Analysis of Selection in Populations Observed over a Sequence of Consecutive Generations

I. Some One Locus Models with a Single, Constant Fitness Component per Genotype

ALAN R. TEMPLETON

Department of Human Genetics and the Society of Fellows, University of Michigan, Ann Arbor, Michigan (USA)

Summary. A statistical model is presented for dealing with genotypic frequency data obtained from a single population observed over a run of consecutive generations. This model takes into account possible correlations that exist between generations by conditioning the marginal probability distribution of any one generation on the previously observed generation. Maximum likelihood estimates of the fitness parameters are derived and a hypothesis testing framework developed. The model is very general, and in this paper is applied to random-mating, selfing, parthenogenetic and mixed random-mating and selfing populations with respect to a single locus, g-allele model with constant genotypic fitness differences with all selection occurring either before or after sampling. The assumptions behind this model are contrasted with those of alternative techniques such as minimum chi-square or "unconditional" maximum likelihood estimation when the marginal likelihoods for any one generation are conditioned only on the initial conditions and not the previous generation. The conditional model is most appropriate when the sample size per generation is large either in an absolute sense or in relation to the total population size. Minimum chi-square and the unconditional likelihood are most appropriate when the population size is effectively infinite and the samples are small. Both models are appropriate when the samples are large and the population size is effectively infinite. Under these last conditions, the conditional model may be preferred because it has greater robustness with respect to small deviations from the underlying assumptions and has a greater simplicity of form. Furthermore, if any genetic drift occurs in the experiment, the minimum chi-square and unconditional likelihood approaches can create spurious evidence for selection while the conditional approach will not. Worked examples are presented.

I. Introduction

A common experimental design used in population genetics is to follow the changes of gene or genotypic frequencies within a population over time in order to discover evidence for selection. In this and subsequent papers I will be concerned with statistical inference in such populations followed over a run of successive generations. This paper in particular will be concerned with g allele, one locus selection models consisting of a single constant fitness component for each genotype in random-mating, selfing, parthenogenetic and mixed random-mating and selfing populations.

Many authors have addressed themselves to the problem of estimating and testing single-locus selection from such data. Some (Workman and Jain, 1966; Lorenz, 1970 a, b, c; Stam, 1971; Allard, Kahler and Weir, 1972) have estimated and tested fitness differences using single generation transitions. When data on several consecutive generations exists, such procedures can be used on each transition observed.

In many situations, however, the population geneticist desires to impose a single fitness model on the entire data set. In such cases, it is inappropriate to estimate fitnesses separately on each generation transition. Furthermore, estimators based on successive pairs of generations contain a common set of observations. Anderson (1969) and Stam (1971) have shown how the covariance caused by the sharing of observations can be calculated, but these covariances still cause serious difficulties in hypothesis testing. Furthermore, there is often another type of dependence in the data that is totally ignored — the Markovian dependence. More of this will be said later.

To avoid some of these difficulties, Wright and Dobzhansky (1946) developed a least squares technique for estimating the fitness parameters for constant fitness and frequency dependent models based on multi-generation data, but their method considers each pair of successive generations independently of all others. Levene et al. (1954) pointed out this causes the fitness estimators to be influenced mainly by the early part of the experiment when the change in gene frequency is large. They stated that, "What is needed is a method of estimation which will, first, consider the observations in their proper sequence and, second, give them their proper weights." They suggested a minimum chi-square procedure in which

the theoretical frequencies for each generation are calculated in a deterministic fashion from the initial gene frequencies using a set of constant fitnesses. Next, a chi-square is calculated for the differences between the observed and the calculated frequencies. Finally, various sets of fitnesses are tried until a set is found which gives the smallest possible chi-square. They noted that this procedure is efficient and essentially equivalent to the method of maximum likelihood. However, these optimal properties are known to exist only when the data points are independent of one another.

A maximum likelihood procedure analogous to this minimum chi-square procedure has been proposed by DuMouchel and Anderson (1968). As in Levene et al., they use the initial gene frequencies and a fixed set of constant fitness values to calculate in a deterministic fashion the expected allele frequencies for each generation. A multinomial sample of the genes is assumed to be taken every generation with the expected allele frequencies in the sample set equal to the theoretical deterministic values. Finally, each generation is treated as if it were statistically independent from all other generations. Maximum likelihood scoring is then used to find a set of fitnesses that maximizes the likelihood for the entire experiment. Others (Cavalli, 1950; Anderson, 1969; Prout, 1969) have also used maximum likelihood procedures, but in all cases independence of the data at each generation is assumed.

However, most evolutionary biologists model evolution as a Markov process such that the probabilities of events at any one generation are highly dependent upon what happened in the past. Consequently, observations made on a single population over time will with some sampling designs represent a highly dependent set of data. Under these conditions statistical techniques which treat the data points as if they were mutually independent can be misleading. For example, Gastwirth and Rubin (1971) have studied this statistical problem for data that has a correlation only between successive data points and Bhat (1961) has studied the distributions of goodness of fit statistics for data from a Markow chain or sequence. These works show that if a null hypothesis is tested with the data treated as if it were independent, the true level of significance is different from the level of significance the experimentor believes he has. When the correlation is positive (which will be the case in the type of Markov processes considered in this paper) the true level of significance is higher than the set level of significance, causing the rejection of the null hypothesis more often than is statistically justified for the set level of significance. To see how such a situation could arise, consider an experiment in which the number of heterozygotes at each of m consecutive generations is observed in a self-mating population. Suppose the experimentor tests for selection acting on the heterozygotes. With

no selection, the expected frequency of heterozygotes at the ith generation is $(1/2)^i p_0$ where p_0 is the initial frequency of heterozygotes. Suppose there is no selection but, just by chance, there is an excess of heterozygotes at generation i. The experimentor could take such random perturbations into account by incorporating a probability distribution on the number of heterozygotes at one generation such as the binomial with a mean of $n_i(1/2)^i p_0$ where n_i is the population size at generation i. However, given an excess of heterozygotes at generation i, an excess of heterozygotes over the value $n_{i+1}(1/2)^{i+1} p_0$ is expected for generation $i + 1$ due solely to the Markovian nature of the process; and similarly for the next few generations. A statistican who treats each generation independently might not attach any special significance to the initial deviation, but could easily conclude there was a significant deviation when the series of generations is combined. The statistician who views this as a Markov process would expect the subsequent deviations even when no selection was operating and therefore would give less significance to them. Because of the positive correlation that exists between generations, the procedures that treat generations as if they were independent tend to accentuate fitness differences and lead to the rejection of the null hypothesis of no selection more often than is justified. It is therefore important to derive estimation and hypothesis testing procedures which explicitly take into account the Markovian dependence that exists in many data sets.

II. Some Basic Theorems and Results

The first step in obtaining maximum likelihood estimators is to obtain the likelihood function for the entire data set. The general data set considered is the array of genotypic numbers (or frequencies since the likelihood function will always be conditioned upon the total sample size every generation) over m consecutive generations. No reference will be made at this point about the mode of reproduction of the population, mating structure, nature of the selection, ploidy or number of loci and alleles. Let X_{ij} be the number of individuals of genotype j at generation i where $\sum_{j=1}^{h} X_{ij} = N_i$ and h is the number of different genotypes. Then let $\tilde{X}_i = (X_{i1}, X_{i2}, ..., X_{ih})$, the genotypic array at generation i, and let $\tilde{X} = (\tilde{X}_1, ..., \tilde{X}_m)$, the entire data set over m generations. $\tilde{X}$ represents a realization from an evolutionary process which is assumed to be Markovian such that
$$f(\tilde{X}_i : \tilde{X}_{i-1}, ..., \tilde{X}_0, \theta) = f(\tilde{X}_i : \tilde{X}_{i-1}, \theta)$$
where $f(\tilde{X}_i : \tilde{X}_{i-1}, \theta)$ denotes the probability density of $\tilde{X}_i$ given $\tilde{X}_{i-1}$, the genotypic array of the previous generation, and some fitness parameters denoted by θ. In all the models considered in this paper it is assumed that the conditional distribution of the genotypic

array at generation i given the genotypic array at the previous generation and the parameters θ is multinomial:

$$f(\widetilde{X}_i: \widetilde{X}_{i-1}, \theta) \propto \prod_{j=1}^{h} [t_{i-1,j}(\theta)]^{X_{ij}},$$

where $t_{i-1,j}(\theta)$ is the probability of obtaining an individual of type j at generation i as a function of $\widetilde{X}_{i-1}$ and θ. These conditional probabilities take into account explicitly the Markovian dependence that exists between generations. The likelihood function for the entire run of m generations, $f(\widetilde{X}:\widetilde{X}_0, \theta)$ obtained from the relation

$$f(\widetilde{X}_i: \widetilde{X}_{i-1}, \theta)\, f(\widetilde{X}_{i-1}: \widetilde{X}_{i-2}, \theta) = f(\widetilde{X}_i, \widetilde{X}_{i-1}: \widetilde{X}_{i-2}, \theta) \tag{2.1}$$

as

$$f(\widetilde{X}: \widetilde{X}_0, \theta) \propto \prod_{i=1}^{m} \prod_{j=1}^{h} [t_{i-1,j}(\theta)]^{X_{ij}} .$$

The log likelihood, $L_m(\theta)$, is proportional to

$$L_m(\theta) \propto \sum_{i=1}^{m} \sum_{j=1}^{h} X_{ij} \ln [t_{i-1,j}(\theta)] . \tag{2.2}$$

θ represents the parameters of the selection model under consideration, and it is assumed that θ ranges over an open subset θ^* of r dimensional Eucledian space. The following condition is imposed on the Markov process (Billingsley, 1961, pg. 23):

Condition 2.1. Inference is limited to the set D of $(\widetilde{X}_{i-1}, \widetilde{X}_i)$ such that $f(\widetilde{X}_i: \widetilde{X}_{i-1}, \theta) > 0$ is independent of θ and has continuous partial derivatives of third order throughout θ^*. Moreover, the $dh \times r$ matrix with elements (d being the number of elements in D)

$$\partial f(\widetilde{X}_i: \widetilde{X}_{i-1}, \theta)/\partial \theta_u \quad (\widetilde{X}_{i-1}, \widetilde{X}_i) \in D, \ u = 1, ..., r$$

has rank r throughout θ^*. For each $\theta \in \theta^*$ there is only one ergodic set and there are no transient states.

For the type of multinomial models used in this paper, the first sentence in Condition 2.1 is meet if the fitnesses assigned to the various genotypes are all greater than zero. Hence, any lethal genotypes must be excluded from the transition matrix before construction of the likelihood function. The second sentence of the condition places certain restrictions on the number of parameters that can be treated, and in particular for the multinomial implies that only relative fitnesses can be dealt with. However, it is possible to have more than one relative fitness component per genotype provided enough transitions have been observed to satisfy the rank condition. The last sentence will in general not be met for multinomial transitions because there can be fixation or loss of certain genotypic classes. Consequently, inference can be made about the relative fitnesses of h genotypes only as long as there are transitions involving all h genotypes. The data set, $\widetilde{X}$, will

therefore refer only to that part of the experiment in which no fixation or loss has occurred.

Condition 2.1 allows the log likelihood to be differentiated, so that the maximum likelihood equations can be obtained as:

$$\partial L_m(\theta)/\partial \theta_u = \sum_{i=1}^{m} \sum_{j=1}^{h} \frac{X_{ij}}{t_{i-1,j}(\theta)} \frac{\partial t_{i-1,j}(\theta)}{\partial \theta_u} = 0$$
$$u = l, ..., r \tag{2.3}$$

The question of importance is when does a consistent solution, $\hat{\theta}$, of 2.3 exist. To answer this, Billingsley (1961) proved the following theorem:

Theorem 2.1. Suppose condition 2.1 is satisfied and that $\theta^°$ is the true value of the parameter. Then there exists a consistent solution, $\hat{\theta}$, of 2.3.

In many situations the experimentor would also like to consider more than one type of selection model. Consequently, in addition to the original parameter space θ^* of dimension r, suppose there is an alternative parameter space Φ^* within θ^* such that Φ^* is an open subset of a c dimensional Eucledian space where $c \leq r$. Billingsley proved the following theorem:

Theorem 2.2. Suppose condition 2.1 is satisfied for Φ^* and θ^*. If $\theta^°$ is the true parameter value, then

$$2\left(\max_{\theta^*} L_m - L_m^0\right) \to \chi_r^2$$
$$2\left(\max_{\Phi^*} L_m - L_m^0\right) \to \chi_c^2,$$
$$2\left(\max_{\theta^*} L_m - \max_{\Phi^*} L_m\right) \to \chi_{r-c}^2,$$

where $\max_p L_m$ denotes the log likelihood evaluated at the maximum likelihood estimators in the parameter space p, L_m^0 the log likelihood evaluated at the null hypothesis of dimension zero (which is the hypothesis of no selection in this paper), χ_v^2 denotes a chi-square distribution with v degrees of freedom and $\to$ denotes convergence in law. Furthermore, the last two statistics are asymptotically independent.

As an example of the use of this theorem, consider the testing of the goodness of fit of a selection model θ. A consistent solution is first found using theorem 2.1 so that the quantity $\max_{\theta^*} L_m$ can be evaluated. To test for goodness of fit, the transition probabilities are re-estimated under a non-parametric space, Ψ^*, such that equation 2.2 now becomes:

$$L_m \propto \sum_{i=1}^{m} \sum_{j=1}^{h} X_{ij} \ln (t_{i-1,j})$$

and 2.3 is obtained by differentiating with respect to each $t_{i-1,j}$ as:

$$X_{ij}/t_{i-1,j} - X_{ih}\bigg/\left[1 - \sum_{j=1}^{h-1} t_{i-1,j}\right] = 0$$
$$i = 1, ..., m, \ j = 1, ..., h-1 .$$

The solution of the above is

$$\hat{t}_{i-1,j} = X_{ij}/N_i .$$

Thus, the estimated transition probabilities are equated to the observed transition frequencies. The dimension of this non-parametric space is $m(h-1)$, so theorem 2.2 yields

$$2\left\{\sum_{i=1}^{m}\sum_{j=1}^{h} X_{ij}\left[\ln(0_{ij}) - \ln\left(e_{ij}(\widetilde{X}_{i-1}, \hat{\theta})\right)\right]\right\} \to \chi^2_{m(h-1)-r}$$

(2.4)

where 0_{ij} is the observed frequency of genotype j at generation i and $e_{ij}(\widetilde{X}_{i-1}, \hat{\theta})$ is the expected frequency of genotype j at generation i under θ^* given $\widetilde{X}_{i-1}$.

Another statistical problem is that of replicate populations. Consider two populations that are followed for m and s generations respectively. The observations in the two populations represent independent observations and the log likelihood for the entire data set is $L(\theta, \theta') = L_m(\theta) + L_s(\theta')$. In particular, a test for homogeneity is desired; that is, a test of the hypothesis $\theta = \theta'$, vs. $\theta \neq \theta'$. Note that the maximum likelihood solution of $L(\theta, \theta')$ when $\theta \neq \theta'$ is simply the maximum likelihood estimates of θ and θ' obtained from each population separately. The maximum likelihood estimate when $\theta = \theta'$ is obtained by solving

$$\partial L(\theta, \theta')/\partial\theta_u = \partial L_m(\theta)/\partial\theta_u + \partial L_s(\theta)/\partial\theta_u = 0$$
$$u = 1, ..., r .$$

Billingsley has shown that the statistic

$$2\left[\max_{\theta^*, \theta'^*} L(\theta, \theta') - \max_{\theta^*} L(\theta, \theta)\right] \to \chi^2_r \qquad (2.6)$$

under the hypothesis that there is some θ which is the true parameter for both populations.

Before discussing the genetic models in detail, one final statistical problem must be dealt with — the problem of incomplete sampling. Until now it has been assumed that both N_i and X_{ij}, the population size at generation i and the number of individuals of genotype j at i respectively, are observed. Often, however, a finite sample of size $n_i \leq N_i$ is taken every generation or the sample is not even drawn from the reproducing population (i. e., the sample individuals never contribute to the next generation) and x_{ij}, the number of individuals with genotype j in the sample, is observed. Consider, for simplicity, the case in which only one independent genotype is observed such that the population consists of X_i individuals of the genotype of interest at generation i and $N_i - X_i$ other individuals. Suppose the distribution of X_i given $X_{i-1}/N_{i-1} = q_{i-1}$ is binomial:

$$f(X_i: q_{i-1}, \theta) = \binom{N_i}{X_i} [t(q_{i-1}, \theta)]^{X_i} [1 - t(q_{i-1}, \theta)]^{N_i-X_i}$$

(2.7)

where $t(q_{i-1}, \theta)$ is the expected frequency of X_i given q_{i-1} and some parameter θ. Suppose a random sample is taken of size n_i at generation i. Then the sampling distribution is

$$\frac{\binom{X_i}{x_i}\binom{N_i - X_i}{n_i - x_i}}{\binom{N_i}{n_i}} . \qquad (2.8)$$

Combining (2.7) and (2.8) and summing over all possible values of the unobserved quantity X_i, the distribution of x_i is obtained as:

$$f(x_i: q_{i-1}, \theta) = \binom{n_i}{x_i} [t(q_{i-1}, \theta)]^{x_i} [1 - t(q_{i-1}, \theta)]^{n_i-x_i} .$$

(2.9)

Consequently, for one generation the distribution of x_i is still binomial. Unfortunately, it depends on the quantity $q_{i-1} = X_{i-1}/N_{i-1}$ which in general is unobserved. Also, relation 2.1 depends on the population values and not the sample values. However, it is possible to estimate the frequency q_i by $\hat{q}_i = x_i/n_i$. This estimate is unbiased and has variance

$$(1 - n_i/N_i)\, q_i(1 - q_i)/n_i , \qquad (2.10)$$

when the sample is drawn from the reproducing population or

$$q_i(1 - q_i)/n_i \qquad (2.11)$$

when the sample is not part of the reproducing population. Equation (2.10) goes to zero when either n_i approaches N_i or n_i gets very large, and (2.11) goes to zero when n_i gets very large. Under these conditions, $\hat{q}_i$ converges in probability to q_i, and the log likelihood (2.2) can be approximated by using the sample frequencies instead of the population frequencies. Given that this approximation is valid, all the previously discussed results follow.

This approximation will be best when the sample is drawn from the reproducing population and n_i represents a substantial proportion of N_i. Consequently, an optimal experimental design is to establish a finite population of such a size that a substantial sample of the population may be taken every generation. However, many experiments have not followed this design and instead have N_i very large and n_i considerably smaller. Under this design the approximation is still good when n_i is large. This implies that N_i is usually extremely large with the result that, given a true fitness model, the population genotypic frequencies at any generation will converge to their deterministic values calculated from the fitnesses and initial conditions. This, coupled with a large n_i every generation, implies that the log likelihood given by 2.2 evaluated at the sample frequencies will be approximately equal to the log likelihood obtained by assuming deterministic expected frequencies with independent, multinomial sampling every generation. Under these conditions (large n_i, $N_i \to \infty$) the model considered here yields as a special case a model analogous to the one employed by DuMouchel and Anderson (1968) for a

random mating population with all selection after sampling.

The final case occurs when the variance of $\hat{q}_i$ cannot be ignored. This will usually occur when $n_i \ll N_i$ and n_i is small or just when n_i is small if the sample is not part of the reproducing population. Under these conditions it would be better to view all the randomness in the process as sampling error and calculate the expected frequencies in a deterministic fashion providing N_i is large. However, a better procedure would be to explicitly take into account both the sampling error and the Markov nature of the evolutionary process, but this will be the subject of a later paper. In all the following sections of this paper it is assumed that either n_i is large with respect to N_i, or that n_i itself is very large.

III. Application to Diploid Populations with Constant Fitness Differences at a Single Locus

The results of the previous section will now be applied to the special case of a diploid population observed with respect to a single locus with g alleles. Instead of designating the genotypes by a single number j as before, it is now more convenient to use a pair of numbers, uv, where $u \leq v$ and u, $v = 1, 2, ..., g$. Thus, $h = (1/2)\ g(g+1)$ in equation 2.2. Let $x_{i,uv}$ be of number of individuals of genotype uv observed at generation i in a sample of size n_i, $q_{i,uv}$ the sample frequency of uv at i, $y_{i,r}$ the number of copies of allele r at i in the sample and $p_{i,r}$ the sample frequency of allele r at i. Suppose further that θ, the fitness model, consists of a single constant fitness parameter per genotype. This model will be designated by the symbol w and w_{uv} is the constant fitness parameter for genotype uv. Since these are relative fitnesses, one of the w's is set equal to one.

Under this type of fitness model, the conditional transition probabilities may be written as $t_{i,uv}(w) = d_{i,uv}(w)/\bar{w}_i$ where

$$\bar{w}_i = \sum_{u=1}^{g} \sum_{v=u}^{g} d_{i,uv}(w) \ .$$

$\bar{w}_i$ represents an average fitness and can always be scaled to equal one when all the w's are one (no selection). The exact form of the d's will, among other things, depend upon the time at which selection operates with respect to sampling. The model restricts all selection to occur either before sampling or after sampling for otherwise more than one fitness parameter per genotype would be required. If all selection occurs before sampling, the d's may be written as

$$d_{i-1,uv} = w_{uv}\, e_{i,uv} \qquad (3.1)$$

where $e_{i,uv}$ is the expected frequency of genotype uv at generation i given no selection and the genotypic state at $i - 1$. The maximum likelihood equations

or scores corresponding to 2.3 now take the form:

$$S_{uv} = \partial L_m / \partial w_{uv} = \sum_{i=1}^{m} (x_{i,uv}/w_{uv} - n_i e_{i,uv}/\bar{w}_{i-1}) = 0$$

$$u \leq v = 1, ..., g \ . \qquad (3.2)$$

An approximate solution when all the $\bar{w}$'s are close to one is

$$\hat{w}_{uv}^0 = \sum_{i=1}^{m} x_{i,uv} \div \sum_{i=1}^{m} n_i\, e_{i,uv} \ .$$

The $\hat{w}_{uv}^0$ closest to one can be set equal to one, and all fitnesses are then measured relative to it. If it is desired to measure fitnesses relative to a specific genotype or if none of the $\hat{w}^0$'s are close to one, all the initial fitness approximations can be divided by the initial fitness estimate of the genotype for which it is desired to have a relative fitness of one.

In order to obtain a more exact solution of 3.2, the following Newton-Raphson iteration is used. Let $\hat{w}_0$ be the vector of initial approximations and let δw_0 be a vector of small additive corrections such that $\hat{w}_0 + \delta w_0$ is a better approximation. Let $D(w)$ be the matrix whose elements are the second partial derivatives of the log likelihood which for the special case of all selection before sampling are

$$D_{uv,jk} = \partial^2 L_m / \partial w_{uv}\, \partial w_{jk} =$$

$$= \sum_{i=1}^{m} (n_i e_{i,uv} e_{i,jk}/\bar{w}_{i-1}^2 - \gamma x_{i,uv}/w_{uv}^2)$$

$$\begin{matrix} u \leq v = 1, ..., g \\ j \leq k = 1, ..., g \end{matrix} \qquad \gamma = \begin{cases} 1 \text{ if } uv = jk \\ 0 \text{ otherwise} \end{cases}$$

The D's are then evaluated at $\hat{w}_0$ and the vector of corrections is

$$\delta w_0 = [- D(\hat{w}_0)]^{-1} S(\hat{w}_0)$$

where $S(\hat{w}_0)$ is the vector of scores given by 3.2 evaluated at $\hat{w}_0$. The next approximation is obtained and the process repeated until convergence to yield the maximum likelihood estimates, $\hat{w}$. The quantity $-D(\hat{w})$ is the precesion matrix which converges with probability one to its expected value, the information matrix (Kendall and Stuart, 1973). Consequently, $[-D(\hat{w})]^{-1}$ estimates the variance-covariance matrix for the fitness estimators.

Once the maximum likelihood estimators have been obtained, hypotheses may be tested. First, the goodness of fit of the null hypothesis may be tested with

$$2 \sum_{i=1}^{m} \sum_{u=1}^{g} \sum_{v=u}^{g} x_{i,uv}[\ln (q_{i,uv}) - \ln (e_{i,uv})] \rightarrow \chi^2_{1/2\ (g+2)\ (g-1)m}.$$

This is simply a special case of statistic 2.4 in which $r = 0$ and $h = 1/2\ g(g+1)$. Similarly, the goodness of fit of the model of constant fitnesses can be tested with statistic 2.4 with $1/2\ (g+2)\ (g-1)\ (m-1)$ degrees of freedom since $1/2\ (g+2)\ (g-1)$ independent w's have been estimated. Finally, theorem 2.2 may be applied to test the constant fitness model vs.

the null hypothesis to yield:

$$2 \sum_{i=1}^{m} \sum_{u=1}^{g} \sum_{v=u}^{g} (x_{i,uv} \ln \hat{w}_{uv} - n_i \ln \hat{\overline{w}}_{i-1}) \rightarrow \chi^2_{1/2(g+2)(g-1)},$$

where $\hat{\overline{w}}_{i-1}$ is the average fitness evaluated at $\hat{w}$.

The exact forms for the e's will depend upon the mode of reproduction and mating pattern of the population. For an automictic parthenogenetic population,

$$\left. \begin{aligned} e_{i,uu} &= q_{i-1,uu} + 1/2 (1 - K) \sum_{\substack{j=1 \\ j \neq u}}^{g} q_{i-1,uj} \\ e_{i,uv} &= K q_{i-1,uv} \quad u \neq v \end{aligned} \right\} \quad (3.3)$$

where K is defined as the probability of a heterozygous adult giving rise to a heterozygous zygote and is a function of the type (or types) of automixis by which diploidy is restored and the linkage relationship of the locus to its centromere (Asher, 1970; Templeton and Rothman, 1973). Equation 3.3 can also be used for a self-mating population by setting $K = 1/2$.

For a random-mating population,

$$\left. \begin{aligned} e_{i,jj} &= p^2_{i-1,j} \\ e_{i,jk} &= 2 p_{i-1,j} p_{i-1,k} \quad j \neq k. \end{aligned} \right\} \quad (3.4)$$

Furthermore, a population reproducing by any mixture of automixis, selfing and random mating may be described by the appropriate mixture of (3.3) and (3.4).

When selection operates after sampling, the d's may not be factored as in (3.1). For an automictic or self-mating population,

$$\left. \begin{aligned} d_{i,uv} &= w_{uv} K q_{i,uv} \quad u \neq v \\ d_{i,uu} &= w_{uu} q_{i,uu} + 1/2 (1 - K) \sum_{\substack{v=1 \\ v \neq u}}^{g} w_{uv} q_{i,uv}. \end{aligned} \right\} \quad (3.5)$$

The maximum likelihood scores are

$$S_{uv} = \sum_{i=1}^{m} q_{i-1,uv} \left\{ \frac{K x_{i,uv}}{d_{i-1,uv}} + \right.$$
$$\left. + 1/2 (1 - K) \left[\frac{x_{i,uu}}{d_{i-1,uu}} + \frac{x_{i,vv}}{d_{i-1,vv}} \right] - \frac{n_i}{\overline{w}_{i-1}} \right\}$$
$$u \leq v = 1, ..., g. \quad (3.6)$$

An initial solution when (3.6) is set equal to zero is

$$\hat{w}^0_{uu} = \sum_{i=1}^{m} n_i q_{i-1,uu} (q_{i,uu} - e_{i,uu}) / e_{i,uu} \div \sum_{i=1}^{m} n_i q^2_{i-1,uu} / e_{i,uu},$$

$$\hat{w}^0_{uv} = \sum_{i=1}^{m} x_{i,uv} \div \sum_{i=1}^{m} n_i q_{i-1,uv} \times$$
$$\times [1 - 1/2 (1 - K) (q_{i,uu} / e_{i,uu} + q_{i,vv} / e_{i,vv})]$$
$$u < v.$$

The maximum likelihood solution can be found through scoring with

$$D_{uv,jk} = \sum_{i=1}^{m} q_{i-1,uv} q_{i-1,jk} \left\{ \frac{n_i}{\overline{w}^2_{i-1}} - \frac{K x_{i,uv}}{d^2_{i-1,uv}} \frac{\partial d_{i-1,uv}}{\partial w_{jk}} - \right.$$
$$\left. - 1/2 (1 - K) \left[\frac{x_{i,uu}}{d_{i-1,uu}} \frac{\partial d_{i-1,uu}}{\partial w_{jk}} + \frac{x_{i,vv}}{d_{i-1,vv}} \frac{\partial d_{i-1,vv}}{\partial w_{jk}} \right] \right\}$$

where

$$\partial d_{i-1,tz} / \partial w_{jk} = \begin{cases} 1 & \text{if } t = z = j = k \\ 1/2 (1 - K) & \text{if } t = z, \ j < k, \\ & \quad t = j \text{ or } t = k \\ K & \text{if } t = j < z = k \\ 0 & \text{otherwise}. \end{cases}$$

The goodness of fit of the null hypothesis and of the constant fitness model are tested with the appropriate forms of statistic (2.4) with $1/2 (g+2) (g-1) m$ and $1/2 (g + 2) (g - 1) (m - 1)$ degrees of freedom respectively. The test of the constant fitness model vs. the null hypothesis is

$$2 \sum_{i=1}^{m} \sum_{u=1}^{g} \sum_{v=u}^{g} x_{i,uv} [\ln(d_{i-1,vv}(\hat{w})) - \ln (e_{i,vv}) -$$
$$- n_i \ln (\hat{\overline{w}}_{i-1})]$$

which is asymptotically chi-square with $1/2 (g + 2) \times (g - 1)$ degrees of freedom.

For a random-mating population with all selection after sampling,

$$\left. \begin{aligned} d_{i,uu} &= p^2_{i-1,u}(w) \\ d_{i,uv} &= 2 p_{i-1,u}(w) \, p_{i-1,v}(w) \quad u < v \end{aligned} \right\} \quad (3.7)$$

where

$$p_{i-1,u}(w) = 1/2 (w_{uu} q_{i-1,uu} + \sum_{v=1}^{g} w_{uv} q_{i-1,uv}).$$

Mixtures of random-mating, automictic and selfing populations may be described with the appropriate mixture of (3.5) and (3.7). However, if the population is reproducing solely by random-mating, the log likelihood corresponding to (2.2) is

$$\sum_{i=1}^{m} \left\{ \sum_{u=1}^{g} \sum_{v=u}^{g} x_{i,uv} \ln (d_{i-1,uv}(w)) - 2 n_i \ln(\overline{w}) \right\} \alpha$$
$$\sum_{i=1}^{m} \left\{ \sum_{t=1}^{g} y_{i,t} \ln(p_{i-1,t}(w)) - 2 n_i \ln(\overline{w}_{i-1}) \right\} \quad (3.8)$$

where

$$\overline{w}_{i-1} = \sum_{t=1}^{g} p_{i-1,t}(w) = \sum_{u=1}^{g} \sum_{v=u}^{g} w_{uv} q_{i-1,uv}.$$

The impact of selection is thus manifested through the y's, the allele numbers, and not the x's, the genotype numbers for this special case. This in essence reduces the amount of information gained per generation since the number of genes is always less than the number of genotypes (unless some genotypes are excluded due to lethality a priori). Because of this, condition (2.1) will not be satisfied unless $m \geq 1/2 \times (g + 2)$ in contrast to the previously discussed cases in which $m \geq 1$ satisfies condition (2.1). It is also more convenient to apply theorems (2.1) and (2.2) to the allele number form of (3.8) as opposed to the genotype number form to yield

$$S_{uv} = \sum_{i=1}^{m} n_i q_{i-1,uv} (p_{i,u} / p_{i-1,u}(w) + p_{i,v} / p_{i-1,v}(w) - 2 / \overline{w}_{i-1})$$

$$D_{uv,jk} = \sum_{i=1}^{m} x_{i-1,uv} x_{i-1,jk} [2n_i \widetilde{w}_{i-1}^2 - \gamma_1 y_{i,u} / \hat{p}_{i-1,u}^2(w) -$$
$$- \gamma_2 y_{i,v} / \hat{p}_{i-1,v}^2(w)]$$

$$\gamma_1 = \begin{cases} 1 & \text{if } u = v = j = k \\ 1/2 & \text{if } u = v, \ j < k, \ u = j, \text{ or } u = k \\ 1/4 & \text{if } u < v, \ j < k, \ u = j \text{ or } u = k \\ 0 & \text{otherwise} \end{cases}$$

$$\gamma_2 = \begin{cases} 1/4 & \text{if } u < v, \ j < k, \ u = j \text{ or } u = k \\ 0 & \text{otherwise}. \end{cases}$$

The goodness of fit of the null hypothesis and of the constant fitness model are tested with statistic (2.4) using observed and expected allele numbers with $(g - 1) m$ and $(g - 1) \left(m - 1/2 (g + 2)\right)$ degrees of freedom respectively. The test of constant fitnesses vs. the null hypothesis is

$$2 \sum_{i=1}^{m} \left\{ \sum_{t=1}^{g} y_{i,t} [\ln (\hat{p}_{i-1,t}(\hat{w})) - \ln (\hat{p}_{i,t})] - 2n_i \ln(\widehat{w}_{i-1}) \right\}$$

which is asymptotically chi-square with $1/2 (g+2) \times (g - 1)$ degrees of freedom.

So far the maximum likelihood theory of Markov processes as developed by Billingsley has been applied to various genetic models. The strength of this theory lies in its general applicability, but when dealing with specific problems it is often possible to supplement this generalized statistical framework with more specialized and powerful statistical tools. For example, Templeton and Rothman (1973; hereafter referred to as T & R) examined in detail the analysis of "heterosis" in an automictic or self-mating population with respect to a single locus for which only two genotypic categories are distinguished: heterozygotes and homozygotes. The homozygotes are assigned a fitness of 1 and the heterozygotes a fitness of $1 + s$. The coefficient s can be regarded as a measure of "heterosis" in a very broad sense, or as the degree to which loss of heterozygosity is retarded or accelerated due to selection.

Assuming that all selection occurs before sampling, this situation is simply a special case of (3.2) and (3.3) with $w_{uv} = 1 + s$ for all $u < v$ and $w_{uu} = 1$ for all u. Under these conditions, the likelihood collapses into the form

$$f(\tilde{x}: s) \approx \prod_{i=1}^{m} \binom{n_i}{x_i} (1 + s)^{x_i} (1 + sKq_{i-1})^{-n_i} \times$$
$$\times (Kq_{i-1})^{x_i} (1 - Kq_{i-1})^{n_i - x_i} \quad (3.9)$$

where x_i and q_i are the number and frequency of heterozygotes observed at i (T & R only considered the case of constant population size, complete sampling and an initial population of heterozygotes only, but the dropping of these restrictions is straightforward under the conditions described at the end of section 2). One can apply Billingsley's theorems to (3.9) to obtain the maximum likelihood estimate of s and test the hypothesis $H_1: s \neq 0$ vs. $H_0: s = 0$, but

T & R also considered two other hypothesis testing tools based on the ratio:

$$f(\tilde{x}: s)/f(\tilde{x}: s = 0) = H(\tilde{x}, s) =$$
$$= (1 + s)^a \exp\left[-\sum_{i=1}^{m} n_i \ln (1 + sKq_{i-1}) \right]$$

where

$$a = \sum_{i=1}^{m} x_i.$$

The first test is to calculate the "odds" in favor of one hypothesis over the other which tells one which hypothesis is more likely and by how much. The odds of H_1 vs. H_0 with the possible values of s weighted by the prior distribution (as described in T & R)

$$f(1 + s) = (1 + s) e^{-(1+s)}$$

are obtained as

$$0(\tilde{x}) \approx e^b (b + 1)^{-a-2} (a + 1)! \approx$$
$$\approx \sqrt{2\pi/(a + 1)} \, [(a + 1)/(b + 1)]^{a+2} e^{b-a-1} \quad (3.10)$$

where

$$b = K \sum_{i=1}^{m} n_i q_{i-1}.$$

The second test was derived by differentiating $H(\tilde{x}, s)$ with respect to s and evaluating the derivative at $s = 0$ to yield the "locally best test" of $H_0': s \leq 0$ vs. $H_1': s > 0$. This test has maximum power out of all possible tests with the same significance level in the vicinity of $s = 0$. After proper normalization this test becomes to reject H_0' whenever $u > c(\alpha)$ where

$$u = \sum_{i=1}^{m} n_i(q_i - Kq_{i-1}) \div \sqrt{\sum_{i=1}^{m} K^i q_0 (1 - K^i q_0)} \quad (3.11)$$

and $c(\alpha)$ is a constant that depends upon α, the significance level of the test. T & R showed that u is normally distributed with mean 0 and variance 1 when sampling is complete and population size constant, and it is a straightforward extension of their proof that (3.11) is also normal zero, one under the sampling conditions discussed at the end of section 2. Consequently, $c(\alpha)$ is obtained from a standard normal table.

However, it is possible to develop even more statistical tools for this problem which give greater flexibility in data analysis. In deriving the distribution of u, it can be shown that the random variables

$$z_i = \sqrt{n_i}(q_i - Kq_{i-1}) \quad i = 1, \ldots, m$$

are asymptotically independent normals with mean zero and variance $K^i q_0 (1 - K^i q_0)$ for large n_i (T & R). Hence,

$$u' = \sum_{i \in I} z_i \div \left[\sum_{i \in I} K^i q_0 (1 - K^i q_0) \right]^{1/2}$$

where I is all or any subset of the generation indices is normal zero, one. Therefore, if data was not ob-

tained for some of the generations between 1 and m, those values of i for which $q_i - Kq_{i-1}$ cannot be evaluated are simply deleted from the sums in u to yield u'. Thus, even with data missing on some generations, it is still possible to test H_0' vs. H_1'. Also, u and u' can test H_0 vs. H_1 by regarding them as two-tailed tests rather than one-tailed tests.

The statistics u and u' are excellent test statistics for distinguishing $s = 0$ vs either $s = $ constant > 0 or $s = $ constant $\neq 0$, but there are some serious shortcomings when testing for the goodness of fit of the null hypothesis against all other alternatives. Suppose, for example, there is indeed selection but that s could be either positive or negative for any given generation. Since both u and u' measure a sum of deviations from the null expected values, a set of positive and negative deviations might yield a low u or u' value and lead to the acceptance of the null hypothesis. To avoid this difficulty, a sum of squares of the deviations can be taken. It follows from the distribution of the z's that when H_0 is true

$$v = \sum_{i=1}^{m} n_i(q_i - Kq_{i-1})^2 \div K^i q_0 (1 - K^i q_0) \to \chi_m^2,$$

$$v' = \sum_{i \in I} n_i(q_i - Kq_{i-1})^2 \div K^i q_0 (1 - K^i q_0) \to \chi_t^2$$

where I is a subset of the generation indices consisting of t elements. These statistics provide a more robust goodness of fit criterion for the null hypothesis than the u and u' statistics. Further illustrations of how these tests may be extended will now be provided in conjunction with the worked examples.

IV. Some Worked Examples

A. A predominately self-mating population

Allard, Kahler and Weir (1972) studied esterase allozymes in a predominately self-mating population of barley. They obtained seed samples ranging in size from 1006 to 4587 seeds per generation for 10 generations. The ten generations are not consecutive, however, but fall into three groups of 3, 4 and 3 consecutive generations respectively with gaps of 7 and 6 generations separating the groups. The seeds were scored for their genotype at four esterase loci and the percent of heterozygotes and gene frequencies were recorded. This data is given in Tables 1 through 4 of Allard et al. The frequency of outcrossing in this population was estimated to be 0.0057. Because this is such a small value, this data will be analyzed using a modification of the methods given in the last part of section III. The effect of outcrossing on the mean frequency of heterozygote given both the heterozygote frequency and gene frequencies of the previous generation is incorporated into the results of that section, but the effect of outcrossing on the higher moments of the conditional heterozygote frequency distribution is ignored; i.e., that portion of the conditional heterozygote frequency at each generation that is due solely to random mating is treated as a constant. This assumption is a good one because the frequency of outcrossing is very low and the model is used only to generate the moments for a *single* generation transition. Thus, the small amount of outcrossing present could have very large effects when compounded over several generations, but the assumptions used here only require that its effect over a single generation be small. This points out an advantage of the conditional approach over the method of using the model to generate the moments for all generations from the initial conditions — that is, *the conditional analysis is more robust to small deviations from the underlying assumptions than the corresponding unconditional model simply because such deviations are not compounded over several generations.*

The modified statistics become:

$$a = \sum_{i=1}^{m} n_i q_i$$

$$b = \sum_{i=1}^{m} n_i e_i$$

$$e_i = 1/2 \,(.9943)\, q_{i-1} + H_i,$$

$$H_i = (.0057) \sum_{j=1}^{g} \sum_{k=j+1}^{g} 2p_{i-1,j} p_{i-1,k},$$

$$0(\tilde{x}) = e^b (b + 1)^{-a-2} (a + 1)!$$

$$u'(\tilde{x}) = \sum_{i \in I} \sqrt{n_i}(q_i - e_i) \div \sqrt{\sum_{i \in I} G^i(q_0) \left(1 - G^i(q_0)\right)}$$

$$v'(\tilde{x}) = \sum_{i \in I} n_i(q_i - e_i)^2 / [G^i(q_0) \left(1 - G^i(q_0)\right)]$$

$$G(q_0) = .49715\, q_0 + H_1,$$

$$G^i(q_0) = G[G^{i-1}(q_0)] = .49715\, [G^{i-1}(q_0)] + H_i.$$

These statistics were applied to the data given in Tables 1, 2 and 3 of Allard et al., (1972) on esterases A, B and C. Each of the three runs of consecutive generations for these loci was treated separately with I corresponding to the indices only in a particular consecutive run and with the first generation in a run corresponding to $i = 0$. The statistics $u'(\tilde{x})$ and $v'(\tilde{x})$ were also used on all the observed generations for each locus with the index $i = 0$ now corresponding to the first generation with data (F_4 in Allard et al.) and going to $i = 22$ (F_{26}) with the missing generations deleted from the summations. However, $G^i(q_0)$ in general depends on some of the H_i from missing generations, so whenever H_i could not be observed directly the maximum H_i over all observed generations was substituted. Such a procedure would tend to bias the test against the hypothesis of "heterosis" ($s > 0$), but since the H's are relatively constant and small for a given locus such a bias is minor.

The results of this analysis are given in Table 1. The A locus shows a consistent pattern of a small amount of heterosis throughout the experiment which

Table 1. *Analysis of esterase heterozygosity in a barley population*[+]

Locus	Generations	Initial Est. of s	Max. Like. Est. of s	Standard Deviation	u Test	Odds[++]	v Test,	df
A	· 4 to 6	0.1774	0.1849	0.1227	1.7463	0.3134	3.7309	2
	14 to 17	0.1729	0.1760	0.1040	1.7971	0.3822	34.1887*	3
	24 to 26	0.0894	0.0908	0.1078	0.9227	0.1299	1.2015	2
	all				$u' = 3.0020*$		$v' = 78.2632*$	7
B	4 to 6	−0.4671	−0.4715	0.1189	−2.8419*	23.6882	8.6972*	2
	14 to 17	0.5147	0.5281	0.1250	6.2016*	7016.5800	376.0675*	3
	24 to 26	−0.1485	−0.1499	0.1086	−1.2358	0.2672	1.5944	2
	all				$u' = 1.0576$		$v' = 1066.7605*$	7
C	4 to 6	0.0118	0.0123	0.1014	−0.1293	0.0906	7.9279*	2
	14 to 17	−0.1644	−0.1685	0.0673	−2.0503*	1.0573	77.4991*	3
	24 to 26	0.1946	0.2046	0.0665	2.8980*	8.3321	197.5070*	2
	all				$u' = 0.4545$		$v' = 1259.0295*$	7

[+] Data from Allard, Kahler and Weir (1972).
[++] Odds less than one favor $s = 0$, odds greater than one favor $s = $ constant $\neq 0$.
* Significant at $P < .05$.

is significant for the experiment as a whole as shown by the u' and v' statistics. The B and C loci have drastic and significant shifts in their s values between runs, but with different patterns for each locus. Note that the u' tests fail to indicate any significant selection for either locus over all generations despite the overwhelming evidence for selection provided by the v' tests. This is attributable to the failure of the assumption of a constant "s" for the B and C loci. This analysis indicates that each of these three loci have experienced drastically different patterns of selection. This confirms and extends the conclusion given in Allard *et al.* that "... selection operated differentially on these three very tightly linked loci."

B. Random-mating with all selection before sampling

A natural population of the moth *Panaxia dominula* has been studied since the year 1939. This population is polymorphic for the *medionigra* gene with all three genotypes being phenotypically distinct in the adult. Data on this population is given in Table 9 of Ford (1971).

The samples were taken from the reproducing adult population each generation. The population size of the entire colony was estimated, and the sample usually amounted to 10 to 25% of the total population. In the years 1939—1961 the sample sizes ranged from a low of 117 to a high of 1,612. This sampling scheme is consistent with the assumptions of the conditional likelihood model for this time period and represents a data set that is undoubtedly affected both by sampling error and the Markovian error or genetic drift of the entire population. The model given by (3.4) will be applied to this data which assumes random-mating, constant fitnesses and all selection occurring before sampling.

The fitness estimators for the 23 generation sequence from 1939 until 1961 is given in Table 2. The goodness of fit of the null hypothesis is very poor, as is that of the constant fitness model. The test of constant fitness vs. the null hypothesis shows that a constant fitness model does not fit the data significantly better than the null hypothesis. Thus, there is evidence for some type of selection, but a constant fitness model is inappropriate. However, as explained in Ford (1971), there is reason to believe there was a drastic change in the fitnesses following the year 1955. Hence, the analysis was reapplied to the 17 gene-

Table 2. *Fitness analysis of a population of Panaxia dominula*[+]

Number of Generations	Genotype	Initial[++] Fitness Estimate	Maximum Like. Estimate	Stand. Dev. of Est.	Corr. of Est.'s	Goodness of fit tests of:				χ^2 Test of	
						null hypothesis	df	constant selection	df	con. selc. vs. null	df
23 (1939—1961)	medionigra	.97	.97	.03	−0.02	128.76*	44	126.14*	42	2.65	2
	bimacula	.81	.80	.16							
17 (1939—1955)	medionigra	.91	.90	.03	−0.02	29.43	32	13.49	30	15.95*	2
	bimacula	.56	.55	.14							

[+] Data from Ford (1971).
[++] All fitnesses are measured relative to the fitness of the *dominula* homozygote.
* Significant at $P < .05$.

Table 3. *Fitness analysis of inversion data on Drosophila ananassae populations* [+]

Initial Freq. A	Repli- cation	Geno- type	Min. χ^{2}[++] Fitness Estimate	Maximum Like. Estimate	Stand. Dev. of Est.	Goodness of fit tests of:				χ^{2} Test of con- stant selc. vs null	df
						null hypothesis	df	constant selection	df		
.1	1	AA	.429	.421	.092	95.96*	6	4.02	4	91.94*	2
		BB	.243	.254	.043						
	2	AA	.263	.237	.092	134.11*	6	9.33	4	124.78*	2
		BB	.172	.160	.033						
	1 & 2	AA	—	.335	.064	230.07*	12	16.72	10	213.35*	2
		BB	—	.205	.027	Homogeneity chi-square of 1 & 2 = 3.36					2
.9	1	AA	.488	.462	.067	71.15*	6	13.72*	4	57.43*	2
		BB	.722	.625	.268						
	2	AA	.400	.403	.064	56.26*	6	7.24	4	49.02*	2
		BB	.315	.377	.181						
	1 & 2	AA	—	.431	.046	127.41*	12	21.56*	10	105.84*	2
		BB	—	.470	.150	Homogeneity chi-square of 1 & 2 = 0.61					2
.1 & .9	all	AA	—	.371	.026	357.47*	24	42.00*	22	315.47*	2
		BB	—	.221	.025	Homogeneity chi-square of all 4 = 7.70					6

[+] Data from Tobari and Kojima (1967).
[++] Calculated by Tobari and Kojima relative to the heterozygote AB.
* Significant at $P < .05$.

rations from 1939 until 1955. The results of this analysis are also given in Table 2 and differ significantly from the previous analysis. The goodness of fit chi-square of both the null hypothesis and constant selection are not significant, but the goodness of fit of constant selection is considerably smaller, indicating an excellent fit of the constant fitness model. The test of constant selection vs. the null hypothesis overwhelmingly rejects the null hypothesis (probability $< .0001$) in favor of constant selection. This analysis therefore confirms Ford's conclusion that "... the *medionigra* gene was consistently subject to a 10% disadvantage compared with *dominula* from 1941 to 1955 ..."

C. Random-mating with all selection after sampling

The final example is also taken from a sexual population with assumed random-mating, but with data gathered with a very different sampling design. The data is given in Tobari and Kojima (1967) on an inversion polymorphism in *Drosophila ananassae*. In their experiment, the populations were kept in cages and quickly reached a very large size. One-hundred and fifty larvae were sampled each generation and scored for their karyotypes. Consequently, the sample flies did not contribute to the next generation. Also, in such experiments, it is customary to assume all selection occurs after sampling. Under such sampling conditions, the unconditional likelihood procedure of DuMouchel and Anderson or the minimum chi-square procedure would be appropriate. However, as argued in section II, the model developed

in this paper is also appropriate if the sample size per generation is large. Under such conditions, this procedure should yield fitness estimates that are statistically equivalent to estimates obtained by the above two procedures. Thus, the model given by (3.8) will be used. The model is applied to four populations given in the appendix of Tobari and Kojima under the designation "2L segregation with 3L = AA." Two replicate populations were started at each of two initial frequencies of "A", .1 and .9. The maximum likelihood estimates, their standard deviations, various tests and the minimum chi-square estimates of Tobari and Kojima are all given in Table 3 for these four populations. As can be seen, in no case did the maximum likelihood estimates differ by more than a half of their standard deviations from the minimum chi-square estimates. Therefore, these two very different methods whose underlying assumptions are quite different yield statistically equivalent estimates, as was predicted.

Statistic (2.6) was used to test for homogeneity with respect to the constant fitness model. Both replicates are homogeneous within an initial gene frequency. Furthermore, there is homogeneity with respect to a constant fitness model for all four populations. Despite the fact that the constant fitness model gave a good fit for three of the four populations, the goodness of fit of a constant fitness model for the data pooled across replicates and initial gene frequencies yields a significant chi-square, indicating that the constant fitness model cannot adequately describe this combined data set.

V. Discussion

The idea of conditioning the likelihood function for any one generation on the previously observed generation is not a new one, although many of the statistical techniques applied to such Markov likelihood functions are of more recent origin. Such a conditional approach was suggested by DuMouchel and Anderson (1968), but was rejected by them because the unconditional method ". . . of a single process originating in the known initial frequencies, consistently yields estimates which fit the observed data best, as tested by a chi-square criterion for goodness of fit." Similar sentiments were expressed in Levene *et al.* (1954) concerning the unconditional minimum chi-square produce which gives ". . . . that set of constant W's that best fits the observations." The flaw of these arguments is the goodness of fit criterion being used. The criterion in the above two papers is the "usual" goodness of fit test in which the

in which the expected values are derived from the fitness estimates and initial conditions only and by statistic 2.4. In every case, the maximum likelihood estimates yield a larger usual goodness of fit chi-square than the minimum chi-square estimates, and the minimum chi-square estimates yield a larger conditional chi-square than the maximum likelihood estimates. Consequently, the minimum chi-square procedure can be rejected in favor of the maximum conditional likelihood procedure since the latter consistently yields estimates which fit the observed data better, as tested by statistic (2.4). This last statement is just as wrong as the previously quoted statements and for the same reason. Both arguments are based on a statistical artifact and demonstrate *only* that the estimation procedure must be consistent with the goodness of fit criterion.

It might be tempting to decide which procedure is "better" for a given data set by comparing the two

Table 4. *A comparison of goodness of fit criteria and estimation procedures on 4 populations of Drosophila ananassae* [+]

Initial Freq. A	Repli- cation	Estimation Procedure	Estimated AA	Fitnesses BB	„Usual" χ^2 [++] based on est. fitnesses	Cond. χ^2 (Eqn. 2.4) based on estimated fitnesses
.1	1	Min. χ^2	.429	.243	4.47	4.17
		Max. Like.	.421	.254	6.02	4.02
	2	Min. χ^2	.263	.172	7.47	9.47
		Max. Like.	.237	.160	7.64	9.33
.9	1	Min. χ^2	.488	.722	14.47*	13.86*
		Max. Like.	.462	.625	17.33*	13.72*
	2	Min. χ^2	.400	.315	3.92	7.48
		Max. Like.	.403	.377	4.10	7.24

[+] Data from Tobari and Kojima (1967).
[++] All tests have 4 degrees of freedom.
* Significant at $P < .05$.

estimated fitnesses are used to predict the expected allele numbers every generation from the initial conditions only. Then, the chi-square is the sum over generations of the sum over alleles of the quantities $(obs.-exp.)^2/exp.$ This goodness of fit criterion is consistent with and rests upon the same assumptions as the unconditional likelihood estimation method of DuMouchel and Anderson and the minimum chi-square estimation procedure of Levene *et al.* However, the goodness of fit criterion appropriate for estimators derived from the conditional likelihood is statistic 2.4. A goodness of fit criterion is appropriate *only* for estimation procedures which share the same underlying assumptions. To do otherwise leads to unjustifiable conclusions. This point is made in Table 4 which gives the fitness estimates for the Tobari and Kojima data under both the conditional likelihood approach and the minimum chi-square approach. Both sets of fitnesses were tested for goodness of fit by the usual goodness of fit criterion

goodness of fit tests evaluated with the estimation procedure consistent with them. However, the two procedures have very different underlying assumptions, and both tests are asymptotic and probably have different rates of convergence. Also, for the special case of random mating with selection occurring after sampling, the conditional likelihood fitness estimates depend both upon the gene and genotype frequencies while the minimum chi-square or unconditional likelihood estimates depend only on gene frequencies. Finally, judging various estimation procedures solely by the results they yield for a particular data set could bias the entire analysis. *The question of which method of analysis is most appropriate should not be answered by the results of that analysis after the data set has been produced, but should be answered before the data is analyzed by inspecting the experimental design and sampling procedures.*

As previously argued, the conditional likelihood approach is most appropriate when the evolving

population itself is followed and the sample includes all or nearly all of this population every generation. In such a situation all of the randomness in the experiment is attributed to the Markov nature of the, evolutionary process; *i. e.*, "genetic drift" in a very broad sense. None is attributed to sampling error. At the other extreme, the evolving population may be extremely large and the samples very small. In such instances, the unconditional approach which assumes all randomness is due to sampling error and which allows no genetic drift is most appropriate. Most situations are intermediate between these two extremes. For example, for the *Panaxia dominula* data, the sample amounted to 10—25% of the evolving population — a substantial proportion but not a complete census. In such instances both the sampling error and genetic drift can contribute significantly to the randomness of the observations. Attributing all of the randomness to the underlying Markov process and none to sampling error exaggerates the importance of correlations between generations and causes the statistical analysis to be somewhat more conservative than it need be for reasons discussed in the introduction. On the other hand, the unconditional approaches which totally ignores the correlation between generations can yield spurious evidence for selection when in fact there is no selection. For these sampling situations it is probably better to adopt the more conservative conditional approach. In deciding whether drift could be a significant factor, it must be noted that drift could accumulate over several generations even if it is negligible over any single generation transition. This accumulation of drift will undermine the assumptions of the unconditional approaches and could therefore be an important factor even for large sized populations if the populations are followed for enough generations.

As the total population size gets larger and the number of generations the population is observed gets smaller, drift becomes less significant and the unconditional models appropriate. However, as long as the sample size is sufficiently large, the conditional likelihood approximates the unconditional likelihood and both types of analyses should yield statistically equivalent results — as they did for Tobari and Kojima's (1967) data on an inversion polymorphism in a population cage based on samples of 150 flies per generation. Thus, when the sample size is large and evolving population size is effectively infinite, either procedure may be used. Even in this situation the conditional approach has some advantages over the unconditional approaches. First as mentioned in IV. A, the conditional approach is more robust in the course of a long experiment to small deviations from the underlying assumptions. Second, the conditional approach in general yields a less complex likelihood function than the unconditional approach because the parameters of the fitness model only enter into single generation transitions. Consequently,

the likelihood function, the scores and second derivatives of the log likelihood can all be written down explicitly. One bonus of this simplicity is that initial estimators of the fitness parameters can often be written down from inspection of the scores. In the worked examples and additional analyses not presented in this paper, these initial estimators were very close to the maximum likelihood estimators. For the unconditional approach the fitness parameters are used recursively to predict all the generation transitions from the initial conditions. Even for very simple models of selection, these relations become very complex after only a few transitions, and the likelihood cannot be written down explicitly. Thus, both the scores and second derivatives must be evaluated through some sort of recursive formula. This means that more effort and time must be employed to evaluate the maximum likelihood estimators for the unconditional case. Thus, even for sampling designs that make both models appropriate, robustness and simplicity favor the conditional likelihood approach over the unconditional likelihood approach.

Literature

Allard, R. W., Kahler, A. L., Weir, B. S.: The effect of selection on esterase allozymes in a barley population. Genetics **72**, 489—503 (1972).

Anderson, W. W.: Selection in experimental populations. I. Lethal genes. Genetics **62**, 653—672 (1969).

Asher, Jr., J. H.: Parthenogenesis and genetic variability. II. One-locus models for various diploid populations. Genetics **66**, 369—391 (1970).

Bhat, B. R.: On the asymptotic distribution of the "psi-squared" goodness of fit criteria for Markov chains and Markov sequences. Ann. Math. Stat. **32**, 49—58 (1961).

Billingsley, P.: Statistical Inference for Markov Processes. Chicago: University of Chicago Press 1961.

Cavalli, L. L.: The analysis of selection curves. Biometrics **6**, 208—220 (1950).

DuMouchel, W. H., Anderson, W. W.: The analysis of selection in experimental populations. Genetics **58**, 435—449 (1968).

Ford, E. B.: Ecological Genetics. London: Chapman and Hall 1971.

Gastwirth, J. L., Rubin, H.: The effect of dependence on the level of some one-sample tests. J. Amer. Stat. Assoc. **66**, 816—820 (1971).

Kendall, M. G., Stuart, A.: The Advanced Theory of Statistics. Vol. II. New York: Hafner 1973.

Levene, H., Pavlovsky, O., Dobzhansky, Th.: Interaction of the adaptive values in polymorphic experimental populations of *Drosophila pseudoobscura*. Evol. **8**, 335—349 (1954).

Lorenz, R. J.: On the determination of relative fitness from frequencies of genotypes in subsequent generations. I. Algebra of reproduction and selection in populations with discrete generations. Theoret. Appl. Genetics **40**, 11—17 (1970a). II. Statistical problems in estimation of fitnesses. Theoret. Appl. Genetics **40**, 76—79 (1097b). III. Simulation experiments. Theoret. Appl. Genetics **40**, 79—87 (1970c).

Prout, T.: The estimation of fitnesses from population data. Genetics **63**, 949—967 (1969).

Stam, P.: Estimators for variability in experimental populations, their variances and covariances. Theoret. Pop. Biol. **2**, 51—59 (1971).

Templeton, A. R., Rothman, E. D.: The population genetics of parthenogenetic strains of *Drosophila mercatorum*. I. One locus model and statistics. Theoret. Appl. Genetics **43**, 204—212 (1973).

Tobari, Y. N., Kojima, K.: Selective modes associated with inversion karyotypes in *Drosophila ananassae*.

I. Frequency-dependent selection. Genetics **57**, 179 to 188 (1967).

Workman, P. L., Jain, S. K.: Zygotic selection under mixed random mating and self-fertilization: Theory and problems of estimation. Genetics **54**, 159—171 (1966).

Wright, S., Dobzhansky, Th.: Genetics of natural populations. XII. Experimental reproduction of some of the changes caused by natural selection in certain populations of *Drosophila pseudoobscura*. Genetics **31**, 125—156 (1946).

Dr. A. R. Templeton
Department of Zoology
University of Texas
Austin, Texas 78712 (USA)