EVOLUTION

INTERNATIONAL JOURNAL OF ORGANIC EVOLUTION

PUBLISHED BY

THE SOCIETY FOR THE STUDY OF EVOLUTION

Vol. 42 **September, 1988** **No. 5**

Evolution, 42(5), 1988, pp. 849–861

ESTIMATING THE FORM OF NATURAL SELECTION ON A QUANTITATIVE TRAIT

DOLPH SCHLUTER

The Ecology Group, Department of Zoology, University of British Columbia, 6270 University Boulevard, Vancouver, BC V6T 2A9, CANADA

Abstract.—The fitness function f relates fitness of individuals to the quantitative trait under natural selection. The function is useful in predicting fitness differences among individuals and in revealing whether an optimum is present within the range of phenotypes in the population. It may also be thought of as describing the ecological environment in terms of the trait. Quadratic regression will approximate the fitness function from data (e.g., Lande and Arnold, 1983), but the method does not reliably indicate features of f such as the presence of modes (stabilizing selection) or dips (disruptive selection). I employ an alternative procedure requiring no a priori model for the function. The method is useful in two ways: it provides a nonparametric estimate of f, of interest by itself, and it can be used to suggest an appropriate parametric model. I also discuss measures of selection intensity based on the fitness function. Analysis of six data sets yields a variety of forms of f and provides new insights for some familiar cases. Low amounts of variation and a low density of data points near the tails of many phenotype distributions emerge as limitations to gaining information on fitness functions. An experimental approach in which the distribution of a quantitative trait is broadened through manipulation would minimize these problems.

Received August 25, 1987. Accepted March 4, 1988

Measurements of natural selection are invaluable in studies of adaptation. Recent theoretical work in this area has successfully addressed two main goals: first, to develop coefficients of selection intensity for quantitative traits, disentangling direct from indirect effects (Lande and Arnold, 1983; Manly, 1985), and second, to predict evolutionary response to observed selection events (Lande, 1979). Data on natural selection can also be used for a third purpose: to estimate the selection surface or fitness function, the unknown function f relating survival and/or reproductive success of individuals (W) to the phenotypic character z under selection:

$$W = f(z) + \text{random error}$$

(Pearson, 1903; Simpson, 1953; Lande, 1979; Lande and Arnold, 1983).

The fitness function is of interest because it is a complete description of selection pressures on individuals. For example, individuals may differentially survive according to body size, in which case W is survival (1 or 0) and $f(z)$ is the probability of survival as a function of size. An estimate of f would thus allow the quantitative prediction of survival probabilities of individuals, the comparison of probabilities among individuals differing in size, and the assessment of whether an optimum body size exists within the range of phenotypes present in the population.

A knowledge of f over a broader range of phenotypes than is actually found in any one population also leads to an estimate of the "adaptive landscape," describing mean fitness in the population $\bar{W}$ as a function of $\bar{z}$, the mean value of the phenotypic trait under selection (Wright, 1932; Lande, 1979). When the trait is heritable and uncorrelated with other traits under selection, then the direction of change in $\bar{z}$ leading to the maximum increase in mean fitness will usually be the direction of evolutionary change in the population.

The fitness function is additionally valuable in other research. For example, optimality theory is employed in many fields to predict features of a fitness surface from hypotheses about how a given trait may be advantageous in a particular environment. DeAngelis et al. (1985) predicted fitness in bivalves as a function of growth strategy, as determined by predation. Schluter and Grant (1984) estimated adaptive landscapes for mean beak size in Galápagos finch populations on the hypothesis that these are determined by seed supply. Powerful tests of such adaptive hypotheses could be made by quantitatively comparing the predicted form of the selection surface with independent estimates based on survival or reproductive success.

Similarly, if the ecological significance of a trait under selection is known, then the form of the fitness function will provide a useful description of the ecological environment in relevant units of fitness. For example, beak size in populations of Galápagos ground finches determines the efficiency with which foods of different size and hardness can be consumed and, consequently, determines diet. Multiple modes in the mean fitness landscapes for mean beak size describe a distribution of alternative feeding niches along a corresponding mean seed size/hardness axis (Schluter and Grant, 1984; Schluter et al., 1985).

A variety of regression methods may be used to estimate f or features of f from data on natural selection. Lande and Arnold (1983) suggested the use of quadratic regression to approximate the fitness function, the coefficients of which correspond to direct measures of selection intensity. This method can give misleading results, such as in indicating the presence of a dip or mode (see below). A more flexible approach would be to compare several regression models (e.g., Manly, 1976, 1985; Mitchell-Olds and Shaw, 1987). Choosing the best model will often be difficult, particularly when error distributions are nonnormal (e.g., survival data). Also, one cannot be assured that the appropriate model is among those initially fitted.

Here, I explore an alternative method to estimate the individual fitness function from data on natural selection, one not requiring that the parametric form of f be specified. The method has two uses: first, it provides a nonparametric estimate of f, of interest by itself, and second, it can suggest an appropriate parametric model if one is desired. I use the method to analyze several familiar data sets involving selection on a single variable and show where new insights are gained. Estimates of selection intensity based on nonparametric estimates of fitness functions are also discussed. Finally, I consider how to measure the sampling variability of the estimate $\hat{f}$. The more difficult problem of estimating surfaces involving two or more characters is ignored here but will be discussed in a forthcoming paper.

Definitions

In the general model for natural selection, survival and/or reproductive success of individuals is related to the phenotype z as $W = f(z) + \epsilon$, where ϵ is a random error. I defined $f(z)$ as individual fitness, because it measures average survival or reproductive success of all individuals having the phenotype z (e.g., Endler, 1986). W is often also referred to as "individual fitness" (Falconer, 1981; Lande and Arnold, 1983; Endler, 1986), but I will refrain from doing so here in order to prevent confusion. Relative fitness is fitness scaled to have a population mean of 1: $f_w(z) = f(z)/\bar{W}$, where $\bar{W}$ is mean survival and/or reproductive success.

Natural selection is defined as variability among phenotypes in fitness. Three types of univariate selection are commonly recognized, on the basis of the form of f over the range of phenotypes in the population (Endler, 1986 pp. 16–18). In directional selection, individuals on one side of the population mean are favored over those on the other. Directional selection may occur with other types, but in pure directional selection, fitness is either nondecreasing or nonincreasing over the range of phenotypes (i.e., f is monotonic). In stabilizing selection, intermediate phenotypes are favored over extremes—the function has a mode, or optimum. In disruptive selection, fitness of extreme individuals is greater than that of intermediates—the surface has a dip, or minimum. Lande and Arnold (1983) redefined these three types of selection (see below), but I use the more traditional meanings here.

Selection Coefficients and the Fitness Function

The most common procedure in the analysis of data on natural selection is to calculate coefficients of selection intensity. Here, I briefly review the most useful coefficients, those of Lande and Arnold (1983), in order to clarify their relationship to the selection surface, f, and to show how their incautious application can mislead.

The fitness function f can be thought of as a feature of the environment, describing the fitnesses of alternative phenotypes but existing independently of their actual distribution. In contrast, selection coefficients, such as those of Lande and Arnold (1983), measure the effects of f on the distribution of phenotypes in the population. To simplify calculations, assume that the single character z under selection has mean $\mu = 0$ and variance $\sigma^2 = 1$. The coefficient of directional selection is then

$$\beta = \mathrm{Cov}(w, z)$$

where $w = W/\bar{W}$, and $\bar{W}$ is mean survival and/or reproductive success. If, in addition, z is normally distributed, then

$$\gamma = \mathrm{Cov}(w, z^2)$$

is the "stabilizing" ($\gamma < 0$) or "disruptive" ($\gamma > 0$) selection coefficient. The value of γ is equal to the difference between the variances of z before and after selection, corrected by β^2, the reduction in variance resulting from directional selection alone: $\gamma = \sigma^2_{\mathrm{after}} - \sigma^2_{\mathrm{before}} + \beta^2$. The constants β and $\gamma/2$ are the linear and quadratic coefficients from the quadratic regression of w on z, when z is normally distributed (Lande and Arnold, 1983).

However, the coefficient γ does not necessarily indicate stabilizing or disruptive selection, as these terms are defined above, when directional selection occurs (see also Mitchell-Olds and Shaw [1987]). For example, consider truncation selection in a normally distributed population with mean 0 and variance 1 (Fig. 1). All individuals having a phenotypic value $z \geq a$ survive,

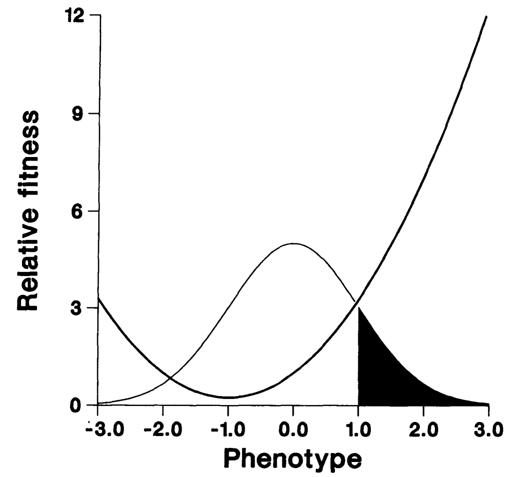


FIG. 1. Misleading approximation of a relative fitness function f_w using quadratic regression. Bell-shaped curve and shaded area indicate truncation selection on a normally distributed trait z with mean 0 and variance 1: all individuals above $z = 1.0$ survive, and all individuals below $z = 1.0$ die. The regression surface (inverted parabola) erroneously suggests the presence of disruptive selection.

while individuals with $z < a$ do not. Selection is purely directional, but it can be shown that the coefficient γ is nonzero whenever a is not the mean phenotype: $\gamma = ae^{-a^2/2}/q\sqrt{2\pi}$, where q is the fraction of individuals in the population surviving. The truncation point in Figure 1 is $a = 1.0$, and γ is large and negative, as indicated by the dip in the best-fit quadratic regression of w on z. A similar situation occurs in the multivariate case, where pure directional selection can result in a significant "correlational" selection coefficient (cf. Lande and Arnold, 1983).

The value of γ fails to indicate the presence of a mode or dip in f because the quantity β^2 does not correct for changes in variance resulting from directional selection. For example, if z is normally distributed as $p(z)$, with $\mu = 0$ and $\sigma^2 = 1$, and if the fitness function f is smooth (i.e., is twice-differentiable [$''$]), then the total change in variance caused by selection, directional or otherwise, is actually

$$\sigma^2_{\mathrm{after}} - \sigma^2_{\mathrm{before}} = -\beta^2 + E[f_w''(z)]$$

where $f_w(z) = f(z)/\bar{W}$. The quantity $E[f_w''(z)] = \int f_w''(z)p(z)\,dz$ measures mean curvature of the surface f_w over the distribution of phenotypes (Lande and Arnold, 1983). Thus,

β^2 will measure the change in variance resulting from directional selection only if the purely directional (monotonic) fitness surface f_w is without mean curvature (e.g., is linear). Many directional fitness surfaces are curved, such as those having an asymptote, and in such cases γ will be nonzero even though stabilizing and disruptive selection are absent.

The absence of a straightforward correspondence between selection coefficients and features of the selection surface does not detract from the importance of the Lande and Arnold (1983) results. Correctly applied and interpreted, their coefficients can successfully measure direct and indirect selection and can relate phenotypic patterns to evolutionary change; knowledge of f is neither necessary nor sufficient to carry out these goals. In order to avoid confusion, γ should merely be renamed (e.g., as the coefficient of nonlinear or quadratic selection). However, the problems underscore the view stressed here, that identification of stabilizing and disruptive (and correlational) selection and identification of other features of f should always be made with reference to a direct estimate of the fitness function. In the following sections, I apply a nonparametric technique to make such an estimate.

Materials and Methods

Nonparametric Estimation of Fitness Functions

The Cubic Spline.—In order to estimate the fitness function f from individual data on survival and/or reproduction (W) and phenotypic measurements (z), a reasonable approach is to begin with maximum-likelihood. Select an estimate $\hat{f}$ from the set of all possible continuous fitness surfaces to maximize the log likelihood

$$\ell(f) = \sum_{i=1}^{n} l(W_i;\, z_i, f)$$

where the sum is over the n individuals in the sample. The term $l(W_i;\, z_i, f)$ is the $\log_e$ probability that $W = W_i$ when $z = z_i$, under the particular choice of f. In the case that W is normally distributed around $f(z)$ with constant variance, $l(W_i;\, z_i, f)$ is proportional to $-[W_i - f(z_i)]^2$, and the $\hat{f}$ that maximizes $\ell(f)$ is the familiar least-squares estimate. Unfortunately, any function connecting all n data points will maximize the likelihood. Such a curve would be extremely rough, zigzagging repeatedly among the different values of W, and would have low predictive value. It would also offend our prejudice that the true fitness surface is smoother and simpler.

The failure of the above method has led to the modified technique of penalized maximum likelihood (Good and Gaskins, 1971). We select the function $\hat{f}$ to maximize the penalized log likelihood

$$\ell(f) = \sum_{i=1}^{n} l(W_i;\, z_i, f) - n\lambda J(f) \quad (1)$$

where λ is a nonnegative constant and $J(f)$ is the summed (squared) curvature of f, a measure of "roughness":

$$J(f) = \int [f''(z)]^2\, dz. \quad (2)$$

Note that the functions considered in the maximization of (1) cannot have sharp creases, or the operation within square brackets in (2) (double differentiation) will not be possible.

Assume for the moment that W is normally distributed around $f(z)$ with constant variance (nonnormal errors are discussed below). For a given λ, the $\hat{f}$ that maximizes (1) is a cubic spline, a function comprising $n + 1$ cubic polynomials joined seamlessly at points corresponding to the n phenotypic values. The value of λ controls the importance of the roughness penalty $J(f)$ to the sum in (1) and, thereby, determines the exact form of the estimate $\hat{f}$. A value for λ must therefore be chosen wisely. When λ is small, little price is paid for roughness, and the best function [the one that maximizes (1)] will nearly fit the individual data points. The penalty term assumes greater weight as λ gets larger, and so to compensate, the best $\hat{f}$ must be smoother. This effect is illustrated using Houde's (1987) data on sexual selection in guppies (Fig. 2), assuming normal errors. In the limit, as λ approaches infinity, the best function will be the linear regression of W on z. The constant λ is often called the "smoothing parameter."

An excellent choice for λ is one providing maximal predictive power in the given data set, computed using the method of cross-validation. For a given λ and an individual i, let $\hat{f}^i$ be the function that maximizes (1), but with individual i excluded from the sum. The function $\hat{f}^i$ is computed for all n individuals in the sample, yielding the sum of squared prediction errors $\Sigma(\hat{f}^i(z_i) - W_i)^2$. Short-cut methods approximate the sum of squared prediction errors and yield the generalized cross-validation (GCV) score (Craven and Wahba, 1979). The value for λ that minimizes the GCV score is known to minimize approximately the sum of squared deviations between the estimate $\hat{f}$ and the true fitness surface f. The GCV score is therefore an appropriate criterion for estimating f. Such a GCV estimate is shown in Figure 2 (solid line) and indicates a possible mode in the sexual-selection surface.

The cubic-spline approach is flexible, allowing the estimation of a wide variety of fitness surfaces. Standard linear, quadratic, and higher-order polynomial functions (e.g., cubic regression) are themselves cubic splines, and such functions will result if the data warrants, but the method is in no way restricted to these forms. Intuition as to how the cubic spline actually yields a smooth estimate of f is provided by Silverman (1984), who shows that the method is approximately equivalent to the weighted moving average, where the number of points included in the average varies with the local density of data points.

The relationship to the moving average clarifies some of the properties of the cubic spline. The spline provides a more local estimate of the regression surface than standard polynomial regression, which yields a global fit to the data. For example, both methods may be applied to a given data set involving a cluster of (W,z) values. If a single outlying point is then added and the regressions are recomputed, the form of the regression within the original cluster of points may be substantially altered in the polynomial case, but not in the spline. A second, less desirable, property is that the spline, like the moving average, can be biased in that the estimate of f tends to be too smooth where the slope of the true fitness function changes rapidly (e.g., at the apex of a narrow peak). Gu (1987) suggests a bootstrap procedure to correct partly for bias. Estimated bias was very small in the data sets that I analyzed, and so I do not correct for it in my results.

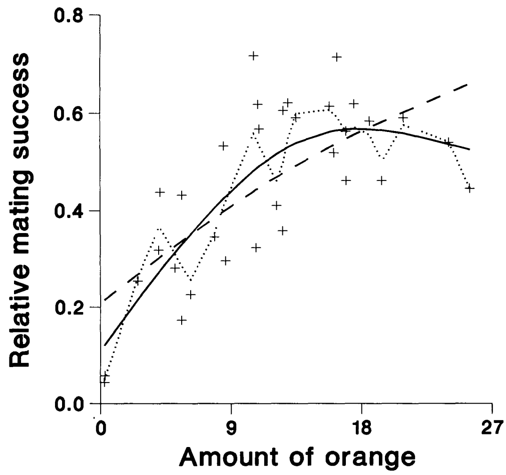


FIG. 2. Effect of λ on the estimate of the fitness function. Data are mating success of male guppies (*Poecilia reticulata*; inferred from data on female behavior) and the percentage of their body area that is colored with orange (taken from figure 3 in Houde [1987]; $N = 32$). Alternative estimates of f represent $\lambda = 0.002$ (dotted), $\lambda = 8.2$ (solid), and $\lambda = 150$ (dashed). The solid line minimizes the GCV score.

Nonnormal Errors.—Measures of survival and reproductive success are often not normally distributed. For example survival is binomial and takes on values 0 or 1. The distribution of number of offspring may, for a given value of z, be approximately Poisson. The main problem with these alternative distributions is that fitness is restricted ($0 \leq f(z) \leq 1$ in the binomial case, and $f(z) \geq 0$ for a Poisson distribution), and that the variance σ_ϵ^2 of the errors around f is not constant but depends on $f(z)$: $\sigma_\epsilon^2 = f(z)(1 - f(z))$ in the binomial case and $\sigma_\epsilon^2 = f(z)$ in Poisson. Data transformations to stabilize the variance are not the answer; for example, the $\log_e$-transformation cannot be used on zeros.

These problems can be solved using standard iterative methods of generalized linear models (McCullagh and Nelder, 1983) in combination with the cubic spline (O'Sullivan et al., 1986). In this case, the modified goal is to maximize the penalized log likelihood $\ell(g)$ rather than $\ell(f)$, where $g(z) =$

$\log_e[f(z)/\{1 - f(z)\}]$ for binomial data and $g(z) = \log_e[f(z)]$ for Poisson data. The method can be thought of as a nonparametric equivalent to logistic and log-linear regression. Note that no regression technique for estimating f, including the cubic spline, requires that the distribution of the phenotypic trait z be normal.

The ability to handle nonnormal errors also allows the method to be used for "cross-sectional" or other data involving two independent samples. For example, Hagen and Gilbertson (1973) made a collection of sticklebacks when individuals were aged 2½ months, and another when they were one year old. Estimates of f can be made by scoring individuals in the first sample as $W = 0$ and those in the second as $W = 1$ and regressing this variable on the trait of interest. The regression will not estimate the actual fitness function, but a function h that is monotonically related to fitness (survival probability) and has all of the same features ($\log_e[h(z)/\{1 - h(z)\}]$ will differ from $g(z) = \log_e[f(z)/\{1 - f(z)\}]$ by a constant). Only traits not subject to growth can be analyzed in this fashion.

Computer Routines.—Efficient FORTRAN77 routines for computing f with normal, binomial and Poisson data were kindly provided by Douglas Nychka (Department of Statistics, North Carolina State University, Raleigh). An easy-to-use, interactive program employing his routines is available from me on request. The program also includes an option to generate bootstrap samples randomly, which is useful in assessing accuracy of $\hat{f}$ (see below). The program is designed to run on a PC-compatible microcomputer, but it should work on any system.

Other programs also exist. The most general, all-purpose routines are included in the public package PGLMPACK (Yandell, 1987), obtainable from Brian Yandell (Department of Statistics, University of Wisconsin, Madison). These will handle multivariate data, and splines other than the cubic. However, they are inefficient in the univariate case and have limited use on a microcomputer. All analysis in the present paper was done on an IBM-XT using Nychka's program.

RESULTS

Here I use the above methods to estimate the fitness function in several other cases. The purpose of the analysis is to see whether new information might be obtained from data that have previously been analyzed in other ways and to indicate the diversity of functions that may be observed in studies of natural selection. I also describe measures of selection intensity based on estimates of f and a way to assess the accuracy of an estimate.

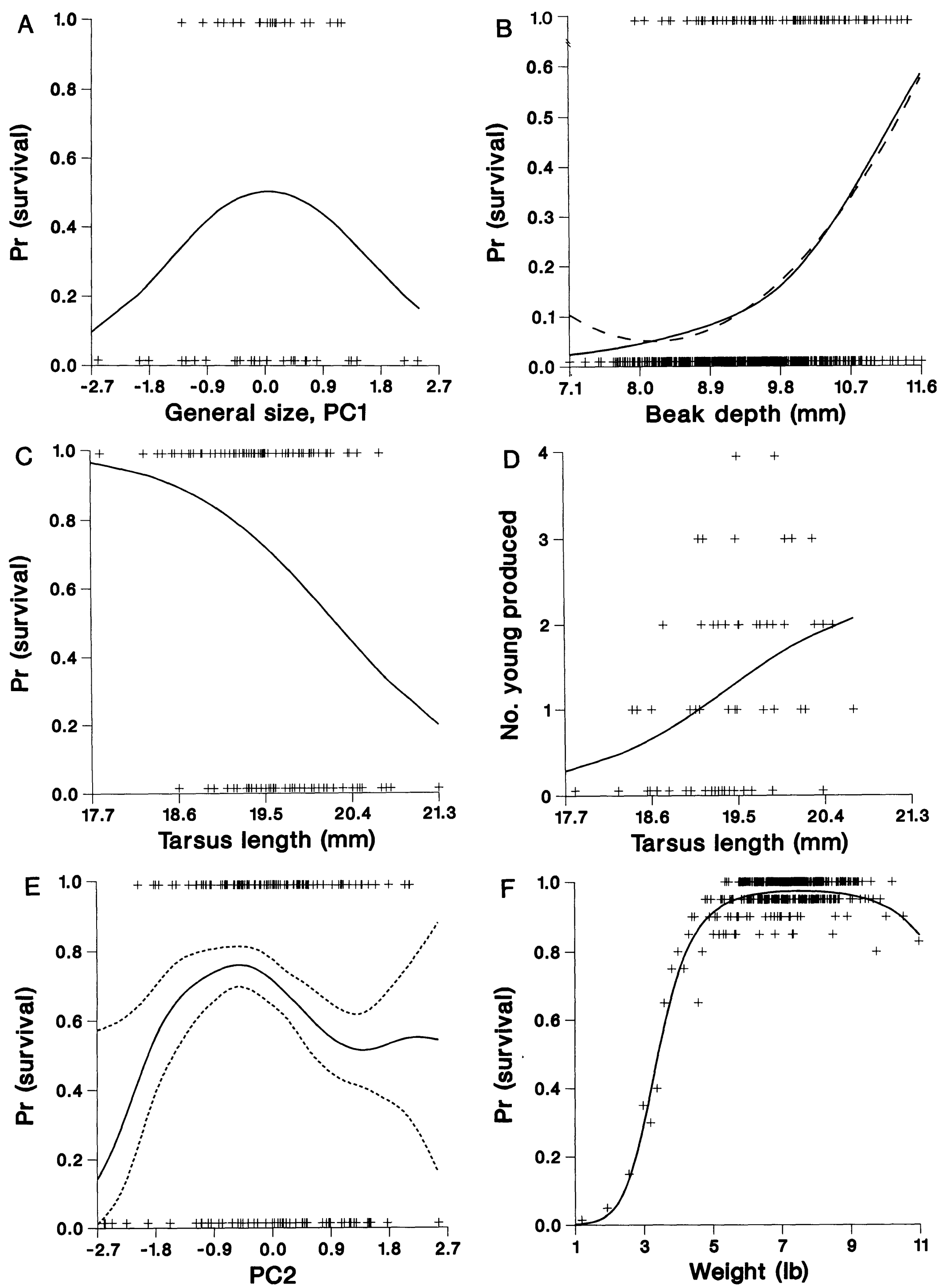
Fitness Functions

House Sparrows.—Bumpus (1899) recorded survival in relation to morphology in 49 female and 87 male house sparrows (*Passer domesticus*) during a New England winter storm. Nine traits were measured, including total body length, wing length, body mass, head length, humerus length, femur length, tibio-tarsus length, skull width, and sternum length. Bumpus published his raw data, and as a consequence they have been frequently analyzed (Grant, 1972; Johnston et al., 1972; O'Donald, 1973; Lande and Arnold, 1983; Manly, 1985). Results have suggested directional selection in males and stabilizing selection in females.

An estimate of the fitness function f in relation to the first principal component in female house sparrows (general size, based on the covariance matrix of all nine traits, $\log_e$-transformed) confirms the presence of stabilizing selection (Fig. 3A). Probability of survival was maximal at the mean phenotype ($\bar{z} = 0$), declining from 0.5 to about 0.1 at the tails of the phenotype distribution.

→

FIG. 3. Nonparametric estimates of fitness functions f in six data sets. The "+" symbols indicate raw data values, except in 3F, where sample size is very large and averages of sequential groups of 20 individuals are given. A) Survival of female house sparrows (Bumpus, 1899; $N = 49$) in relation to standardized PC1. B) Survival in the Galápagos finch *Geospiza fortis* (Boag and Grant, 1981; $N = 642$) in relation to beak size. The dashed curve is the estimate of the surface using quadratic regression. C) Overwinter survival in juvenile female song sparrows ($N = 145$) as a function of tarsus length. Data are from four winters combined (Schluter and Smith,



1986). D) Reproductive success in female song sparrows in 1979 ($N = 57$) as a function of tarsus length (Schluter and Smith, 1986). E) Overwinter survival in juvenile male song sparrows ($N = 152$) in relation to standardized PC2. Data are from four winters combined (Schluter and Smith, 1986). Dashed curves indicate ± 1 SE of predicted values $\hat{f}(z)$ from 200 bootstrap replicates of the fitness function. F) Survival in male human infants (Karn and Penrose, 1951; $N = 7{,}037$) as a function of birth weight.

The function is roughly parabolic and might be approximated by a quadratic (e.g., Lande and Arnold, 1983) or a normal model (e.g., O'Donald, 1973).

Galápagos Finches.—Boag and Grant (1981, 1984) described nonrandom survival in relation to various beak and body dimensions in 642 individuals (males and females combined) of the Galápagos ground finch, *Geospiza fortis,* over a prolonged dry season. Price et al. (1984), using the methods of Lande and Arnold (1983), showed that directional selection was particularly strong on beak depth, beak width, and body mass. Schluter et al. (1985), also using the methods of Lande and Arnold (1983), suggested that some disruptive selection was present in addition to directional selection, as shown by a quadratic regression of survival against beak depth (Fig. 3B).

The spline estimate of the fitness function in *G. fortis* showed probability of survival increasing exponentially with increasing beak depth over the entire range of phenotypes (Fig. 3B). Thus, selection was apparently purely directional, and disruptive selection was absent, in contrast to the earlier result. The spline is probably closer to the true function: inspection of the data shows that the quadratic curve turns upward at small beak sizes, even though the very smallest individuals did not survive. The upturn is a good illustration of the fact that a quadratic regression may provide a reasonable overall fit to the data, but its form is not necessarily responsive to local features (see also Fig. 1).

The revised estimate of f in *G. fortis* does not alter a main conclusion of Schluter et al. (1985), that the small and large seeds favor different beak sizes in the population. Large size was favored, but some small-beaked individuals survived solely by efficiently exploiting small seeds (Price, 1987). Data from other life stages (Price and Grant, 1984) and years (Gibbs and Grant, 1987) confirm that selection may at other times favor small size instead of large size. However, the conclusion of disruptive selection in Schluter et al. (1985) is not supported by reanalysis.

Song Sparrows.—Schluter and Smith (1986) described natural selection at different life stages over a five-year period for the resident song sparrow (*Melospiza melodia*) of Mandarte Island, British Columbia. Strong selection was found in females, particularly for tarsus length and beak length. Estimates are given in Figure 3C and D of the relation between fitness (probability of juvenile recruitment and mean number of young produced) and tarsus length in females. Recruitment was a binomial variable, while the number of young produced by a given female was assumed to have a Poisson distribution. The two curves show the opposing relations between fitness and the trait at different life stages (Schluter and Smith, 1986). They also show the enormous differences in fitness accompanying slight differences in tarsus length. Over a 3.6-mm range, estimated probability of recruitment declined from nearly 1.0 to 0.2, while estimated reproductive success increased more than seven-fold. Selection was purely directional in both cases, but f was not quite linear, particularly in survival, where an asymptote was approached at low values of tarsus length (Fig. 3C).

Analysis of male song sparrows had suggested the presence of stabilizing selection on the second principal component (PC2) in association with juvenile survival (Schluter and Smith, 1986). PC2 was obtained from the covariance matrix of six external morphological traits and indicated wing length, tarsus length, and body mass relative to length, depth, and width of the beak. A nonparametric estimate of f (Fig. 3E) confirms the presence of a mode to the left of the mean of the phenotype distribution ($\bar{z} = 0.0$). However, the function was not symmetric but slowly equilibrated to an intermediate value of fitness at the large end of the phenotype distribution. A tiny second mode, probably not real, was also present at large values of PC2.

Human Infants.—Karn and Penrose (1951) presented data on survival to 28 days in newborn humans and its relation to weight at birth. They fitted the data to a logistic regression model with a quadratic term, which indicated an optimum birth weight at about 8 lb (3.6 kg) in both males and females. Their result is a classic example of stabilizing selection in our species. Similar studies are reviewed by Van Valen and Mellin (1967) and Ulizzi and Terrenato (1987).

The relation between probability of survival and weight at birth in 7,037 male infants was estimated using the cubic spline (Fig. 3F). I was initially unable to obtain a sensible result using this method—no minimum value for the GCV score was observed. The problem probably resulted in part from the small number of unique values for weight (23), since Karn and Penrose (1951) presented their observations rounded to the nearest half pound. Rounding seems to render the method especially sensitive to values at the ends of the distribution, where data are sparse and, as a result of rounding, separated from the remaining data by gaps. The problem disappeared when I grouped the largest three infants (11.5, 11.5, and 13 lb) with those in the 11.0-lb class.

The nonparametric estimate of f agreed with the earlier analysis and indicated stabilizing selection (Fig. 3F). However, the mode in $\hat{f}$ was very broad, and probability of survival exceeded 0.90 over the range 5.0–10.0 lb. Estimated fitness declined only slightly at larger infant sizes, in contrast to very low probabilities of survival accompanying the smallest birth weights. Thus, most selection appeared to be directional. Maximum probability of survival (0.97) was observed at 7.5 lb, but this must be considered a very tentative estimate, given the flatness of the mode.

Selection Intensity

The present paper is mainly concerned with estimating the selection function, but such an estimate can be employed further. For example it can be used to estimate selection "intensity," a useful quantity for comparing different selection episodes. Selection intensity is usually defined as a function of intrapopulation variation in fitness (Haldane, 1954; Van Valen, 1965; O'Donald, 1970; Manly, 1977) or of covariation between fitness and the phenotypic traits (Lande and Arnold, 1983). Intensity is not a property of the fitness function alone, but of both the function and the distribution of phenotypes in the population.

Natural selection is defined as variation in fitness, and hence, a straightforward measure of total selection intensity is variance in relative fitness itself,

$$V = \mathrm{Var}(f_w(z)).$$

TABLE 1. Estimated selection intensity V in six episodes (cf. Fig. 3).

Population	Fitness measure	Intensity
Female house sparrows	survival	0.054
Galápagos finches	survival	0.453
Female song sparrows	survival	0.049
Female song sparrows	reproduction	0.114
Male song sparrows	survival	0.031
Male human infants	survival	0.012

Variance in relative fitness as defined here should not be confused with Var(w), the variance in relative survival or reproductive success, known as the opportunity for selection (Arnold and Wade, 1984). V is estimated as the squared coefficient of variation among individuals in the predicted values $\hat{f}(z)$. The measure will work for nonparametric as well as parametric forms of f, provided that the function is based on the entire population or a random sample. V is similar to Manly's (1977) index, but is not wedded to a particular selection model and is measured directly on $\hat{f}(z)$ and not on a transformed scale. The measure does not reflect selection on specific parameters (moments) of the phenotype distribution, as do those of Lande and Arnold (1983), and hence, it cannot be used to predict evolutionary change in the parameters. Rather, it reflects selection on all the moments of the phenotype distribution combined and is purely descriptive.

Estimated selection intensities for the populations illustrated in Figure 3 are given in Table 1. Values show weak selection in human infants and male song sparrows and relatively intense selection associated with survival in Galápagos finches and with reproduction in female song sparrows. I did not calculate V for the guppy data (Fig. 2), since the distribution of the phenotypic trait in the males tested differed from that in a random sample of males (Houde, 1987).

It is further possible to calculate an intensity of stabilizing (or disruptive) selection by measuring the contribution of a mode (or dip) in $\hat{f}$ to the total selection experienced by a population. The simplest approach is to compare total selection intensity V with variance in relative fitness recomputed after the mode is removed (V_0). The difference between V and V_0 will mea-

TABLE 2. The intensity of stabilizing selection V_s in two populations (cf. Figs. 3A,F). Total selection intensity V is provided for comparison.

Measure	House sparrows	Human infants
V_s	0.027	0.000
V	0.056	0.012

sure the fraction of the total variance in relative fitness among individuals that results directly from the mode:

$$V_s = V - V_0.$$

As with total selection intensity, the measure is purely descriptive and cannot be used to predict evolutionary change.

One way to remove a mode in $\hat{f}$ is to increase the value of the smoothing parameter λ [Eq. (1)] until the observed mode has barely disappeared. An advantage of this method is that if a mode is symmetric about $\bar{z}$, the constrained $\hat{f}$ will be a horizontal line, and $V_s = V$. A disadvantage is that the variances are not additive: V_0 may exceed V and give negative values of V_s. I used an alternative approach, taking the predicted values $\hat{f}(z)$ from the GCV estimate of f and fitting the maximum-likelihood monotonic function using the "pool adjacent violators" algorithm of Barlow et al. (1972) (of the two monotonic functions that can be fit, nondecreasing and nonincreasing, choose the one yielding greater variance in predicted relative fitness). One disadvantage of this method is that if a mode is symmetric about $\bar{z}$ the constrained $\hat{f}$ will not be flat, and $V_s < V$ (in practice, V_s will be roughly $V/2$).

I calculated V_s in female house sparrows and male human infants, which show alternative extreme effects of a mode in $\hat{f}$ on a population (Table 2). In house sparrows, the mode is nearly symmetric about $\bar{z}$ (Fig. 3A), and V_s is large relative to V (i.e., is $V/2$), indicating that variability among individuals in fitness resulted primarily from stabilizing selection. In contrast, the mode in $\hat{f}$ is broad in human infants, and fitness shows a substantial decline only at the lower end of the phenotype distribution (Fig. 3F). Thus, V_s is small relative to V, indicating that stabilizing selection was very weak and that most variability in fitness in the population was the result of directional selection.

The above method can be generalized to other measures of selection and other features of $\hat{f}$. For example, the intensity of selection associated with a second of two modes can be estimated by fitting a function constrained to have only one. The amount of "quadratic" selection in the population $\gamma = \mathrm{Cov}(w, z^2) \approx \mathrm{Cov}(f_w, z^2)$ (Lande and Arnold, 1983) attributable to the presence of a mode can similarly be computed.

Accuracy

Finally, it is desirable to consider the sampling variability in a nonparametric estimate of the fitness function. A simple way to judge variability is by using the bootstrap (Efron, 1982; Efron and Tibshirani, 1986), where regression surfaces are repeatedly estimated on data resampled from the original set. The performance of the bootstrap in the case of splines has not yet been tested using simulation. Hence, bootstrap results should be considered approximate and interpreted cautiously.

Two hundred bootstrap regressions were calculated using survival in male song sparrows (Fig. 3E). Each repetition involved two steps: i) the random generation of a new W_i (0 or 1) for each observed z_i, where $\Pr(W_i = 1) = \hat{f}(z_i)$ and where $\hat{f}(z_i)$ is the predicted value in the original estimate of f; ii) a search for the cubic spline minimizing the GCV score. Standard errors for the 200 resulting predicted values for each z_i are shown in Figure 3E. Standard errors are not symmetric about $\hat{f}$, since they were computed on the logistic scale $\log_e[\hat{f}(z)/\{1 - \hat{f}(z)\}]$.

The bootstrap results illustrate two points. First, estimates of f may be quite variable, although this depends on sample size and the proximity of $f(z)$ to the endpoints (0 and 1). Nonparametric estimates have more sampling variation than parametric regression surfaces, and this is the price paid for not knowing a priori the parametric form of f. Second, uncertainty is greatest at the ends of the phenotype distribution, a consequence of a reduced density of data points. This is not a limitation specific to the nonparametric method: in all regressions, the data are most informative where they are most abundant. However, it is a feature of many natural phenotype distributions that the fitness function can be estimated with accuracy over only a limited phenotypic range.

The bootstrap can also be used to test hypotheses concerning features of f. For example, the presence of a mode can be tested by counting the frequency of repetitions in which the mode is present. In carrying out such tests, it is wisest to restrict the mode search to a predefined region of z values, as modes may occasionally appear as a result of sampling error elsewhere, particularly in the tails of the phenotype distribution. For example, I carried out 200 bootstrap replications of the guppy data (Fig. 2), recording the number of modes in $\hat{f}$ appearing between $z = 10$ and $z = 26$. The frequency of replicates in which no mode appeared was 30%, suggesting that a null hypothesis of purely directional selection would not be rejected by these data.

When testing for accuracy of functions of $\hat{f}$ and the data (e.g., variance in relative fitness), an alternative resampling procedure from the one suggested above should be employed, one which resamples also the z values, in order to account for sampling variability in z.

Discussion

Many questions about the workings of natural selection on quantitative traits are directly concerned with features of the fitness function (Lande and Arnold, 1983; Endler, 1986). Do modes in fitness commonly occur within the range of phenotypes in the population? How frequent is disruptive selection? Are multiple peaks present? What is the range in fitness in a typical population? Do functions vary between life stages, years, or geographic localities? I have summarized a nonparametric method which can be generally used to estimate fitness functions and, thus, could be used to answer the above questions.

Regression using splines can be highly revealing, as demonstrated by insights into some familiar and well analyzed cases of natural selection. The Galápagos finch *Geospiza fortis* was previously thought to represent one of the few cases of disruptive selection in nature (Schluter et al., 1985; Endler, 1986), but the spline suggests that selection was purely directional (monotonic). Analysis of the human infant data of Karn and Penrose (1951) gave the unexpected result that very little stabilizing selection was present and that the majority of variance in fitness is attributable to directional selection. Other known cases of disruptive and stabilizing selection might be reevaluated in this light. Another insight is the range of fitnesses within populations. The selection event in *G. fortis* may be the most intense example from a vertebrate (Boag and Grant, 1981; e.g., Table 1), yet the estimated range in probability of survival was as high or higher in some other species (Fig. 3). Strong selection in *G. fortis* resulted mainly from a lower mean fitness.

In theory, fitness functions may assume a variety of shapes, and surfaces predicted from ecological or biomechanical considerations often show a complex pattern of peaks and valleys (Schluter and Grant, 1984; DeAngelis et al., 1985). An advantage of the nonparametric method is that estimates of the function are not restricted a priori to a particular model of selection—one that usually assumes the presence of at most a single mode. If functions in fact have a complex pattern of peaks and valleys, then a nonparametric approach is the best way to reveal it. However, the present method will also estimate simple forms of f and could be used to suggest a parametric model for the function, if one is desired. Several of the estimates of f shown in Figure 3 might subsequently be fit to a simple parametric model (e.g., normal, exponential, logistic, or quadratic).

Though the spline is flexible, estimates of fitness functions based on measurements of natural selection were usually simple (Fig. 3), involving only directional and/or stabilizing selection. This simplicity of observed forms of f, when complex patterns are possible and even expected, may result from small sample size in some cases. It is simply not possible to reveal a complex pattern when the density of data points available to estimate a local feature of f is small. Also, if the fitness function is indeed complex and interesting, the range of phenotypes in the population may be too narrow to detect it. Most of my example data sets are from birds, where coefficients of variation for morphological traits are typically less than 5%. Frequency of individuals is generally low toward the ends of the phenotype distribution, and this further restricts the range of phenotype values over which accurate information might be gained about the form of f (Fig. 3E).

If the form of f is of interest and if levels of phenotypic variation are low, researchers might consider alternative strategies, including deliberate experimental manipulations of the mean and variance of the quantitative trait. Examples of this approach include manipulations of clutch size (e.g., Lessels, 1986; Boyce and Perrins, 1987; Finke et al., 1987) and tail length (Andersson, 1982) in birds and of leaf area in plants (Willson and Price, 1980). A population with inflated phenotypic variance would be a useful experimental "probe" of the selective environment, allowing estimation of f over a broad range. It would also allow calculation of mean fitness $\bar{W}$ as a function of mean value of the character, useful in predicting the outcome of selection and evolution. Manipulated distributions should be relatively flat (uniform) rather than normal, so that equally reliable information can be gained over most of the range of z. Some instability at the tails of the distribution will nonetheless be present, owing to the necessary decline in the number of points falling within any fixed region of z as the edge is included. This effect can be minimized by further inflating the frequency of individuals at the tails.

The above strategy would assume that natural selection might reasonably be expected to occur within the timespan of a typical ecological study. The large number of cases of natural selection that have already been observed involving only natural levels of phenotypic variation (Endler, 1986) is highly encouraging. Indeed, the realization that detectable natural selection is a frequent event in ecological time may herald a new era of controlled experimental studies of the process. In this context, the present methods to estimate fitness functions should prove especially useful.

Acknowledgments

Thanks to N. Heckman and T. Price for many hours of discussion and to D. Nychka and B. Yandell for providing the computer programs. P. Grant and J. Smith kindly provided the data on Galápagos finches and song sparrows. T. Price suggested the example in Figure 1. Thanks also to S. Arnold, P. Boag, P. Grant, N. Heckman, T. Mitchell-Olds, D. Nychka, T. Price, G. Pogson, J. Smith, and C. Wehrhahn for comments on the manuscript. This work was supported by the Natural Sciences and Engineering Research Council of Canada (#U0459).

Literature Cited

Andersson, M. 1982. Female choice selects for extreme tail length in a widowbird. Nature 299:818–820.

Arnold, S. J., and M. J. Wade. 1984. On the measurement of natural and sexual selection: Theory. Evolution 38:709–719.

Barlow, R. T., D. J. Bartholomew, J. M. Bremner, and H. D. Brunk. 1972. Statistical Inference Under Order Restrictions. Wiley, London, U.K.

Boag, P. T., and P. R. Grant. 1981. Intense natural selection in a population of Darwin's finches (Geospizinae) in the Galápagos. Science 214:82–85.

———. 1984. The classical case of character release: Darwin's finches (*Geospiza*) on Isla Daphne Major, Galápagos. Biol. J. Linn. Soc. 22:243–287.

Boyce, M. S., and C. M. Perrins. 1987. Optimizing clutch size in a fluctuating environment. Ecology 68:142–153.

Bumpus, H. C. 1899. The elimination of the unfit as illustrated by the introduced sparrow, *Passer domesticus*. Biol. Lect. Woods Hole Mar. Biol. Sta. 6: 209–226.

Craven, P., and G. Wahba. 1979. Smoothing noisy data with spline functions: Estimating the correct degree of smoothing by the method of generalized cross-validation. Numer. Math. 31:377–403.

DeAngelis, D. L., J. A. Kitchell, and W. M. Post. 1985. The influence of naticid predation on evolutionary strategies of bivalve prey: Conclusions from a model. Amer. Natur. 126:817–842.

Efron, B. 1982. The Jackknife, the Bootstrap, and Other Resampling Plans. Society for Industrial and Applied Mathematics, Philadelphia, PA.

Efron, B., and R. Tibshirani. 1986. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. Stat. Sci. 1:54–77.

Endler, J. A. 1986. Natural Selection in the Wild. Princeton Univ. Press, Princeton, NJ.

Falconer, D. S. 1981. Introduction to Quantitative Genetics, 2nd Ed. Longman, N.Y.

Finke, M. A., D. J. Milinkovich, and C. F. Tompson. 1987. Evolution of clutch size: An experimental test in the house wren (*Troglodytes aedon*). J. Anim. Ecol. 56:99–114.

Gibbs, H. L., and P. R. Grant. 1987. Oscillating selection on Darwin's finches. Nature 327:511–513.

Good, I. J., and R. A. Gaskins. 1971. Non-parametric roughness penalties for probability densities. Biometrika 58:255–277.

Grant, P. R. 1972. Centripetal selection and the house sparrow. Syst. Zool. 21:23–30.

GU, C. 1987. What happens when bootstrapping the smoothing spline? Comm. Stat. Theor. Meth. 16: 3275–3284.

HAGEN, D. W., AND L. G. GILBERTSON. 1973. Selective predation and the intensity of selection acting upon the lateral plates of threespine sticklebacks. Heredity 30:273–287.

HALDANE, J. B. S. 1954. The measurement of natural selection. Proc. IX Internat. Congr. Genet. 1:480–487.

HOUDE, A. E. 1987. Mate choice based upon naturally occurring color-pattern variation in a guppy population. Evolution 41:1–10.

JOHNSTON, R. F., D. M. NILES, AND S. A. ROHWER. 1972. Hermon Bumpus and natural selection in the house sparrow *Passer domesticus*. Evolution 26: 20–31.

KARN, M. N., AND L. S. PENROSE. 1951. Birth weight and gestation time in relation to maternal age, parity and infant survival. Ann. Eugen. 16:147–164.

LANDE, R. 1979. Quantitative genetic analysis of multivariate evolution, applied to brain:body size allometry. Evolution 33:402–416.

LANDE, R., AND S. J. ARNOLD. 1983. The measurement of selection on correlated characters. Evolution 37:1210–1226.

LESSELS, C. M. 1986. Brood size in Canada geese: A manipulation experiment. J. Anim. Ecol. 55:669–689.

MANLY, B. F. J. 1976. Some examples of double exponential fitness functions. Heredity 36:229–234.

———. 1977. A new index for the intensity of natural selection. Heredity 38:321–328.

———. 1985. The Statistics of Natural Selection on Animal Populations. Chapman and Hall, N.Y.

MCCULLAGH, P., AND J. A. NELDER. 1983. Generalized Linear Models. Chapman and Hall, N.Y.

MITCHELL-OLDS, T., AND R. G. SHAW. 1987. Regression analysis of natural selection: Statistical inference and biological interpretation. Evolution 41: 1149–1161.

O'DONALD, P. 1970. Change of fitness by selection for a quantitative character. Theoret. Popul. Biol. 1:219–232.

———. 1973. A further analysis of Bumpus' data: The intensity of natural selection. Evolution 27: 398–404.

O'SULLIVAN, F., B. S. YANDELL, AND W. J. RAYNOR, JR. 1986. Automatic smoothing of regression functions in generalized linear models. J. Amer. Statist. Assoc. 81:96–103.

PEARSON, K. 1903. Mathematical contributions to the theory of evolution. XI. On the influence of natural selection on the variability and correlation of organs. Phil. Trans. Roy. Soc. Lond. A 200:1–66.

PRICE, T. D. 1987. Diet variation in a population of Darwin's finches. Ecology 68:1015–1028.

PRICE, T. D., AND P. R. GRANT. 1984. Life history traits and natural selection for small body size in a population of Darwin's finches. Evolution 38:483–494.

PRICE, T. D., P. R. GRANT, H. L. GIBBS, AND P. T. BOAG. 1984. Recurrent patterns of natural selection in a population of Darwin's finches. Nature 309:787–789.

SCHLUTER, D., AND P. R. GRANT. 1984. Determinants of morphological patterns in communities of Darwin's finches. Amer. Natur. 123:175–196.

SCHLUTER, D., T. D. PRICE, AND P. R. GRANT. 1985. Ecological character displacement in Darwin's finches. Science 227:1056–1059.

SCHLUTER, D., AND J. N. M. SMITH. 1986. Natural selection on beak and body size in the song sparrow. Evolution 40:221–231.

SILVERMAN, B. W. 1984. Spline smoothing: The equivalent variable kernel method. Ann. Statist. 12: 898–916.

SIMPSON, G. G. 1953. The Major Features of Evolution. Columbia Univ. Press, N.Y.

ULIZZI, L., AND L. TERRENATO. 1987. Natural selection associated with birth weight V. The secular relaxation of the stabilizing selection component. Ann. Hum. Genet. 51:205–210.

VAN VALEN, L. 1965. Selection in natural populations. III. Measurement and estimation. Evolution 19:514–528.

VAN VALEN, L., AND G. W. MELLIN. 1967. Selection in natural populations. VII. New York babies (fetal life study). Ann. Hum. Genet. 31:109–127.

WILLSON, M. F., AND P. W. PRICE. 1980. Resource limitation of fruit and seed production in some *Asclepias* species. Can. J. Bot. 58:2229–2233.

WRIGHT, S. J. 1932. The roles of mutation, inbreeding, crossbreeding, and selection in evolution. Proc. VI Internat. Congr. Genet. 1:356–366.

YANDELL, B. S. 1987. Algorithms for multidimensional semiparametric GLM's. Technical Report 813. Department of Statistics, Univ. Wisconsin, Madison.

Corresponding Editor: J. M. Cheverud