THE ANALYSIS OF SELECTION IN EXPERIMENTAL POPULATIONS

WILLIAM H. DuMOUCHEL AND WYATT W. ANDERSON

Department of Statistics and Department of Biology, Yale University

Received August 3, 1967

THE theory of evolution is the unifying concept of biology. At the heart of the modern theory of evolution is the idea of *natural selection*, of differential reproduction among genotypes. Many factors contribute to the leaving of different numbers of offspring; survival, longevity, and fertility are some of the most important. The biological theory of evolution—the "modern synthesis"—was profoundly influenced by the mathematical formulations of population genetics by FISHER, WRIGHT, and HALDANE. Many laboratory experiments have been devised to test various aspects of the theory. The demonstration that some of the genetic changes in nature could be reproduced in experimental populations opened the way for a many-faceted attack on the genetic processes in evolution. We will consider the problem of measuring selection in experimental populations and show how changes in the frequencies of the alleles at a single locus can be used to estimate the selection acting on each genotype. Our considerations are restricted to sexually reproducing, diploid organisms. The technique was devised for experiments with flies of the genus Drosophila, but it is applicable to many other organisms.

The Models: Consider k alleles $A_1, A_2, \ldots A_k$ at a single autosomal locus. Let the frequency of A_i in the experimental population at the beginning of generation t be $p_i(t)$. The k alleles may be combined in $k(k+1)/2$ different ways to form genotypes, each with two alleles. The adaptive value of a genotype in a given environment is defined as the expected number of offspring produced by this genotype; this definition includes all the factors of selection mentioned previously. The terms "fitness" and "selective value" are synonyms for adaptive value. Let the adaptive value of genotype A_iA_j be w_{ij}. For convenience in formulating the model, let us distinguish between genotypes A_iA_j and A_jA_i, but let $w_{ij} \equiv w_{ji}$. We shall consider only constant adaptive values. Since we have worked with data involving four or five alleles, we necessarily adopted this simplest model of constant adaptive values in order to keep the number of parameters estimated to a manageable level.

Under random mating in a large population, the frequency of genotype A_iA_j at the beginning of the t^{th} generation is $p_i(t)\ p_j(t)$. (This paper assumes throughout that the size of the population is so large that differences between actual population frequencies and expected population frequencies can be safely neglected, in contrast to frequencies in samples taken from that population.) Under the evolutionary model, the relative frequency of the offspring of genotype A_iA_j among all those produced at generation $t+1$ is given by $p_i(t)\ p_j(t)\ w_{ij}/\sum_{ij} p_i(t)$.

$p_j(t)\ w_{ij}$. Half of these offspring will have received allele i and half allele j from the A_iA_j parent under consideration. Then $p_i(t+1)$, the frequency of A_i at the beginning of generation number $t+1$, is determined by the contributions of each genotype of the previous generation:

$$p_i(t+1) = \text{constant} \cdot \Sigma_j[(1/2)p_i(t)p_j(t)w_{ij} + (1/2)p_j(t)p_i(t)w_{ji}]$$

Remembering that by definition $w_{ij} \equiv w_{ji}$,

$$p_i(t+1) = \text{constant} \cdot p_i(t)\ \Sigma_j p_j(t) w_{ij}\ .$$

The constant is chosen so that $p_1(t+1) + \ldots + p_k(t+1) = 1$, giving for every choice of i

$$(1) \qquad p_i(t+1) = p_i(t)\ \Sigma_j p_j(t) w_{ij} \,/\, \Sigma_{l,m} p_l(t) p_m(t) w_{lm}.$$

This recurrence relation is the basic formula of the genetic models of selection. The formula involves discrete generations; that is, it is constructed with specific, nonoverlapping periods for reproduction. The organisms are assumed to participate in only one reproductive period, and it is assumed to be the first one after their birth. This model is realistic for organisms with short adult lifespans and specific breeding times. The experiments for which our analysis was devised (see accompanying paper) were designed with generations strictly discrete, so that the above assumptions hold.

Genetic models for populations with continuous breeding and overlapping generations have been formulated (see KIMURA 1958 and HASOFER 1966), but they are generally only approximations which depend upon a number of assumptions. One of the most important of these assumptions is that selection be rather weak. The frequencies of the alleles can then be represented as continuous functions of time. HALDANE (1927) considered the case of slow selection against a completely recessive allele and showed that the changes in the frequencies of the alleles were nearly the same under a model of overlapping generations and continuous change as with the model of discrete generations. The relationship between continuous and discrete models has not been investigated for the case of strong selection. The continuous model itself has not been formulated for rapid changes in allele frequencies under strong selection. The age structure of populations, and the age-specific reproductive capacities of the genotypes must also be included in a general genetic model of continuous change, although these factors are usually ignored because they are so difficult to handle analytically. Even where generations do overlap and breeding is continuous, we shall use the model of discrete generations, remembering that it is only an approximation to the real situation.

The basic recurrence relation (1) can be applied in at least two ways in attempting an experimental analysis of selection. We can assume, for a fixed set of adaptive values $\{w_{ij}\}$, a single deterministic process and, beginning with the known initial frequencies, apply the formula repeatedly to find the expected allele frequencies at any generation. The expected frequencies at any generation t depend only on the initial frequencies of the alleles, the adaptive values, and the generations elapsed since the population was initiated.

Another method of analysis can also be used. We calculate the expected frequencies of the alleles at any particular generation by inserting the *sampling*

frequencies from the previous sample into relation (1) and evaluating it for the number of generations since the last sample. Thus, to calculate the expected allele frequencies at generation t_x, we insert the sampling frequencies observed in the previous sample, at generation t_y, into the right side of relation (1) and evaluate the recurrence relation for t_x-t_y generations. In the first method described above, the expected allele frequencies at generation t_x would be computed by inserting the known *initial* frequencies into relation (1) and repeating the recursion t_x times. It was originally hoped that the second method would be more robust in the course of a long experiment, in the sense of being less easily influenced by shifts in the allele frequencies caused by insufficient experimental control or to some other factor unaccounted for by the theory. This second means of analysis is incorrect in that it assumes, in effect, that the frequencies we estimate from a small sample of the population are the true values. The expected frequencies in the next generation, which are predicted from these estimates, will be influenced by the sampling errors. We have generated estimates of the adaptive values using both analyses. The first method, that 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. LEVENE (in LEVENE, PAVLOVSKY, and DOBZHANSKY 1954) previously commented on the superiority of the first analysis. Throughout the rest of the paper, we will consider the first analysis only.

In many experiments on selection among several alleles in Drosophila, large populations are maintained in population boxes. These boxes contain food cups which are regularly rotated; each cup is kept in the population for a full generation so that the eggs deposited in it have an opportunity to develop to adults. DOBZHANSKY and his associates obtain egg samples and rear the larvae which hatch from them under nearly optimal conditions. Thus the frequencies of the alleles are determined after random mating but *before* selection in the new generation. It is important to note that our formula (1) is designed to fit the details of this type of experiment, in which allele frequencies are determined before selection. Formula (1) would *not* describe the changes in frequencies of the alleles if the samples were taken among the adults *after* part of the total selection had taken place.

Estimation by Maximum Likelihood: Several techniques for measuring selection have been proposed in the past. All were developed for the system of inversions on the third chromosome of *Drosophila pseudoobscura*. These inversions act as supergenes, locking together blocks of genes. Mathematically they behave as different alleles at a single locus. WRIGHT (in WRIGHT and DOBZHANSKY 1946) developed a least squares technique. Actually, he developed two techniques. For the case of two alleles, he minimized the sum of the squares of the difference between the observed and expected *changes* in allele frequency. For the case of three alleles, he minimized the sum of the squares of the difference between the observed and expected allele frequencies. The two approaches give very similar, although not identical, results. CAVALLI (1950) considered the case of heterotic selection (heterozygote advantage) for two alleles; he utilized a model of con-

tinuous change and found the maximum likelihood estimates of the selection coefficients by FISHER's method of scoring. CAVALLI's technique can readily be extended to any two allele case, but does not generalize to greater numbers of alleles. In his continuous model the allele frequencies are implicitly defined and must be found by some iterative procedure. LEVENE (in DOBZHANSKY and LEVENE 1951) designed a graphic technique for the case of two alleles. Later, LEVENE (in LEVENE, PAVLOVSKY, and DOBZHANSKY 1954) estimated adaptive values by the method of minimum chi-square. His procedure was based on trial and error fitting of the allele frequencies expected with various amounts of selection to the observed frequencies. We have developed a general technique for obtaining the maximum likelihood estimates of adaptive values and their covariance matrices.

There is a well-known statistical theory of the precision of maximum likelihood estimation. This theory says that in large samples no other method of estimation can be materially more precise. Moreover, the theory provides a good approximation for the variances and the covariances of the maximum likelihood estimates, namely the inverse of the FISHER information matrix.

The Likelihood Function: The likelihood is a complicated function of the w_{ij}. Let k be the number of alleles under consideration. At generation t the probability of sampling exactly $x_i(t)$ alleles of type i (i=1, . . . ,k) if the total number of genes sampled is $\Sigma_i x_i(t)$, is proportional to the product

$$f(t) = \prod_i p_i(t)^{x_i(t)} \tag{2}$$

where the allele frequencies at time t, the $p_i(t)$, are the functions of the adaptive values and the initial allele frequencies given by formula (1). The expected allele frequencies at any generation may be calculated with relation (1). The likelihood function for an entire experiment is the product of the likelihoods at each generation. That is, it is the product of the expressions (2), for $t = t_1, \ldots, t_n$, where the t_s are the numbers of the generations at which samples were taken and n is the total number of different generations that samples were taken. The basic formula (1) assumes that the t_s are integers and, thus, that samples are taken only at intervals of one or more generations. For continuously breeding populations, samples are often taken at irregular intervals convenient to the experimenter. We have found the expected $p_i(t)$ by simple linear interpolation when t is not an integer. Although this linear approximation is not strictly correct, it has given satisfactory results in application.

Taking the logarithm of the likelihood function, we have

$$L(\text{data}|\text{selective values}) = \log_e f = \text{constant} + \sum_s \sum_i x_i(t_s) \log_e p_i(t_s); \tag{3}$$

we want to find that set of adaptive values which maximizes L. It should be pointed out here that any set of adaptive values is only unique up to a multiplicative constant. If $W = \{w_{ij} \mid i,j=1, \ldots ,k\}$ is one set of adaptive values, the set $\alpha W = \{\alpha w_{ij} \mid i,j=1, \ldots ,k\}$ will produce exactly the same changes in allele frequency for any $\alpha > 0$. This equivalence follows from our basic formula (1), which we see contains a w_{ij} in each term of the numerator and denominator. The absolute sizes of the adaptive values determine the population size, that is, the

number of individuals. The relative sizes of the adaptive values determine the allele frequencies, which are all our samples permit us to estimate. The common practice in the past has been to set one adaptive value arbitrarily equal to unity and define the others relative to it. This scaling disrupts the symmetry of the formulae, particularly when a standard error for every estimate is desired. We have therefore estimated all $k(k+1)/2$ different adaptive values under the constraint that they should sum to the constant $k(k+1)/2$, so that, if there were no selection, every adaptive value would equal unity.

In order to maximize the likelihood we need the first partial derivatives of the logarithm of the likelihood, L, with respect to the w_{ij}:

$$\frac{\partial L}{\partial w_{lm}} = \sum_s \sum_i \frac{x_i(t_s)}{p_i(t_s)} \cdot \frac{\partial p_i(t_s)}{\partial w_{lm}} . \tag{4}$$

The Recursive Definition of the Derivatives of Allele Frequencies: The $\frac{\partial p_i(t_s)}{\partial w_{lm}}$ were evaluated in the same manner as were the $p_i(t_s)$, through the recurrence relation (1). Let us denote the numerator of (1) by $F_i(t)$:

$F_i(t) = \Sigma_j p_i(t) p_j(t) w_{ij}$.

Similarly, let the denominator of (1) be $D(t)$:

$D(t) = \Sigma_h F_h(t)$.

Introducing another abbreviation, let

$$p'_{ilm}(t) = \frac{\partial p_i(t)}{\partial w_{lm}} . \tag{5}$$

Then we have, by differentiation,

$$\frac{\partial F_i(t)}{\partial w_{lm}} = \Sigma_j [p_i(t) p_j(t) \frac{\partial w_{ij}}{\partial w_{lm}} + w_{ij}(p_i(t) p'_{jlm}(t) + p_j(t) p'_{ilm}(t))]$$

$$\text{and } \frac{\partial D(t)}{\partial w_{lm}} = \Sigma_h \partial F_h(t)/\partial w_{lm} .$$

Finally, since $p_i(t+1) = F_i(t)/D(t)$,

$$p'_{ilm}(t+1) = \frac{\partial p_i(t+1)}{\partial w_{lm}} = \frac{\partial F_i(t)}{\partial w_{lm}} \cdot \frac{1}{D(t)} - \frac{F_i(t)}{[D(t)]^2} \cdot \frac{\partial D(t)}{\partial w_{lm}} . \tag{6}$$

The derivatives (5) are zero at generation zero, since the initial frequencies are known constants. Thus we may evaluate, for any given set of adaptive values, the $p'_{ilm}(t_s)$ for any generation number t_s by recurrence relation (6). Hence $\frac{\partial L}{\partial w_{lm}}$ can be determined. In continuously breeding populations where samples are taken at irregular intervals and the t_s are not integers, we use linear interpolation to find the $p'_{ilm}(t_s)$.

The Iterative Solution of the Likelihood Equations: In order to maximize L, we try to find a vector of adaptive values, $\hat{W}$, such that $\partial L(\text{data}|\hat{W})/\partial w_{lm} = 0$ for all (l,m). (Actually this derivative is defined for only $N=k(k+1)/2-1$ independent pairs (l,m); w_{kk} is defined implicitly by the equation $\sum_{i=1}^{k} \sum_{j=i}^{k} w_{ij} = k(k+1)/2$, and derivatives with respect to w_{kk} are not taken.) An explicit solution of the likelihood equations is not possible, so we have resorted to two iterative techniques: FISHER's method of maximum likelihood scoring and the method of

steepest ascent. Both procedures were accomplished with the aid of the high speed digital computer facilities at Yale University. See RAO (1952) and BAILEY (1961) for a more detailed description of maximum likelihood scoring as it was used here. The technique is essentially as follows. A trial vector of w's, $W^{(1)}$, is chosen. We used either the vector whose components are all unity or else the vector which resulted from the technique of steepest ascent which is described below. The values of $\frac{\partial L}{\partial w_{lm}}$ and the information matrix I are then computed at $W^{(1)}$. Let S be the vector whose elements are the $\partial L/\partial w_{lm}$. For notational convenience, let us renumber the adaptive values so that each is indexed by a single subscript. If L is the logarithm of the likelihood function, then the elements of the FISHER information matrix are defined by

$$I_{lm} = E(-\partial^2 L/\partial w_l \partial w_m);\ l,m=1,2,\ldots,N. \tag{7}$$

It can be shown (see the above references) that the I_{lm} can be represented as simple functions of the expected allele frequencies $p_i(t_s)$ and of their derivatives $\partial p_i(t_s)/\partial w_l$ $(i=1,\ldots,k;s=1,\ldots,n;l=1,\ldots,N)$. These are given by

$$I_{lm} = \Sigma_s n(s) \Sigma_i (\partial p_i(t_s)/\partial w_l)(\partial p_i(t_s)/\partial w_m)/p_i(t_s), \tag{7'}$$

where $n(s)$ is the number of genes sampled at time t_s. Then the improved vector of estimates $W^{(2)}$ is given by

$$W^{(2)} = W^{(1)} + I^{-1}S.$$

We then re-evaluate S and I at $W^{(2)}$ and repeat the process above to give $W^{(3)}$. The iteration is continued until the elements of S are close to zero. The last set of estimates is used to evaluate the information matrix and to obtain the approximate covariance matrix I^{-1}. The expected values of the allele frequencies are generated from the known initial frequencies of the alleles and the final set of estimated adaptive values.

For the populations with two or three alleles, this method of scoring was very efficient in locating the maximum of L. For $k=4$ or 5, however, the successive values of $W^{(n)}$ often did not converge. The method of steepest ascent was used in these cases. Beginning at a trial value $W^{(1)}$, the vector of derivatives, S, is formed. Then the next estimate is given by

$$W^{(2)} = W^{(1)} + \text{constant} \cdot S, \text{ constant} > 0.$$

The constant at each stage is determined by various inner criteria, depending on whether or not a maximum point of L is being approached. This method of steepest ascent is patterned after that used by KRUSKAL (1964); his article details ways of choosing the constant at each step. For a more general review of steepest ascent see SPANG (1962). The approach by steepest ascent was usually continued until the Euclidean length of the vector S was reduced to about one percent of the value it had in the first few steps. Then the scoring technique was attempted. If the scoring method still did not lead to convergence, the process was halted and the last set of estimates from steepest ascent was used to obtain the information matrix, as well as the expected values of the allele frequencies.

Both of these maximization techniques are liable to converge to a merely local maximum of the likelihood function, since we were not able to prove rigorously that the likelihood must have a unique maximum. A more thorough approach

would be to start with several widely differing values of $W^{(1)}$ and see that the same final estimates are reached. We have not done so, but we expect that the large sample sizes involved would produce a relatively smooth likelihood function with a single maximum to which our estimates will converge. Three hundred genes were typically sampled from each of the populations for which we devised this procedure, at each of seven to twelve different generations. Convergence to some other critical point, such as a minimum or a saddle point, is not a danger, since at each step the value of the log likelihood is printed out to ensure that the desired maximization is actually taking place.

It is more probable that there should be no clearly defined maximum than that the maximum be local only. Where the scoring technique is successful, its very success indicates that the maximum is quite well defined. The scoring technique was not usually effective, however, for large numbers of alleles. Each vector of adaptive values defines a point in an N-dimensional simplex, where N is the number of independent adaptive values being estimated. The likelihood is a function of these points, and it may attain a maximum or near-maximum along a line or a plane or even a higher dimensional subspace in the N-dimensional space. All that can be said in this case is that the data point equally toward any of the points, or vectors of adaptive values, on the subspace. Of course, vectors which contain negative adaptive values are biologically meaningless and are excluded from the outset—they correspond to points in the space not on the simplex. This consideration can reduce the uncertainty in the choice of the maximum likelihood estimates. This topic is developed further below (p. 443).

The fit of the expected allele frequencies generated from the maximum likelihood estimates of the adaptive values to the allele frequencies actually observed can be appraised by a chi-square statistic for goodness of fit. There are $(k-1)n - k(k+1)/2 + 1$ degrees of freedom for a population with k alleles and samples at n times after the initiation of the population.

We have checked our computer program in the following way. Several sets of artificial data were produced using formula (1) with known adaptive values. These data were then used as input to the estimation program. The original adaptive values were returned as output to an accuracy of at least one percent in all cases.

TABLE 1

Analysis of selection in two-allele population number 204 of PAVLOVSKY *and* DOBZHANSKY *(1966)*

Genotype	Estimated adaptive values	Standard deviations	Correlation matrix AR/AR	AR/PP	PP/PP
AR/AR	1.203	.016	1.000	.909	—.963
AR/PP	1.156	.027	.909	1.000	—.988
PP/PP	.641	.042	—.963	—.988	1.000

A chi-square statistic for goodness of fit of the observed gene frequencies was computed under the hypothesis that the estimated adaptive values were the true fitnesses. Estimation of two independent fitnesses from frequencies observed at 12 generations left 10 degrees of freedom. The value of the chi-square was 4.9.

TABLE 2

Analysis of selection in three-allele population number 215 of PAVLOVSKY *and* DOBZHANSKY *(1966)*

Genotype	Estimated adaptive values	Standard deviations	Matrix of correlations ST/ST	ST/AR	ST/PP	AR/AR	AR/PP	PP/PP
ST/ST	.506	1.907	1.000	—1.000	—.924	.999	—.986	—.941
ST/AR	2.166	3.041	—1.000	1.000	.918	—1.000	.989	.938
ST/PP	1.115	.234	—.924	.918	1.000	—.912	.855	.861
AR/AR	.383	2.231	.999	—1.000	—.912	1.000	—.990	—.938
AR/PP	1.289	.705	—.986	.989	.855	—.990	1.000	.904
PP/PP	.542	.196	—.941	.938	.861	—.938	.904	1.000

A chi-square statistic for goodness of fit of the observed gene frequencies was computed under the hypothesis that the estimated selective values were the true fitnesses. Estimation of five independent fitnesses from observed frequencies at 10 generations, each generation contributing two degrees of freedom, left 15 degrees of freedom for the chi-square. The contributions to the χ^2 statistic from computations involving each of the three allele frequencies are given, along with the total χ^2.

Contribution of ST	13.0
Contribution of AR	14.3
Contribution of PP	9.7
TOTAL	37.0

TABLE 3

Analysis of selection in four-allele population control II of ANDERSON *et al. (1968)*

Genotype	Estimated fitness	Standard deviation	Correlation matrix ST/ST	ST/AR	ST/CH	ST/PP	AR/AR	AR/CH	AR/PP	CH/CH	CH/PP	PP/PP
ST/ST	1.56	.98	1.00	.99	.94	.97	.76	.90	.28	—.98	.99	—1.00
ST/AR	1.38	.90	.99	1.00	.92	.97	.70	.92	.29	—.97	.98	—.99
ST/CH	1.06	.54	.94	.92	1.00	.88	.82	.73	.45	—.88	.93	—.96
ST/PP	1.12	.93	.97	.97	.88	1.00	.77	.94	.08	—.99	.98	—.96
AR/AR	.87	.60	.76	.70	.82	.77	1.00	.53	.04	—.74	.81	—.79
AR/CH	1.05	1.38	.90	.92	.73	.94	.53	1.00	.00	—.95	.90	—.86
AR/PP	1.49	.81	.28	.29	.45	.08	.04	.00	1.00	—.09	.18	—.33
CH/CH	1.09	3.81	—.98	—.97	—.88	—.99	—.74	—.95	—.09	1.00	—.99	.96
CH/PP	.17	5.85	.99	.98	.93	.98	.81	.90	.18	—.99	1.00	—.99
PP/PP	.20	7.32	—1.00	—.99	—.96	—.96	—.79	—.86	—.33	.96	—.99	1.00

In the chi-square statistic for goodness of fit, estimation of nine independent fitnesses from frequencies observed at 8 generations, each generation contributing three degrees of freedom, left 15 degrees of freedom for the chi-square.

Contribution of ST	12.7
Contribution of AR	27.1
Contribution of CH	7.1
Contribution of PP	6.4
Total	53.4

Examples: Tables 1, 2, and 3 show some results of analyses of data from the literature. It is clear that the sizes of the standard deviations of the adaptive values increase dramatically with an increase in the number of alleles. This increase is not altogether surprising, since the number of adaptive values being estimated increases in proportion to the square of the number of alleles. Of particular interest are the very high correlations between the estimates. These correlations are typically extreme, close to plus or minus one, in all of the thirty populations we have investigated. It is these large correlations which make more information available about the adaptive values as a whole in a population than is indicated by the large standard deviations. We will elaborate on this point later.

Data from the two allele populations we have analyzed usually fit the model satisfactorily, as evidenced by the chi-square tests of goodness of fit between observed and expected allele frequencies. The populations with three or more alleles usually do not yield estimates which account satisfactorily for the data. Interactions between genotypes and frequency-dependent selection are probably factors in the departure of the data from our model of change governed by constant adaptive values. Unrecorded changes in the environment during the course of the usually long experiments may alter adaptive values and produce irregularities in the data for which no simple model can account. More complex models would be difficult to formulate analytically, but once formulated, they could be analyzed by adaptation of the techniques we have used. The model we have assumed promises to give a fair first approximation to the true situation, and it may be used to see how we should proceed to set up more intricate models. The size of the contribution of each allele to the goodness-of-fit chi-squares indicates roughly which alleles are influenced most by interactions, frequency-dependent selection, and the other factors which depart from our simple model.

The Problem of Indeterminancy: In the case of 4 or 5 alleles, our program fairly often indicated that the function $L(\text{data}|W)$, given by (3), was maximized at a point where one or more of the adaptive values was negative. Such an estimate is, of course, meaningless both biologically and statistically. These negative estimates could be accounted for by sampling error or they may point to a possible defect in the mathematical model which supposedly governs the experiment, as already noted. We should ask: For what non-negative values of the w_l is the likelihood function maximized? Properly, this involves a nonlinear programming approach to the problem of maximizing L, and we have made no attempt to do this. But we have been able to use the "formal" (negative) estimates and their information matrices to obtain other, non-negative, estimates of the adaptive values for which the value of the likelihood is virtually identical. This fact is the basis for the earlier statement that the likelihood forms a "ridge" or a "plateau" in the N-dimensional space. Also, the large standard deviations coupled with the extreme correlations in the FISHER covariance matrices show that the experiments yield very little information concerning certain linear combinations of the fitnesses. This corresponds to the uncertainty as to where, along the ridge, the point representing the true adaptive values lies.

This uncertainty is a disappointing outgrowth of our analyses. It indicates that,

in many cases at least, a single experiment can never enable one to estimate every adaptive value accurately. Just why this should be so is not clear, although certain special cases seem explicable. These cases rest on the fact that once an equilibrium in allele frequencies is reached, estimation of the adaptive values is no longer possible. There are only $k-1$ degrees of freedom involved in observing the allele frequencies at any one generation. In order to estimate the N independent fitnesses, several generations must be observed. If an equilibrium in the allele frequencies obtains, then the successive generations contribute no new information for purposes of estimation. Thus, for example, if $k=5$ and an equilibrium were established before the fourth set of observations, then at most $3 \cdot (k-1) = 12$ degrees of freedom would be available for estimation, insufficient to estimate the $N=14$ parameters involved.

The idea of a partial equilibrium of allele frequencies is important also. If alleles A_i and A_j are such that $p_i(t) = \alpha p_j(t)$, $\alpha > 0$, α independent of t from some time on, then let us say that alleles A_i and A_j have reached an equilibrium with respect to each other. Consider equation (1), assuming that such a partial equilibrium holds between A_1 and A_2, say. In this case, the fitnesses w_{11}, w_{12}, w_{22} occur in (1) only in the combinations

$$S_1(t) = p_1(t)\, p_2(t)\, w_{12} + p_1(t)^2\, w_{11}$$
$$S_2(t) = p_1(t)\, p_2(t)\, w_{12} + p_2^2(t)\, w_{22}\ .$$

We assume that for every t, $p_2(t) = \alpha p_1(t)$ so that

$$S_1(t) = p_1^2(t)\ (\alpha w_{12} + w_{11})$$
$$S_2(t) = p_1^2(t)\ (\alpha w_{12} + \alpha^2 w_{22}).$$

This shows that the three parameters w_{12}, w_{11} and w_{22} only enter the likelihood equations in the two expressions $\alpha w_{12} + w_{11} = z_1$, say, and $\alpha w_{12} + \alpha^2 w_{22} = z_2$, and so any triple (w_{11}, w_{12}, w_{22}) which preserves z_1 and z_2 will not change the likelihood equation. Thus in the case stated, with $p_2(t) = \alpha p_1(t)$, w_{11}, w_{12}, and w_{22} are confounded with each other, and only z_1 and z_2 can be estimated. This would produce a "ridge" on the likelihood surface in the N-dimensional space.

There are even further considerations. Suppose that $w_{1j} = w_{2j}$, $j = 1, 2, \ldots, k$. In this case, no matter what the initial allele frequencies $p_1(0), \ldots, p_k(0)$ were, it can be seen from (1) that the frequencies $p_1(t)$, $p_2(t)$ will remain in the ratio $\alpha_0 = p_2(0)/p_1(0)$. Thus if A_1 and A_2 are observed to be in a partial equilibrium with each other *including the zero*$^{\text{th}}$ *generation*, then one of two cases exist: Either $w_{1j} = w_{2j}$, $j = 1, \ldots, k$, or the initial value of α, $\alpha_0 = p_2(0)/p_1(0)$ just happened to be that at which A_1 and A_2 are in equilibrium with one another. The first case, that alleles A_1 and A_2 act identically for purposes of selection, is *a priori* much more probable, assuming that $p_1(0)$ and $p_2(0)$ were chosen by the experimenter without regard to any considerations bearing on such a partial equilibrium. Therefore, in such circumstances, there is a strong case for estimating the w_{ij} under the restriction that $w_{1j} = w_{2j}$ for every j or, equivalently, treating A_1 and A_2 as different manifestations of the same allele, and, by adding the observed frequencies of A_1 and A_2, reduce the problem of estimation to that of $k-1$ alleles rather than k. This assumption will usually greatly reduce the vari-

TABLE 4

Analysis of selection when a partial equilibrium has occurred. Five-allele population DDT-2 of ANDERSON *et al. (1968)*

Genotype	Estimate (1) Fitness	St. dev.	Estimate (2) Fitness	St. dev.	Estimate (3) Fitness	St. dev.
ST/ST	.97	13.0	1.36	1.60	1.477	.267
ST/AR	2.06	29.2	1.72	2.03	1.522	.410
ST/CH	1.71	12.1	1.34	1.49	1.404	.111
ST/PP	1.57	32.3	1.45	2.43	1.404	.111
ST/TL	1.27	16.2	1.34	1.49	1.404	.111
AR/AR	1.27	17.7	1.59	1.76	1.751	.273
AR/CH	.79	17.1	.99	1.37	1.079	.055
AR/PP	1.08	10.2	1.22	2.27	1.079	.055
AR/TL	1.15	17.0	.99	1.37	1.079	.055
CH/CH	.74	219.	.49	5.11	.476	.058
CH/PP	.20	196.	.73	10.4	.476	.058
CH/TL	.42	35.4	.49	5.11	.476	.058
PP/PP	.66	155.	.00	16.1	.476	.058
PP/TL	.36	37.4	.73	10.4	.476	.058
TL/TL	.75	20.4	.49	5.11	.476	.058
χ^2	30.1		31.2		34.0	
d.f.*	22		27		31	

(1) Maximum likelihood estimates with no special assumptions.

(2) Estimates made under the assumption that alleles CH and TL act alike.

(3) Estimates made under assumption that alleles CH, PP and TL act alike.

* Frequencies were observed at 9 generations, each generation contributing four degrees of freedom.

ances of the estimated adaptive values, since the parameter space is reduced by k dimensions. Table 4 presents an example of the use of this method.

In more general cases, where such partial equilibria have not been observed, the phenomenon of a "ridge" on the likelihood surface has still been observed. In these cases, if it is desired to use some point on the ridge as an estimate of the vector of adaptive values, it is, of course, necessary to use a point which is also on the simplex, i.e., where every adaptive value is nonnegative. Among such points, unless there are *a priori* considerations, there is very little reason to choose one point on the "ridge" over another. The situation would be quite different if two independent experiments were made, under identical conditions except for different initial allele frequencies. Then, although each experiment would have associated with it a "ridge" of ambiguity, the combined experiments would point strongly toward the adaptive values represented by the intersection of the "ridges." But in single experiments, with large numbers of alleles, it must be kept in mind that any single vector of estimates for the fitnesses cannot adequately describe the results of the experiment without all of the associated covariances.

Combining and comparing vectors of adaptive values from different popula-

tions: When an experiment is replicated and more than one vector of estimates is obtained for the same adaptive values, a combined or averaged set of estimates will usually be desired. The average is formed by weighting each vector of estimates by its information matrix. If independent experiments yield the vectors $\hat{W}^{(1)}, \hat{W}^{(2)}, \ldots$ as estimates of the adaptive values, and if the information matrices are $I^{(1)}, I^{(2)}, \ldots$ respectively, then the best average is

$$W^* = (I^{(1)} + I^{(2)} + \ldots)^{-1} (I^{(1)}\hat{W}^{(1)} + I^{(2)}\hat{W}^{(2)} + \ldots). \tag{8}$$

To prove this, assume that the likelihood function of the n^{th} experiment is well approximated by the multivariate normal distribution with mean vector $\hat{W}^{(n)}$ and covariance matrix $[I^{(n)}]^{-1}$. Then the n^{th} log likelihood is, up to an additive constant,

$$L^{(n)} = -\tfrac{1}{2}(W-\hat{W}^{(n)})' I^{(n)} (W-\hat{W}^{(n)}),$$

where the prime denotes transpose. (See, for example, RAO (1952) for more information about the multivariate normal distribution.) Since the experiments were assumed to be independent, L^*, the logarithm of the likelihood of the combined experiments, is the sum of the $L^{(n)}$.

$$L^* = -\tfrac{1}{2} \Sigma_n (W-\hat{W}^{(n)})' I^{(n)} (W-\hat{W}^{(n)}).$$

The maximum likelihood estimate for the combined experiments, W^*, is found by equating the vector derivative of L^* with respect to W to the zero vector.

$$-dL/dW = \Sigma_n I^{(n)} (W^*-\hat{W}^{(n)}) = 0.$$

From this it follows that

$$\Sigma_n I^{(n)} W^* = \Sigma_n I^{(n)} \hat{W}^{(n)},$$
$$W^* = [\Sigma_n I^{(n)}]^{-1} \Sigma_n I^{(n)} \hat{W}^{(n)}.$$

The last line is identical with equation (8). Also, since $I^{(n)}$ is defined as $E[-d^2 L^{(n)}/dW^2]$, then $I^* = E(-d^2L^*/dW^2) = \Sigma_n I^{(n)}$, since both expectation and differentiation are additive. The covariance matrix for the averaged set of estimates W^* is thus $C^* = I^{*-1} = (I^{(1)} + I^{(2)} + \ldots)^{-1}$.

Using the assumption of normality, useful statistics for comparing adaptive values from the same and different populations can be derived from $\hat{W}$ and I. For example, in the population described in Table 1, the estimates of the fitnesses of genotypes labeled *AR/AR* and *AR/PP* are $\hat{w}_{11} = 1.203$ and $\hat{w}_{12} = 1.156$, respectively. If the *PP* gene were completely recessive, then w_{11} would equal w_{12}. We can base a statistic on the hypothesis that this is so. From Table 1, the variance of $\hat{w}_{11}$, $V(\hat{w}_{11}) = .016^2 = .000256$, $V(\hat{w}_{12}) = .027^2 = .000729$, and the covariance of $\hat{w}_{11}$ and $\hat{w}_{12}$ is $Cov(\hat{w}_{11}, \hat{w}_{12}) = (.909)(.027)(.016) = .000393$. Then the standard deviation of $\hat{w}_{11} - \hat{w}_{12}$ is given by

$$\begin{aligned}\sigma(\hat{w}_{11}-\hat{w}_{12}) &= [V(\hat{w}_{11}) + V(\hat{w}_{12}) - 2\,\mathrm{Cov}(\hat{w}_{11}, \hat{w}_{12})]^{1/2} \\ &= [.000256 + .000729 - .000786]^{1/2} \\ &= .014.\end{aligned}$$

The observed value of $\hat{w}_{11}-\hat{w}_{12}$ is .047, a figure which is $.047/.014 = 3.3$ standard deviations away from zero. This is strong evidence that w_{11} is actually larger than w_{12} so that the PP allele is not wholly recessive.

In many cases, one must decide whether selection as a whole has operated differently in differently treated populations. Each vector of adaptive values defines a point on an N-dimensional simplex, where N is the number of independent

estimates of the adaptive values. This point is an overall representation of selection in the population. To see whether selection has been different in different populations means we must test the differences between the two vectors of adaptive values. The effect of the high correlations among the estimates is to concentrate the "confidence region" around the point in the vicinity of a straight line or some higher dimensional subspace. The large standard deviations mean that the location of the point along this line cannot be stated very precisely. But the overall performance of the population is quite likely represented by a point somewhere *close* to the straight line. Thus, differences in the total selection between populations can be much easier to demonstrate than differences between individual adaptive values.

As an example, consider the data in Table 3 for a population with four alleles. Comparing the standard deviations with the estimates themselves, there is little evidence that the vector of estimates is significantly different from the vector whose elements are all unity. This vector of ones represents a total lack of selection. Assume that the estimated adaptive values are normally distributed with covariance matrix I^{-1}. Then we may construct a test of the hypothesis that the actual adaptive values are all equal to unity. Denote by 1 the nine-dimensional vector, all of whose elements are unity, and by $\hat{W}$ the vector of the first nine adaptive values in Table 3. Denote by I the FISHER information matrix of the adaptive values. The standard deviations and correlations of Table 3 were obtained by inverting I and using the relation $\Sigma_{i \geq j} w_{ij} = 10$ to compute the last row and column of the covariance matrix. Under the hypothesis, the vector $\hat{W}$ is normally distributed with mean 1 and covariance I^{-1}, so the product

$$(\hat{W}-1)' I (\hat{W}-1) = \chi^2$$

will be distributed as a chi-square variable with nine degrees of freedom. In our example the actual value of the chi square is 516, making it unthinkable that no selection occurred. Of course, a glance at the observed gene frequencies for this population, given in the accompanying paper, also makes it obvious that some selection has occurred.

A similar test can be made of the hypothesis that vectors of adaptive values from different populations are equal. Let $\hat{W}^{(1)}$, $\hat{W}^{(2)}$, $I^{(1)}$, $I^{(2)}$ be the vectors of estimates and the information matrices for the two populations. Then we want to test the hypothesis that the vector $\delta = \hat{W}^{(1)} - \hat{W}^{(2)}$ has as its mean the vector whose elements are all zeros. The covariance matrix of δ will be given by*

$$H^{-1} = I^{(1)-1} + I^{(2)-1}$$

and δ will be normally distributed on the assumption that $\hat{W}^{(1)}$ and $\hat{W}^{(2)}$ are. Therefore the product

$$\delta' H \delta = \chi^2$$

will have a chi-square distribution with number of degrees of freedom equal to the number of independent adaptive values being estimated $(k(k+1)/2-1)$. This technique was used to test for the effect of two insecticides on selection in polymorphic populations of *Drosophila pseudoobscura* in the accompanying paper.

* One must be very careful when attempting to invert such ill-conditioned matrices as seem to arise in this problem. All matrix inversions were done in double precision arithmetic, which carries 16 significant figures.

Advice to the Experimenter: This study has implications for the design of experiments on selection. The first is that unrealistically large sample sizes are needed to measure adaptive values accurately when as many as four or five alleles are present. It might be better to conduct several experiments with different combinations of the alleles taken two or three at a time. Little or no information about selection can be extracted once an equilibrium of the allele frequencies is reached. It is better to run two shorter experiments than to continue one long experiment after an equilibrium has been reached. If it seems desirable to continue the experiment to ensure that an equilibrium is in fact maintained, the sample sizes can be reduced, and the time between samples increased, for the latter part of the experiment. It is the *changes* in allele frequencies which allow us to estimate adaptive values. Hence, the initial frequencies should be chosen as far as possible from the experimenter's best guess for the equilibrium frequencies. Also, different replications of an experiment should start off with widely different initial frequencies. Little extra effort is required to check whether the *genotype* frequencies fit the multinomial distribution predicted by the Hardy-Weinberg law. Since the mathematical model for selection assumes so heavily that they do, a check with a simple chi-square test can prevent an unjustified analysis. The programs we have written for extracting the maximum likelihood estimates of the adaptive values and their covariance matrices, and for combining and comparing selection in different populations, are available from us.

This research was supported in part by the United States Army, Navy, Air Force and NASA under a contract administered by the United States Office of Naval Research. The authors wish to thank PROFESSOR LEONARD J. SAVAGE, Department of Statistics, Yale University, for his many suggestions and discussions concerning this paper.

SUMMARY

An analysis of selection is described in the case of experimental populations which are segregating for alleles at a single gene locus. Iterative procedures for obtaining the maximum likelihood estimates of adaptive values are described. The resulting vectors of estimated adaptive values and their corresponding FISHER information matrices are used to test the differences in selection between different populations. Examples of the procedure are presented, and suggestions to future experimenters are included.

LITERATURE CITED

ANDERSON, W. W., C. OSHIMA, T. WATANABE, TH. DOBZHANSKY, and O. PAVLOVSKY, 1968 Genetics of Natural Populations XXXIX. A test of the possible influence of two insecticides on the chromosomal polymorphism in *Drosophila pseudoobscura*. Genetics **58**: 423–434.

BAILEY, N. T. J., 1961 *Introduction to the mathematical theory of genetic linkage*. Clarendon Press, Oxford.

CAVALLI, L. L., 1950 The analysis of selection curves. Biometrics **6**: 208–220.

DOBZHANSKY, TH., and H. LEVENE, 1951 Development of heterosis through natural selection in experimental populations of *Drosophila pseudoobscura*. Am. Naturalist **85**: 247–264.

HALDANE, J. B. S., 1927 A mathematical theory of natural and artificial selection. Part IV. Proc. Cambridge Phil. Soc. **23**: 607–615.

HASOFER, A. M., 1966 A continuous-time model in population genetics. J. Theoret. Biol. **11**: 150–163.

KIMURA, M., 1958 On the change of population fitness by natural selection. Heredity **12**: 145–167.

KRUSKAL, J. B., 1964 Nonmetric multidimensional scaling: a numerical method. Psychometrika **29**: 115–129.

LEVENE, H., O. PAVLOVSKY, and TH. DOBZHANSKY, 1954 Interaction of the adaptive values in polymorphic experimental populations of *Drosophila pseudoobscura*. Evolution **8**: 335–349.

PAVLOVSKY, O., and TH. DOBZHANSKY, 1966 Genetics of Natural Populations XXXVII. The coadapted system of chromosomal variants in a population of *Drosophila pseudoobscura*. Genetics **53**: 843–854.

RAO, C. R., 1952 *Advanced Statistical Methods in Biometric Research*. John Wiley and Sons, New York.

SPANG, H. A., III, 1962 A review of minimization techniques for nonlinear functions. SIAM Rev. **4**: 343–365.

WRIGHT, S., and TH. DOBZHANSKY, 1946 Genetics of Natural Populations XII. Expreimental reproduction of some of the changes caused by natural selection in certain populations of *Drosophila pseudoobscura*. Genetics **31**: 142–160.