d
30/2	120	60	7	59, 276	116	57	5	59, 257	0·046	7
15/3	121	64	15	56, 287	117	59	13	58, 268	0·049	17
10/4	121	62	12	56, 279	117	56	12	58, 264	0·049	12
2/16	120	57	10	59, 267	115	51	6	59, 248	0·040	10
15/2 + 1/16	123	67	7	58, 291	116	55	5	59, 256	0·044	8
5/4 + 1/16	121	63	15	57, 282	116	55	12	59, 259	0·044	16
15/2 + 5/4	122	65	12	57, 292	117	58	12	58, 269	0·051	15
3/2 + 4/4 + 1/16	122	64	16	57, 293	116	54	12	59, 263	0·046	18
9/2 + 2/4 + 1/16	123	68	21	57, 293	116	56	10	59, 261	0·046	22
12/2 + 1/4 + 1/16	123	65	8	58, 293	116	54	5	59, 256	0·045	8

The simulated populations have a true effective size of $N_e = 100$. Various combinations of loci with different numbers of alleles, giving the same total number (30) of independent alleles, were used. The initial frequencies are 0·0625 for $k-1$ alleles and $1-0.0625(k-1)$ for the remaining allele at a locus. Samples of 100 individuals were taken with replacement from the population at generations 0 and 4. See Table 2 for the explanations for d, μ, σ and $95\%R$, and Table 1 for the explanation of P_{PL}.
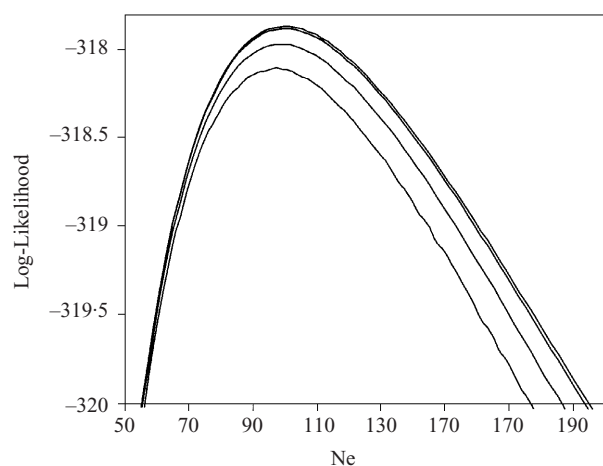

Fig. 1. Log-likelihood curves for a simulated data set using different threshold values in likelihood calculation. The curves upward from the bottom are generated using threshold values of 0·01, 0·005, 0·001 and 0·0001, respectively. The data set was generated by simulation, with samples of 100 individuals taken at generations 0 and 4 from a population with $N_e = 100$. Four loci, each with 10 co-dominant alleles (9 of the alleles have an initial frequency of 0·0625), were used.

(a) Allele frequencies

It has been shown that the F-statistic estimator typically performs poorly in the presence of rare alleles (e.g. Waples, 1989; Luikart et al., 1999), but little is known about the likelihood method. Table 2 compares the performance of the two estimators for a wide range of initial allele frequencies. The main difference in performance between the two estimators occurs when rare alleles are involved. As the initial frequency of the uncommon allele (q_0) decreases below 0·04, the F-statistic method leads to an increasingly larger upward bias, while the likelihood method gives increasingly more accurate estimates. The precision of the likelihood estimator is also higher than the F-statistic estimator across the range of allele frequencies, the contrast being evident especially when q_0 is small. The likelihood estimator performs only slightly better than the F-statistic estimator if no rare alleles are involved ($q_0 > 0.04$).

Simulations for populations with other parameters were also run. The occurrence of a sample containing rare alleles depends on t/N_e and sample sizes (n). The smaller is the value of nN_e/t, the more likely are samples with rare alleles for a given initial allele frequency. For a population with parameters $N_e = 20$, $t = 10$, $n = 10$, and 20 bi-allelic loci, for example, the F-statistic method gives much worse estimates ($\hat{N}_e = 31 \pm 15$, $95\%R = 14$–71) than the likelihood method ($\hat{N}_e = 21 \pm 10$, $95\%R = 10$–46) even if the initial frequency of the uncommon allele is as high as 0·16.

The changes in performance of the two estimators with allele frequencies are expected because the F-statistic estimator uses only the first two moments of the allele frequency distribution, while the likelihood estimator uses all the information on the distribution. With an intermediate value of the population allele frequency, the distribution of sample allele frequency is close to symmetric and can be reasonably described by the first two moments. With population allele frequency approaching 0 or 1, the sample allele frequency distribution becomes more and more skewed, and higher orders of moments are required for a better description of the distribution.

Although a prior uniform distribution of q_0 is assumed in the likelihood method, the likelihood
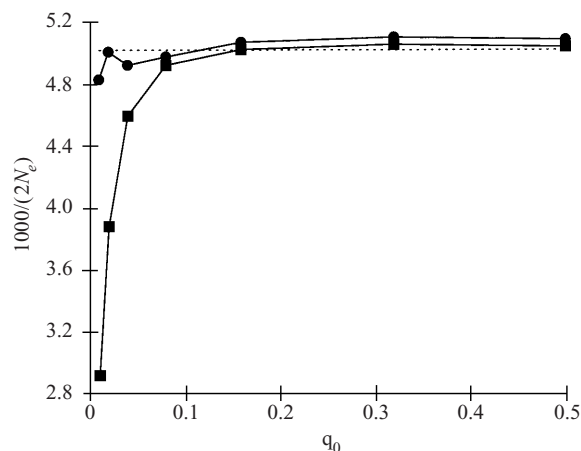
Fig. 2. Changes in the mean estimates of $1/(2N_e)$ from likelihood and F-statistic methods with initial allele frequencies (q_0). The simulated populations have a true value of $1/(2N_e) = 0.005$. Twenty bi-allelic loci, each with the initial frequency of the uncommon allele being q_0, were used. Samples of 100 individuals were taken with replacement from the population at generations 0 and 4. The means were taken over 10000 replicates, with $\hat{N}_e > 10N_e$ for the likelihood method and $\hat{N}_e < 0$ for the F-statistic method being regarded as $\hat{N}_e \to \infty$ and $1/(2\hat{N}_e) = 0$. Circles, likelihood estimate; squares, F-statistic estimate; dashed line, true value.

estimator performs well for any initial allele frequency across the whole range (Table 2). This demonstrates that the likelihood estimator is robust to violations of this assumption (W&S99).

Both estimators tend to overestimate N_e. This is generally true for various values of the parameters involved (Tables 2 and those as shown below). They tend to give much smaller biases for the estimate of $1/(2N_e)$, however, as shown in Fig. 2 The likelihood estimates are essentially unbiased for any initial allele frequency, while the F-statistic estimates are downward-biased substantially for $1/(2N_e)$ only when rare alleles are included. The much smaller bias of estimates for $1/(2N_e)$ than for N_e is generally true for wide ranges of values of different parameters. This is because both methods are actually estimating the average $1/(2N_e)$ rather than N_e from the changes in allele frequency at several loci. Fortunately, in most formulae it is $1/(2N_e)$ that occurs and matters, not N_e. Therefore, we need not worry about the small upward bias of N_e estimates.

(b) *Numbers of marker loci and alleles per locus*

Various combinations of marker loci with different numbers of alleles (k), resulting in the same total number of independent alleles (30), are compared in Table 3 for the accuracy and precision of the estimates from the F-statistic and the pseudo-likelihood methods. For all the 10 sets of parameters, the initial frequency of each of the $k-1$ alleles is assumed to be

0.0625 while that of the remaining allele is $1 - 0.0625(k-1)$ at each locus. As can be seen, the pseudo-likelihood method gives more accurate and precise estimates than the F-statistic method for all the combinations of loci. For a given total number of independent alleles, the numbers of loci and alleles do not affect the performance of either estimator. The approximation in the pseudo-likelihood by converting a multi-allelic locus to bi-allelic 'loci' is satisfactory, because both the accuracy and precision (as indicated by σ_{PL}, d_{PL}, $95\% R_{PL}$ and P_{PL}) of the estimates are essentially the same among combinations of loci with different numbers of alleles.

The above assumption of initial allele frequencies is aimed at minimizing their effects on the comparison among different combinations of loci used in the estimation. In reality, loci with more alleles are more likely to have rare alleles. If the initial allele frequencies were drawn from a uniform distribution, then the performance of the F-statistic method decreases rapidly and that of the pseudo-likelihood method keeps almost constant with an increasing number of alleles per locus on average among the 10 combinations of loci (results not shown). To reduce the large upward bias of the F-statistic estimator, rare alleles can be pooled. Pooling rare alleles, however, could result in less precise estimates with more frequent extreme values when few loci with many alleles are used in the estimation, and the low precision is even more problematic than bias in practice.

With an increasing number of independent alleles, the precision and accuracy of both methods improve and the difference in performance between the two methods diminishes (data not shown), as expected. The contrast of the two estimators is more obvious with a smaller data set (smaller samples of fewer independent alleles) and small t/N_e.

(c) *Sample size*

With the systematic forces excluded, the observed temporal changes in allele frequency come from genetic drift accumulating over generations in the population and the sampling. The effect of the latter could be reduced to get a better estimate of N_e by increasing sample size. Table 4 compares the performance of the two methods for different diploid sample sizes. With increasing sample sizes, both methods improve their accuracy and precision, as expected. At all sample sizes, the pseudo-likelihood estimator always has a better performance than the F-statistic estimator, but the difference decreases as the sample size increases.

The small sample size can be compensated by using more markers. With the same parameters as those in the first row of Table 4 but using 20 marker loci, the estimates are improved greatly to $\hat{N}_e = 149 \pm 97$,

Table 4. *Comparison between the* F-*statistic and pseudo-likelihood estimators*: *sample sizes*

	F-statistic estimate				Pseudo-likelihood estimate				
n	μ_F	σ_F	d_F	$95\%R_F$	μ_{PL}	σ_{PL}	d_{PL}	$95\%R_{PL}$	d
25	162	125	493	55, ∞	125	90	211	50, 769	554
50	139	74	17	67, 325	111	42	5	62, 217	18
100	126	38	0	76, 220	108	25	0	72, 170	0
200	121	27	0	80, 184	109	19	0	80, 152	0
400	118	23	0	81, 171	111	16	0	84, 147	0

The simulated populations have a true effective size of $N_e = 100$. Ten loci, each having 8 alleles with their initial frequencies drawn from a uniform distribution, were used in the estimations. Samples of different numbers (n) of diploid individuals were taken with replacement from the population at generations 0 and 4. See Table 2 for the explanations for d, μ, σ and $95\%R$.

Table 5. *Comparison between the* F-*statistic and pseudo-likelihood estimators*: t/N$_e$

	F-statistic estimate				Pseudo-likelihood estimate				
t	μ_F	σ_F	d_F	$95\%R_F$	μ_{PL}	σ_{PL}	d_{PL}	$95\%R_{PL}$	d
1	151	135	969	42, ∞	153	113	690	58, ∞	1135
2	147	107	176	54, 702	130	73	92	60, 415	197
4	138	71	13	66, 319	114	44	2	63, 229	14
8	140	51	0	75, 263	107	30	0	65, 181	0
16	153	50	0	85, 274	104	26	0	65, 167	0

The simulated populations have a true effective size of $N_e = 100$. Five loci, each having 8 alleles with their initial frequencies drawn from a uniform distribution, were used in the estimations. Samples of 100 individuals were taken with replacement from the population at generations 0 and t. See Table 2 for the explanations for d, μ, σ and $95\%R$.

$95\%R = 66$–491 and $d_F = 83$ for the F-statistic method, and to $\hat{N}_e = 115 \pm 60$, $95\%R = 58$–276 and $d_{PL} = 15$ for the pseudo-likelihood method.

For both estimators, simulations demonstrate that the performance is determined mainly by the harmonic mean of sample sizes (data not shown). For a given total sampling input, therefore, it is better to equalize sample sizes.

(d) *Interval of sampling* (t) *and* N$_e$

The extent of genetic drift (changes in allele frequency) that occurred in the population between sampling is determined directly by t/N_e, rather than the absolute values of N_e and t. Better performance is expected for both estimators with large nt/N_e (n: sample size) so that the changes in allele frequencies are caused mainly by genetic drift rather than sampling. In Table 5, the two estimators are compared for a population with $N_e = 100$ and different intervals between sampling. Prolonging the sampling interval increases both the accuracy and the precision of the pseudo-likelihood method, and the precision only of the F-statistic method. The upward bias of the F-statistic method first decreases, then increases with t. This is because a large t results in more rare alleles in the second sample, which leads to an overestimation of N_e for the F-statistic method.

Varying N_e for a given t has similar effects on the performance of the two estimators (data not shown), as the changes in population allele frequency due to drift are determined by t/N_e.

For a population with small t/N_e, a large data set (more independent alleles and large samples) would be necessary to get a reasonably good estimate. The worst estimate listed in row 1 of Table 5 can be improved to $\hat{N}_e = 122 \pm 68$, $95\%R = 55$–300 and $d_F = 21$ for the F-statistic estimator and to $\hat{N}_e = 129 \pm 54$, $95\%R = 73$–265 and $d_{PL} = 9$ for the pseudo-likelihood estimator, when samples of 300 individuals (instead of 100) are used.

(e) *Three samples*

With three samples taken sequentially at generations 0, t_1 and t_2, a joint likelihood estimation of $\hat{N}_{e1}$ and $\hat{N}_{e2}$ for the periods 0–t_1 and t_1–t_2, respectively, can be obtained. The harmonic mean of $\hat{N}_{e1}$ and $\hat{N}_{e2}$, weighted by t_1 and t_2–t_1 respectively, gives the average effective size ($\hat{N}_e$) during the whole period (0–t_2). For the F-statistic method, $\hat{N}_{e1}$ and $\hat{N}_{e2}$ can be obtained by using the corresponding samples. Two F-statistic estimates of N_e can be calculated. One is estimated directly from samples taken at generations 0 and t_2, and the other is the harmonic mean of $\hat{N}_{e1}$ and $\hat{N}_{e2}$ if both are positive. The final estimate of $\hat{N}_e$ is calculated as the average of

Table 6. *Comparison between the* F-*statistic and pseudo-likelihood estimators: three samples*

	F-statistic estimate				Pseudo-likelihood estimate				
Population	μ_F	σ_F	d_F	$95\%R_F$	μ_{PL}	σ_{PL}	d_{PL}	$95\%R_{PL}$	d
Population 1									
$N_{e1} = 100$	144	116	271	50, ∞	124	83	126	51, 548	291
$N_{e2} = 20$	25	9	0	13, 49	21	6	0	12, 36	0
$N_e = 33$	42	13	0	24, 73	35	8	0	22, 55	0
Population 2									
$N_{e1} = 20$	24	8	0	13, 43	21	5	0	13, 33	0
$N_{e2} = 100$	144	120	385	45, ∞	133	105	246	46, 998	428
$N_e = 33$	39	11	0	24, 66	34	8	0	23, 53	0

Ten loci, each having 4 alleles with their initial frequencies drawn from a uniform distribution, were used in the estimations. Samples of 50 individuals were taken with replacement from the population at generations 0, 4 and 8. N_{e1}, N_{e2} and N_e are true effective sizes for generations 0–4, 5–8 and 0–8, respectively. See Table 2 for the explanations for d, μ, σ and $95\%R$.

Table 7. *Pseudo-likelihood estimates of the initial effective size* (N_{e0}) *and exponential growth rate* (r)

		$\hat{N}_{e0}$			$\hat{r}$			
Population	No. of samples	μ_{PL}	σ_{PL}	$95\%R_{PL}$	μ_{PL}	σ_{PL}	$95\%R_{PL}$	d
Population 1: $N_{e0} = 20$, $r = 1\cdot2$								
	3	22	12	9, 53	1·24	0·16	0·94, 1·60	22
	4	21	10	10, 45	1·23	0·14	0·97, 1·54	8
Population 2: $N_{e0} = 100$, $r = 0\cdot8$								
	3	134	111	40, 435	0·80	0·11	0·59, 1·04	39
	4	122	80	45, 320	0·80	0·09	0·62, 0·99	4

Ten loci, each having 4 alleles with their initial frequencies drawn from a uniform distribution, were used in the estimations. Samples of 100 individuals were taken with replacement from the population at either generations 0, 3 and 6 (for the case of three sample), or generations 0, 2, 4 and 6 (four samples). N_{e0} and r are true initial effective size and exponential growth rate. Replicates with either $\hat{N}_{e0} > 10N_{e0}$, $\hat{r} > 1\cdot5r$ or $\hat{r} < 0\cdot5r$ were discarded, and the number of discarded replicates is denoted as d. See Table 2 for the explanations for μ, σ and $95\%R$.

the two estimates if both are in the proper range (positive and smaller than $10N_e$); otherwise the estimate in the proper range is selected. The F-statistic and pseudo-likelihood estimators are compared in Table 6 for two populations with either increasing or decreasing N_e.

As can be seen, the pseudo-likelihood method gives more accurate and precise estimates for all three effective sizes of each of the two populations. For the same data, pseudo-likelihood estimates using two samples were also obtained. They are better than F-statistic estimates but are slightly worse than the joint likelihood estimates using three samples simultaneously. The difference is obvious, especially for N_{e1} of population 1 and N_{e2} of population 2, where more marker information is required in the estimation. The pseudo-likelihood estimates using two samples are $\hat{N}_{e1} = 126 \pm 89$, $95\%R = 49$–654 and $d_{PL} = 164$ for population 1, and $\hat{N}_{e2} = 134 \pm 105$, $95\%R = 44$–∞ and $d_{PL} = 285$ for population 2.

The pseudo-likelihood estimator using three samples is better than that using two samples because the former utilizes all the information available.

Temporal samples from the same population are inter-related and each has implications for all the others. If an allele is not observed in the first two samples but appears in the third for example, then we know that it exists in the population during the whole observation period. The two-sample likelihood estimator cannot use this kind of information for estimating N_{e1}, in contrast to the three-sample estimator. With more rare alleles involved in the samples, therefore, we expect greater difference in performance between the estimators.

(f) *Continuous growth model*

Using three or more temporal samples, a continuous growth model can be fitted to the data and pseudo-likelihood estimates of growth parameters can be obtained. The numerical examples for the simple exponential growth are shown in Table 7. With a reasonable amount of marker information, both initial effective size (N_{e0}) and growth rate (r) are satisfactorily estimated. For a growing population, r tends to be overestimated, while for a shrinking population N_{e0} is

overestimated. The estimates for both parameters in both populations improve with more temporal samples. Similarly, increasing marker information for each sample (number of independent marker alleles, sample size) also results in better estimations (data not shown).

More generally, the pseudo-likelihood method can be used to test different hypotheses about the population dynamics using three or more samples. For a growing population, for example, different growth models (e.g. exponential, logistic) can be fitted to the same data and the one with the largest maximum likelihood would be the likely pattern of growth for the population.

5. Analysis of an empirical data set

A real data set that has been analysed several times by the *F*-statistic method (Krimbas & Tsakas, 1971; Nei & Tajima, 1981; Pollak, 1983; Waples, 1989) was re-analysed by the pseudo-likelihood method. Three samples were taken in September or October of 1966, 1967 and 1968, respectively, from an isolated population of the olive fly, *Dacus oleae*, infesting an orchard of about 2000 olive trees near Athens, Greece (Krimbas & Tsakas, 1971). The samples were analysed for the gene frequencies at two esterase loci, A and B, which were highly polymorphic and included 18 and 13 electrophoretic alleles, respectively. The sample sizes for years 1966, 1967 and 1968, respectively, were 474, 312 and 400 for locus A, and 469, 281 and 409 for locus B. For this population, there were about four generations in one year (Krimbas & Tsakas, 1971). When the samples were taken (autumn), the actual size of this population was apparently very large compared with the effective size. Therefore, the samples could be regarded as taken with replacement (Krimbas & Tsakas, 1971; Nei & Tajima, 1981).

The effective sizes of the population during periods 1966–1967 and 1967–1968, N_{e6-7} and N_{e7-8}, and the average, N_{e6-8}, can be estimated using either two samples each time or three samples simultaneously by the pseudo-likelihood method. The 95% confidence interval for the two-sample likelihood method can be calculated as the range of support associated with a drop of 2 in the log-likelihood. For the three-sample likelihood method, the confidence intervals are obtained with likelihood ratio tests. Consider the confidence interval of N_{e6-7} as an example. The likelihood tests begin by defining two hypotheses – H$_1$: $N_{e6-7} > 0$, $N_{e7-8} > 0$ and H$_2$: $N_{e6-7} = N_{Alt}$, $N_{e7-8} > 0$. Let us define L_1 and L_2 as the highest likelihood possible under H$_1$ and H$_2$, respectively. L_1 is obtained with the maximum likelihood estimates of N_{e6-7} and N_{e7-8}, and L_2 is obtained by selecting the value of N_{e7-8} that maximizes the likelihood given N_{Alt}. A 95% confidence interval for N_{e6-7} can be obtained (Rice,

1995) by finding the highest and lowest values of N_{Alt} that conform to the inequality $(2\ln(L_1/L_2) \leqslant \chi^2_{0.05, 1})$. The confidence intervals for N_{e7-8} can be obtained similarly. The confidence intervals for N_{e6-8} are calculated from those for N_{e6-7} and N_{e7-8}. The results of the pseudo-likelihood estimates are shown in Table 8, together with the *F*-statistic estimates from Nei & Tajima (1981) and Pollak (1983) using different *F* estimators. Several interesting things emerge from Table 8.

First, the estimated effective sizes for 1967–1968 are much smaller than those for 1966–1967. This is true irrespective of the marker loci and the methods used in the estimation. The explanation is that extensive spray of insecticide was applied to the population in 1968 and, as a result, the census size was reduced substantially in that year (Krimbas & Tsakas, 1971).

Second, the two *F*-statistic estimators from Nei & Tajima (1981) and Pollak (1983), as well as that of Krimbas & Tsakas (1971) (not listed in Table 8), give essentially the same results. Extensive simulations also showed that different *F*-statistic estimators result in similar estimates (e.g. Waples, 1989). Compared with *F*-statistic estimators, the pseudo-likelihood method using either two or three samples gives much more consistent estimates of each of three effective sizes (N_{e6-7}, N_{e7-8} and N_{e6-8}) among locus A, locus B and both loci. Also, the pseudo-likelihood method results in much narrower 95% confidence intervals. All these findings indicate that the pseudo-likelihood method gives more precise estimates for this data set. This is not surprising given the simulation results shown above and the fact that several rare alleles are involved in the samples. There are four alleles on locus A and two alleles on locus B that are observed in only one of the three samples.

Third, the average effective size for the period 1966–1968 ($\hat{N}_{e6-8}$), calculated as the harmonic mean of $\hat{N}_{e6-7}$ and $\hat{N}_{e7-8}$, tends to be smaller than that estimated directly using samples of years 1966 and 1968 for both the pseudo-likelihood and *F*-statistic methods. This trend is especially evident for locus B. The pseudo-likelihood method using three samples simultaneously results in, therefore, much larger estimates of $\hat{N}_{e6-7}$ compared with that using two samples each time. The reason that $\hat{N}_{e6-7}$ rather than $\hat{N}_{e7-8}$ is increased by using three samples is that the former has a much larger confidence interval than the latter.

There are indications that $\hat{N}_{e6-7}$ was underestimated by the *F*-statistic method. Krimbas & Tsakas (1971) noticed that, before the spray of insecticide in 1968, the census population size of olive flies became minimum in winter and, at that time, there were still on average two flies per tree. If this estimate is correct, the minimum census size of this population is roughly 4000, much larger than the *F*-statistic estimate of N_{e6-7}. Assuming that the population was in equi-

Table 8. *Pseudo-likelihood and F-statistic estimates of the effective size and 95 % confidence intervals (in parentheses) of an olive fly population*

	F-statistic method		Pseudo-likelihood method	
Locus used	Nei & Tajima	Pollak	Two samples	Three samples
1966–1967				
A	583 (240, ∞)	576	1013 (376, 17466)	2553 (712, ∞)
B	1056 (314, ∞)	923	1022 (259, ∞)	2316 (433, ∞)
A + B	722 (332, 7408)	687	1016 (433, 5976)	2455 (817, ∞)
1967–1968				
A	168 (86, 538)	186	238(146, 421)	223 (130, 401)
B	234 (100, 1984)	227	263 (124, 659)	292 (126, 767)
A + B	189 (108, 446)	200	246 (162, 392)	242 (153, 396)
1966–1968				
A	291 (145, 965)	289	401 (223, 758)	410 (220, 802)
B	762 (313, 12069)	400	592 (231, 1796)	519 (195, 1534)
A + B	400 (225, 961)	326	450 (271, 769)	442 (257, 791)

The data were from Krimbas & Tsakas (1971). The *F*-statistic estimates were computed by Nei & Tajima (1981) and Pollak (1983). The pseudo-likelihood estimates were obtained using two samples each time or three samples simultaneously.
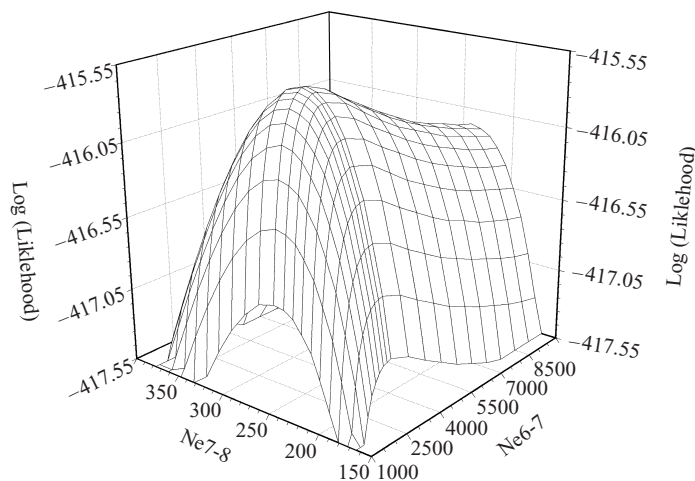

Fig. 3. Changes of log-likelihood with estimated effective sizes for years 1966–1967 (N_{e6-7}) and 1967–1968 (N_{e7-8}). The data are from Krimbas & Tsakas (1971).

librium before 1968, they also estimated the mutation rate, *u*, by the formula

$$n_e = 4N_e u + 1$$

using their estimates of N_{e6-7} (about 500) and the effective number of alleles, n_e, which was estimated to be 4·2 and 2·4 for loci A and B, respectively. The estimated mutation rates for loci A (1·6 × 10⁻³) and B (7·0 × 10⁻⁴) are both quite high because of the low estimate of N_{e6-7} (Krimbas & Tsakas, 1971).

Nei & Tajima (1981) also suspected that N_{e6-7} was underestimated by the *F*-statistic method, and one possible cause was ascribed to local sampling. With the three-sample pseudo-likelihood method, however, $\hat{N}_{e6-7}$ is now increased to 2455. It could be larger, because the likelihood changes slowly with increasing $\hat{N}_{e6-7}$ above 2455, as shown in Fig. 3. As can be seen,

the 95 % confidence interval for $\hat{N}_{e7-8}$ is narrow, but that for $\hat{N}_{e6-7}$ is quite broad. Actually, there is no upper limit of the 95 % confidence interval for $\hat{N}_{e6-7}$. With $\hat{N}_{e7-8} = 242$, even $\hat{N}_{e6-7} = 50000$ gives a log likelihood of −416·59, smaller than the maximum value of −415·55 at $\hat{N}_{e6-7} = 2455$ by only about 1.

It may seem to be strange that the pseudo-likelihood estimates of N_e for this data set are larger than the *F*-statistic estimates regardless of the time period and number of samples used, given the extensive simulation results which show the opposite trend. However, we should notice that the simulation results are averages over a large number of replicates. For any parameter combination there are always some replicates in which the likelihood estimates are smaller than the *F*-statistic estimates. In an extreme case of row 2 in Table 2 where the uncommon allele frequency

is 0·02, for example, the average likelihood estimate is much smaller than the *F*-statistic estimate due to the large proportion of rare alleles involved in the samples. However, there is still a proportion of 16 % replicates where the likelihood estimates are larger than the *F*-statistic estimates. Compared with the example, the proportion of rare alleles in the olive fly data set is much smaller and the chances of likelihood estimates being larger than the *F*-statistic estimates should be greater. Another possible explanation is that there is one null allele at each of the loci used and this is not accounted for in the pseudo-likelihood calculation. The data for the likelihood method (allelic counts) were actually converted from the allele frequency data listed in Table 1 of Krimbas & Tsakas (1970).

Using the present pseudo-likelihood method (program), the analysis of the data set is completed without computational difficulty. It takes about 60–120 min for a PIII PC to obtain the likelihood estimates of effective size and their confidence intervals, with $2N_e$ taking values as high as 100 000 in the iterations.

6. Discussion

In the present study the previous maximum likelihood approach to estimating effective population size using temporal data on gene frequencies (W&S99) is extended by simplifying the computation with several efficient algorithms, and employing a reasonable approximation to the likelihood function for multi-allelic loci. These developments reduce the computational and storage demands of the likelihood method considerably, and thus make it possible to assess the performance of the likelihood method against the *F*-statistic method over a wide range of parameters by extensive simulations. They also enable the use of the likelihood method to estimate the effective size of large populations using highly polymorphic loci (e.g. microsatellites) in practice without much computing difficulty.

This study confirms the conclusion that the likelihood estimator gives more accurate and precise estimates of N_e than the *F*-statistic estimator (W&S99) under a much wider range of parameters. I have shown that inclusion of rare alleles is the main cause for the performance difference. This is understood since the sample frequency of rare alleles has a more skewed distribution that cannot be described satisfactorily by the first two moments only. Considering the widespread use of highly polymorphic markers (e.g. microsatellites), which inevitably result in a substantial proportion of rare alleles in samples of reasonable sizes, the likelihood estimator is obviously more desirable in practical applications.

The present study has also demonstrated that more than two alleles per locus can be manipulated in the

likelihood estimator by considering each allele and pooling all the others, similar to the *F*-statistic estimator. Although this treatment does not utilize fully the marker information available and the calculated likelihood is only an approximation because the converted 'loci' are not completely independent, it makes the computational problem manageable and yet provides satisfactory estimates (Tables 1, 3). Simulations show that the pseudo-likelihood method gives essentially the same results (in terms of precision and accuracy) as the full likelihood method for the case of tri-allelic loci (Table 1). Simulations involving the same total number of independent alleles but different combinations of loci with different numbers of alleles (Table 3) also indicate that the pseudo-likelihood method works well. It should be noted that the dependence of allele frequencies at a locus is partially accounted for by the pseudo-likelihood method in which the log-likelihood for a *k*-allelic locus is calculated as the sum of the log-likelihoods over the *k* converted bi-allelic 'loci' multiplied by a correction factor $(k-1)/k$. In practice, Anderson *et al.*'s (2000) multi-allele likelihood method can be applied to small problems because of the high computational demand, while the pseudo-likelihood method is appropriate for large problems involving highly polymorphic loci (many alleles per locus), large population size or many samples.

The performance of any estimator based on the temporal changes in gene frequency depends on the proportion of such changes due to genetic drift against sampling, and the amount of marker information. The former is determined approximated by nt/N_e, while the latter depends mainly on the number of independent alleles in the samples. For the *F*-statistic estimators, gene frequency is also a decisive factor when it is close to one or zero. To increase the precision and accuracy of the temporal estimators, the sampling intervals, sample size and number of independent alleles need to be large. These factors compensate each other to a certain extent. With small nt/N_e, for example, a good estimation of N_e could still be possible if many independent alleles are used. It should be pointed out that the temporal approach applies equally well to populations of all sizes, large or small. The important determinants are nt/N_e, the number of independent alleles used, and gene frequency for the *F*-statistic estimator.

Both the *F*-statistic and likelihood methods apply to non-ideal populations (data not shown). One non-ideal aspect, non-random mating or population structure, needs to be considered more carefully, however, for both likelihood and *F*-statistic methods. When matings occur more frequently than at random between closely related individuals, the genotypic frequencies depart from Hardy–Weinberg proportions. To estimate the gene frequency, therefore,

samples need to be taken representatively from the whole population. Local sampling could result in biased estimation of N_e for both estimators. When samples from a local population are used in the estimation, the estimate would be close to the effective size of the local population sampled if the migration rate and sampling interval are small so that no migrant genes are sampled. On the contrary, the estimate would be close to the effective size of the entire population if the migration rate and/or sampling interval are/is large even though samples are drawn from a local population. Without knowing the genetic structure of the population under study, it is better to draw samples from the entire population (Nei & Tajima, 1981).

Except for sampling and genetic drift, systematic forces can also cause a shift in gene frequency over time. The assumptions of the absence of these systematic forces, mutation, selection and migration (as discussed above), seem to be not very restrictive to the application of the temporal approach in most situations. Due to the low mutation rate and usually short sampling interval, mutation can be safely ignored (Nei & Tajima, 1981; W&S99). Beneficial or deleterious genes or neutral genes closely linked to strongly selected genes can change in frequency as a result of selection. As long as the change is small relative to that due to genetic drift because of the small N_e, however, selection can also be neglected without causing much error.

References

Anderson, E. C., Williamson, E. G. & Thompson, E. A. (2000). Monte Carlo evaluation of the likelihood for N_e from temporally spaced samples. *Genetics* **156**, 2109–2118.

Baum, L. E. (1972). An inequality and associated maximization technique in statistical estimation for probabilistic function of Markov processes. In *Inequalities-III: Proceedings of the Third Symposium on Inequalities held at the University of California, Los Angeles, September 1–9, 1969* (ed. O. Shisha), pp. 1–8. New York: Academic Press.

Caballero, A. (1994). Developments in the prediction of effective population size. *Heredity* **73**, 657–679.

Crow, J. F. & Kimura, M. (1970). *An Introduction to Population Genetics Theory*. New York: Harper & Row.

Dobzhansky, T. & Wright, S. (1941). Genetics of natural populations. V. Relations between mutation rate and accumulation of lethals in populations of *D. pseudoobscura*. *Genetics* **26**, 23–51.

Ewens, W. J. (1979). *Mathematical Population Genetics*. New York: Springer.

Feller, W. (1950). *An Introduction to Probability Theory and Its Applications*. New York: Wiley.

Fiumera, A. C., Wu, L., Parker, P. G. & Fuerst, P. A. (1999). Effective population size in the captive breeding program of the Lake Victoria cichlid *Paralabidochromis chilotes*. *Zoo Biology* **18**, 215–222.

Fiumera, A. C., Parker, P. G. & Fuerst, P. A. (2000). Effective population size and maintenance of genetic diversity in captive-bred populations of a Lake Victoria cichlid. *Conservation Biology* **14**, 886–892.

Frankham, R. (1995). Conservation genetics. *Annual Review of Genetics* **29**, 305–327.

Funk, W. C., Tallmon, D. A. & Allendorf, F. W. (1999). Small effective population size in the long-toed salamander. *Molecular Ecology* **8**, 1633–1640.

Hill, W. G. (1981). Estimation of effective population size from data on linkage disequilibrium. *Genetical Research* **38**, 209–216.

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

Kantanen, J., Olsaker, I., Adalsteinsson, S., Sandberg, K., Eythorsdottir, E., Pirhonen, K. & Holm, L. E. (1999). Temporal changes in variation of North European cattle breeds. *Animal Genetics* **30**, 16–27.

Krimbas, C. B. & Tsakas, S. (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.

Labate, J. A., Lamkey, K. R., Lee, M. & Woodman, W. L. (1999). Temporal changes in allele frequencies in two reciprocally selected maize populations. *Theoretical and Applied Genetics* **99**, 1166–1178.

Luikart, G. & Cornuet, J. M. (1999). Estimating the effective number of breeders from heterozygote excess in progeny. *Genetics* **151**, 1211–1216.

Luikart, G., Cornuet, J. M. & Allendorf, F. W. (1999). Temporal changes in allele frequencies provide estimates of population bottleneck size. *Conservation Biology* **13**, 523–530.

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

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

Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. (1992). *Numerical Recipes in Fortran 77*, 2nd edn. Cambridge: Cambridge University Press.

Pudovikin, A. I., Zaykin, D. V. & Hedgecock, D. (1996). On the potential for estimating the effective number of breeders from heterozygote-excess in progeny. *Genetics* **144**, 383–387.

Rice, J. A. (1995). *Mathematical Statistics and Data Analysis* 2nd edn. Belmont: Wadsworth.

Roff, D. A. (1997). *Evolutionary Quantitative Genetics*. London: Chapman & Hall.

Schwartz, M. K., Tallmon, D. A. & Luikart, G. (1999). Using genetics to estimate the size of wild populations: many methods, much potential, uncertain utility. *Animal Conservation* **2**, 321–323.

Turner, T. F., Richardson, L. R. & Gold, J. R. (1999). Temporal genetic variation of mitochondrial DNA and the female effective population size of red drum (*Sciaenops ocellatus*) in the northern Gulf of Mexico. *Molecular Ecology* **8**, 1223–1229.

Wang, J. & Caballero, A. (1999). Developments in predicting the effective size of subdivided populations. *Heredity* **82**, 212–226.

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

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

Wright, S. (1931). Evolution in Mendelian populations. *Genetics* **16**, 97–159.