

Likelihood-based confidence intervals of relative fitness for a common experimental design

Steven T. Kalinowski and Mark L. Taper

Abstract: Statistical inferences concerning the relative fitness of different types of individuals in a population have not been well developed. We present a method for calculating confidence intervals for maximum likelihood estimates of relative fitness obtained from an experimental design that is common in the fisheries literature. Analysis and simulation show that these confidence limits are reliable. We also show that the bias of the estimates is low for realistic sample sizes.

Résumé : Les inférences statistiques sur la fitness relative des différents types d'individus dans une population sont encore peu élaborées. Nous présentons une méthode pour calculer les intervalles de confiance des estimations de vraisemblance maximale de la fitness relative obtenues d'un plan d'expériences couramment utilisé dans la littérature halieutique. L'analyse et la simulation montrent que ces intervalles de confiance sont fiables. Nous établissons aussi que l'erreur des estimations est faible pour des tailles d'échantillons réalistes.

[Traduit par la Rédaction]

Introduction

Measuring fitness in the wild is essential for understanding evolution and often important for managing populations. It is also notoriously difficult to do in natural populations. Molecular markers, such as microsatellites, however, offer a relatively new tool for estimating reproductive success. The method is general but has been used most often in the fisheries literature to estimate the relative fitness of different strains of anadromous fishes (e.g., Garant et al. 2003; Hendry et al. 2003; McLean et al. 2004). In this literature, studies are often conducted as follows. Adult salmon are intercepted during their upstream migration and genotyped at multiple microsatellite loci. Once sampled, the fish are released to continue their upstream migration, after which they spawn in the wild. The offspring from these fish are then captured in subsequent years and genotyped. These genotypes identify the parents of each fish, and the fitness of the parents can be estimated. This allows the fitness of different types of fish (e.g., hatchery vs. wild) to be compared (McLean et al. 2004).

These studies have brought to light limitations in the statistical methods used for inference concerning fitness. In particular, there is no report of a reliable method to produce confidence limits for estimates of relative fitnesses (e.g., Haldane 1956; Manly 1985). Wald confidence intervals based on the t distribution have been constructed by estimating the sampling variance of relative fitnesses (e.g., Manly 1972, 1985). However, the sampling distribution of estimates of relative fitness is not normal and is even asymmetric, so confidence limits estimated this way are unlikely to perform well. Hinrichsen (2003), for example, derived an expression

for the approximate sampling variance for estimates of relative fitness, but did not use this variance in his power analysis. Authors reporting the relative fitness of different stocks of fish have usually not included measures of uncertainty in their work (for examples see Hinrichsen 2003).

The purpose of this article is twofold: (1) to present a profile-likelihood method for constructing confidence limits around estimates of relative fitness, and (2) to evaluate the bias and sampling error of the maximum likelihood estimates themselves.

Estimating relative fitness

We begin by summarizing the conventional population genetic model for how natural selection affects the frequencies of different types of individuals in a population (for a more detailed presentation and examples of studies that estimate the fitness of specific genotypes see Hedrick 2000). Let k represent the number of different types of individuals in a population and p_i represent the frequency of the i th type of individual, i.e., $\sum_{i=1}^k p_i = 1.0$. By type of individual, we mean genotype, phenotype, stock, life history, morph, or any other feature that natural selection might act upon. For example, p_1 and p_2 might represent the frequency of hatchery- and wild-born salmon migrating up a river. To simplify our presentation, we use "stock" as shorthand for "type of individual".

Let p'_i represent the frequency of the i th stock after selection. This will be equal to

$$(1) \quad p'_i = \frac{\omega_i p_i}{\bar{\omega}}$$

Received 11 June 2004. Accepted 30 November 2004. Published on the NRC Research Press Web site at <http://cjfas.nrc.ca> on 18 April 2005.

J18171

S.T. Kalinowski¹ and M.L. Taper. Department of Ecology, Montana State University, Bozeman, MT 59717, USA.

¹Corresponding author (e-mail: skalinowski@montana.edu).

where ω_i denotes the fitness of the i th stock and $\bar{\omega}$ is the average fitness in the population, $\bar{\omega} = \sum_i p_i \omega_i$. Equation 1 has two notable characteristics: it can model the effects of both differential viability and fecundity and it applies to both absolute measures of fitness as well as relative fitness measures.

In many study designs, estimating the actual fitness of each stock is not possible. This is because eq. 1 depends on ratios of fitnesses and not their magnitudes. Therefore, fitness is often measured relative to a standard stock. Let w_i represent the relative fitness of the i th stock:

$$(2) \quad w_i = \frac{\omega_i}{\omega_0}$$

where ω_0 is the fitness of the stock that is being used as a standard of comparison. Substituting eq. 1 into eq. 2 gives the relationship between relative fitness and the frequencies of the stocks before and after selection:

$$(3) \quad w_i = (p'_i/p_i)/(p'_0/p_0)$$

where p_0 and p'_0 are the frequencies of the standard stock before and after selection.

Obtaining maximum likelihood estimates of relative fitness is straightforward. Let \mathbf{w} represent a vector of the relative fitnesses for the stocks in a population, i.e., $\{w_1, w_2, \dots, w_k\}$. If we assume that the frequencies of each stock before selection are known without error, the likelihood is multinomial with the probabilities for each stock given by eq. 1:

$$(4a) \quad L(\mathbf{w}) = \left(\frac{N!}{\prod_i n_i!} \right) \prod_i \left(\frac{p_i w_i}{\bar{w}} \right)^{n_i}$$

where n_i is the number of individuals from the i th stock found in the sample after selection and N is the total sample size, $N = \sum_i n_i$. By definition, the values of \mathbf{w} that maximize the right-hand half of eq. 4a are the maximum likelihood estimates of the relative fitnesses of the stocks. Expression 4a can be modified in two ways that do not alter the location of the maximum, but facilitate computation. First, the multinomial coefficient may be dropped, and second, the logarithm of each side of the equation can be taken. Therefore, maximizing the right-hand side of the following expression is sufficient to find maximum likelihood estimates of \mathbf{w} :

$$(4b) \quad \text{Ln}L(\mathbf{w}) \propto \sum_{i=1}^k n_i \ln \left(\frac{p_i w_i}{\bar{w}} \right)$$

This expression (eq. 4b) is maximized by

$$(5) \quad \hat{w}_i = \frac{\hat{p}'_i p_0}{\hat{p}'_0 p_i}$$

where \hat{p}'_i is the proportion of stock i in the samples taken after selection, $\hat{p}'_i = n_i/N$ (e.g., Manly 1985).

In some studies, combining results from different experiments may be desired. This can be done by assuming that

the relative fitness of each stock is constant. With this assumption, the log-likelihood is proportional to

$$(6) \quad \text{Ln}L(\mathbf{w}) \propto \sum_j \sum_{i=1}^k n_{i,j} \ln \left(\frac{p_{i,j} w_i}{\bar{w}_j} \right)$$

where j indexes experiments. This equation is equivalent to eq. 1 of Hinrichsen (2003). Our paper will concentrate on studies with data from one experiment because this study design has the lowest power and highest bias (Hinrichsen 2003).

Confidence limits for the estimates of relative fitness can be obtained by the hypothesis test inversion method (Casella and Berger 2001). This technique produces confidence limits by finding the set of all points in a univariate or multivariate parameter space where the likelihood ratio comparing the maximum likelihood estimates with a point in the set just equals the critical value for a test of the desired size. Let \mathbf{w}_A be a point in the confidence set. The set is defined as all points satisfying the equality:

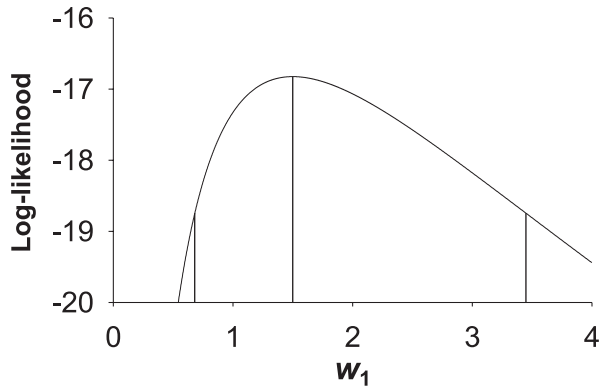
$$(7) \quad 2[\text{Ln}L(\mathbf{w}_{\text{ML}}) - \text{Ln}L(\mathbf{w}_A)] = \chi_{1,\alpha}^2$$

where $\chi_{1,\alpha}^2$ is the χ^2 value with 1 degree of freedom, and $\text{Ln}L(\mathbf{w}_{\text{ML}})$ is the log-likelihood (eq. 4b) evaluated at the maximum likelihood values of \mathbf{w} . A $1 - \alpha$ confidence interval for each of the stocks is given by the largest and smallest values of w_i in the confidence set (Pawitan 2000). Such intervals are known as profile-likelihood intervals and are generally considered superior to Wald confidence intervals constructed with parameter variance estimated on the basis of the inverse of the Fisher information matrix. The simplest approach to making joint confidence intervals that have a $1 - \alpha'$ probability of capturing all parameters (as opposed to each interval having a $1 - \alpha$ probability of capturing one parameter) is to use the Dunn-Sidak probability, $\alpha' = 1 - (1 - \alpha)^{1/k}$, in eq. 7 (Sokal and Rohlf 1995). Alternatively, a joint confidence region can be constructed using eq. 7 with $k - 1$ degrees of freedom (McCullagh and Nelder 1989; Meeker and Escobar 1995).

We illustrate the profile-likelihood method with a pair of examples, one with two stocks and one with three. In the first case, assume that there are two stocks in a river and that they both had a frequency of 0.5 before selection. A sample of 25 fish is taken after selection, and this sample contains 15 fish from stock 1 and 10 fish from stock 2 (Fig. 1 shows the likelihood profile of these data). The likelihood profile has a maximum at $\hat{w}_1 = 1.5$, which means that 1.5 is the maximum likelihood estimate of w_1 . At this point, the log-likelihood is approximately -16.83 . We search downwards from 1.5 for a value of w_1 satisfying eq. 7 that is our lower confidence limit. Searching upwards from 1.5 yields the upper confidence limit. The 95% confidence interval for \hat{w}_1 is approximately [0.68, 3.45]. Note the asymmetry in this interval.

The case of three stocks is a little more complicated, because two relative fitnesses must be estimated: w_1 and w_2 . Consider an example in which each of three stocks had a frequency of 0.3333 before selection, and that a sample of 100 fish is collected from a river after selection that contains 51 fish from stock 1, 17 fish from stock 2, and 32 fish from

Fig. 1. The log-likelihood of w_1 for a sample of 25 individuals containing 15 individuals from stock 1 and 10 individuals from stock 2 (assuming that both stocks had a frequency of 0.5 before selection).



stock 3 (the stock with which relative fitnesses are compared). The confidence set is constructed as before, but is now a closed curve in a two-dimensional parameter space. Profile confidence intervals for w_1 and w_2 are obtained by finding the smallest and largest values of w_1 and w_2 within this region.

Methods

Now that we have shown how to estimate relative fitness and how to obtain confidence intervals for these estimates, we will evaluate (i) the bias of these estimates, (ii) the variability of the estimates, and (iii) the accuracy of the confidence intervals.

Statistical properties of estimates of relative fitness

To evaluate how effectively samples estimate relative fitness, we used several metrics: bias, geometric mean bias, confidence-interval coverage, and geometric mean error. We examined two types of studies, those with two stocks (the most common application in the fisheries literature) and those with four stocks (less common). For studies with two stocks, we calculated each of the four metrics analytically. An example illustrates how we did this and why we chose the statistics listed.

Consider two stocks that have equal frequency before selection. If the relative fitness of the first stock is twice that of the second, the frequencies of the stock after selection will be 0.6667 and 0.3333 (from eq. 1). Now assume that the relative fitness of stocks is estimated by sampling 25 fish from the system after selection. If, for example, this sample consists of 15 fish from the first stock and 10 fish from the second stock, the maximum likelihood estimate of relative fitness will be 1.5 (eq. 5). If, however, the sample consists of 16 fish from stock 1 and 9 fish from stock 2, then the maximum likelihood estimate of relative fitness for stock 1 is 1.78 (which is a little closer to the parametric value of 2.0). The probability for each of these samples (and every other possible sample) can be calculated from the binomial formula. The probability of observing k fish from stock 1 is

$$(8) \quad P(n_1 = k) = \left(\frac{N!}{k!(N-k)!} \right) (0.6667)^k (0.3333)^{N-k}$$

where N is the total sample size (25 in our case). For example, there is approximately a 12.6% chance that 15 fish from stock 1 will be in observed in the sample (Table 1).

Table 1 provides all the information required to calculate the sampling statistics we used to evaluate the statistical properties of maximum likelihood estimates of relative fitness (Table 1). For example, statistical bias is the difference between the expected value of an estimate and the actual value of the parameter being estimated:

$$(9) \quad \text{bias} = E(\hat{w}_i) - w_i$$

The expected value of \hat{w}_i (directly from Table 1) is

$$(10) \quad E(\hat{w}_i) = \sum P(\hat{w}_i) \hat{w}_i$$

where summation is taken over all possible samples. For the example shown (Table 1), the expected value of \hat{w}_1 is approximately 2.3, which is about 15% too high. The calculation is approximate because there is a small chance that \hat{w}_1 will equal infinity (this occurs if no individuals from stock 2 are observed in the sample). Strictly speaking, this means that estimates of relative fitness are infinitely biased. However, the expected value of \hat{w}_1 is relatively insensitive to this problem. For example, the expected value of \hat{w}_1 is approximately 2.31 if the problem sample is not included in the calculation and 2.32 if 99 is used as an estimate when no individuals from stock 2 are included in the sample.

The bias is caused by an asymmetric sampling distribution for \hat{w}_1 (Fig. 2). The sampling distribution has a mode of approximately 2.0 (the parametric value being estimated), but has a long tail to the right (large values). The expectation of \hat{w}_1 is the arithmetic average value of \hat{w}_1 , so the tail on the right raises the expected value above 2.0. However, further examination of the example shows that the observed bias is not a problem — and is actually a by-product of a desirable property of maximum likelihood estimates of relative fitness. Consider two potential estimates of \hat{w}_1 : 1 and 4. They are equally poor. The former is half of the actual value of 2.0, while the latter is twice as large. If these estimates were equally likely, the sampling distribution for \hat{w}_1 would look something like Fig. 2a, and the arithmetic mean estimate would be greater than 2. The geometric mean, however, would be equal to 2. The geometric mean is a more appropriate measure of central tendency because relative fitnesses are used multiplicatively (and not additively) in eq. 1. In the example shown (Table 1), the geometric mean of the estimates is 2.07 (see Fig. 2b), which is substantially closer to the parametric value than the arithmetic mean of 2.3.

Table 1 provides useful information regarding other statistical properties of estimates of relative fitness (Table 1). For example, there is a 96.8% chance that the nominal 95% confidence intervals for \hat{w}_1 contain the parametric value, which is quite close to the coverage they promise.

The variability of estimates of relative fitness is also of interest. The sampling variance or standard error is usually used for this purpose, but either statistic would be misleading for the same reasons that the arithmetic mean is not an appropriate measure of central tendency. Therefore, we cal-

Table 1. Sampling distribution of estimates of relative fitness for a sample of 25 individuals taken after two stocks underwent selection.

n_1^a	n_2^b	Probability ^c	\hat{w}^d	\hat{w}_{\min}^e	\hat{w}_{\max}^f	CI successful? ^g	$\hat{w}_{\min}^?h$
0	25	1.18×10^{-12}	0.00	0.00	0.08		
1	24	5.90×10^{-11}	0.04	0.00	0.20		
2	23	1.42×10^{-9}	0.09	0.01	0.29		
3	22	2.17×10^{-8}	0.14	0.03	0.39		
4	21	2.39×10^{-7}	0.19	0.06	0.50		
5	20	2.01×10^{-6}	0.25	0.08	0.62		
6	19	1.34×10^{-5}	0.32	0.12	0.75		
7	18	7.26×10^{-5}	0.39	0.15	0.89		
8	17	3.27×10^{-4}	0.47	0.19	1.06		
9	16	0.0012	0.56	0.24	1.25		
10	15	0.0040	0.67	0.29	1.47		
11	14	0.0108	0.79	0.35	1.73		
12	13	0.0251	0.92	0.42	2.04	Yes	
13	12	0.0503	1.08	0.49	2.41	Yes	Yes
14	11	0.0862	1.27	0.58	2.87	Yes	Yes
15	10	0.1264	1.50	0.68	3.45	Yes	Yes
16	9	0.1580	1.78	0.80	4.20	Yes	Yes
17	8	0.1673	2.13	0.95	5.21	Yes	Yes
18	7	0.1487	2.57	1.12	6.61	Yes	Yes
19	6	0.1096	3.17	1.34	8.69	Yes	Yes
20	5	0.0658	4.00	1.62	12.0	Yes	Yes
21	4	0.0313	5.25	2.00	18.0	Yes	Yes
22	3	0.0114	7.33	2.54	30.9		Yes
23	2	0.0030	11.5	3.40	71.6		Yes
24	1	4.95×10^{-4}	24.0	5.08	99		Yes
25	0	3.96×10^{-5}	∞	—	—		Yes

Note: Both stocks had a frequency of 0.5 before selection. Stock 1 had a parametric relative fitness of 2.0.

^aNumber of individuals from stock 1 in the sample.

^bNumber of individuals from stock 2 in the sample.

^cProbability for the sample.

^dEstimate of relative fitness for the sample.

^eLower end of the 95% confidence interval for \hat{w} .

^fUpper end of the 95% confidence interval for \hat{w} .

^gDoes the confidence interval (CI) for \hat{w} contain the parametric value of 2.0?

^hIs \hat{w} significantly greater than 1.0?

culated the geometric mean multiplicative error (GMME) for each study design that we evaluated (see below). GMME is the average (geometric mean) factor by which estimates of relative fitness are in error. For example, a GMME of 1.25 indicates that the average estimate is in error by a factor of 1.25 (i.e., needs to be multiplied or divided by 1.25 to get the parametric fitness). GMME is defined as

$$(11) \quad \text{GMME} = \exp(E|\ln(\hat{w}_i/w_i)|)$$

where $\exp(\cdot)$ denotes the exponential function and $E(\cdot)$ denotes expectation.

The enumeration method for calculating bias and confidence-interval coverage described above can be extended to study designs with multiple stocks. However, the number of possible samples quickly becomes prohibitively large and the method becomes computationally impracticable. In such cases, computer simulation can be used to estimate bias, confidence-interval coverage, etc.

Evaluation of study designs

We examined two sets of study designs, those with two stocks, and those with five stocks. The set of study designs

we examined that had two stocks are listed in Table 2. In brief, we varied sample size from 50 to 400, varied relative fitness from 0.1 to 10, and varied initial frequencies from 0.10 to 0.90. We used the analytic method described above to analyze study designs with two stocks. Computer simulation was used to evaluate a study design with five stocks that had relative fitnesses of {1.0, 0.8, 0.9, 1.1, 1.2} for sample sizes of 50, 100, 200, and 400.

For each study design we calculated the expected value of estimates of relative fitness, geometric means of estimates, the error rate for 95% confidence intervals, and GMME.

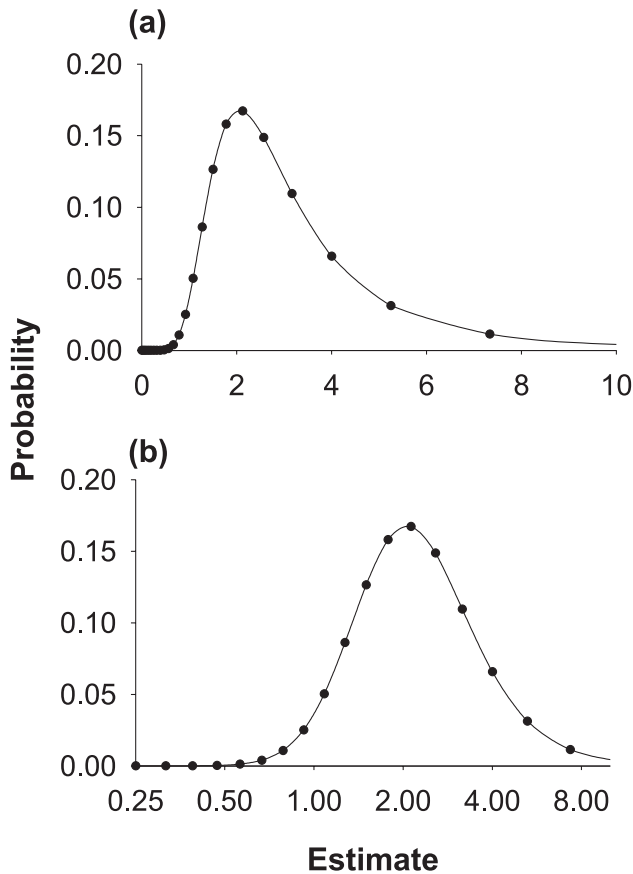
Software

A computer program, relative fitness analysis, is available to apply the statistical methods described above. The program estimates relative fitnesses (and their confidence limits) for empirical data sets and simulates data. Relative fitness analysis is available for download from www.montana.edu/kalinowski (Kalinowski 2004).

Results

Maximum likelihood estimates of relative fitness had low

Fig. 2. The sampling distribution of estimates of relative fitness for two stocks having parametric relative fitnesses of 2.0 and 1.0 when both stocks were equally abundant before selection and a sample size of 25 individuals is collected after selection. The scale is linear in *a* and logarithmic in *b*.



bias, accurate confidence intervals, and reasonable variability. The results for the study designs with two stocks are provided in Tables 3–6. Small sample sizes showed a modest amount of bias, which usually declined to a negligible amount when samples were expected to include at least 10 individuals in each of the stocks. Furthermore, the geometric mean estimate was almost always quite close to the parametric value being estimated. The 95% confidence intervals were very reliable across the entire range of sample sizes and relative fitnesses. GMME also appears acceptable. For a wide range of relative fitnesses, GMME was approximately 1.3 for a sample size of 50 — and lower for larger samples.

Estimates of relative fitness for the study designs with five stocks also had low bias, accurate confidence intervals, etc. (results not shown).

Discussion

This work shows that maximum likelihood estimates of relative fitness are relatively unbiased, have reasonable sampling error, and have nearly true sized confidence limits — even for sample sizes as low as 10 individuals for the least abundant stock. This is reassuring, for it indicates that small samples can be informative. However, this does not mean that small samples will be sufficient to precisely estimate

Table 2. Expected numbers of stock 1 in the two-stock study designs examined in this study as a function of the initial frequencies of two stocks (*p*), the relative fitness of stock 1 (*w*₁), and the number of fish sampled after selection (*N*).

<i>p</i> ₁ : <i>p</i> ₂	<i>w</i> ₁	<i>N</i>			
		50	100	200	400
0.5 : 0.5	0.10	5*	9*	18	36
	0.50	17	33	67	133
	0.80				
	1	25	50	100	200
	1.2	27	54	109	218
	2	33	67	133	267
0.25 : 0.75	10	45	91	181	363
	0.80	11	21	42	84
	1	13	25	50	100
	1.2	14	28	57	114
0.75 : 0.25	0.80	35	70	141	282
	1	38	75	150	300
	1.2	39	79	156	313
0.10 : 0.90	0.80	4*	8*	16	32
	1	5*	10*	20	40
	1.2	6*	12	23	47
0.90 : 0.10	0.80	44*	87	175	351
	1	45*	90	180	360
	1.2	46*	92*	183	366

*The expected number of individuals in the sample for one of the stocks is 10.0 or fewer.

relative fitnesses. Larger sample sizes give increased power and tighter confidence intervals.

The profile confidence intervals that we describe above are expected to perform better than the more traditional Wald intervals. Wald intervals are constructed by assuming that the likelihood is distributed normally with a variance/covariance matrix approximated by the inverse of the Fisher information matrix. Confidence intervals can then be constructed using *t*, *z*, or χ^2 distributions. At large sample sizes such intervals will be equivalent to profile-likelihood intervals, but at smaller sample sizes the estimated likelihood will generally not be normally distributed, and profile-likelihood intervals will be superior (Kalbfleisch and Sprott 1970; Venzon and Moolgavkar 1988; for reviews see Meeker and Escobar 1995; Pawitan 2000). Our investigations (result not reported) have shown this trend — the coverage properties of the Wald interval for relative fitness generally were poorer than the profile interval. Undesirable properties of Wald intervals included high error rates, bias, and intervals that extended below zero. In the past, Wald intervals have been reasonably justified on the basis of their greater ease of computation. With current computation power, this issue is much less relevant than before.

The model (eq. 1) for selection that we have examined is the most basic available for studying evolution (e.g., Hedrick 2000). Its advantage is simplicity and insight into how two groups compare. One disadvantage is that it usually will not indicate which genetic or physical traits contribute to fitness. If selection gradients are of interest they can be estimated by fitting a curve to a plot of fitness versus phenotype for the individuals sampled (e.g., Smouse et al. 1999; Morgan and Conner 2001; Hendry et al. 2003).

Table 3. Expected values of \hat{w}_i as a function of the initial frequencies of two stocks (p), the relative fitness of stock 1 (w_1), and the number of fish sampled after selection (N).

$p_1 : p_2$	w_1	N			
		50	100	200	400
0.5 : 0.5	0.10	0.10*	0.10*	0.10	0.10
	0.50	0.52	0.51	0.50	0.50
	0.80	0.83	0.81	0.81	0.81
	1	1.04	1.02	1.01	1.00
	1.2	1.26	1.23	1.21	1.21
	2	2.13	2.06	2.03	2.02
0.25 : 0.75	10	13.0	11.4	10.6	10.3
	0.80	0.82	0.81	0.81	0.80
	1	1.02	1.01	1.01	1.00
0.75 : 0.25	1.2	1.23	1.22	1.21	1.20
	0.80	0.86	0.83	0.81	0.81
	1	1.09	1.04	1.02	1.01
0.10 : 0.90	1.2	1.33	1.26	1.23	1.21
	0.80	0.82*	0.81*	0.80	0.80
	1	1.02*	1.01*	1.00	1.00
0.90 : 0.10	1.2	1.22*	1.21	1.21	1.20
	0.80	0.99*	0.88	0.84	0.82
	1	1.28*	1.13	1.06	1.03
	1.2	1.57*	1.39*	1.28	1.24

*The expected number of individuals in the sample for one of the stocks is 10.0 or fewer.

Table 4. Geometric mean values of \hat{w}_i as a function of the initial frequencies of two stocks (p), the relative fitness of stock 1 (w_1), and the number of fish sampled after selection (N).

$p_1 : p_2$	$w_1 : w_2$	N			
		50	100	200	400
0.5 : 0.5	0.10	0.09*	0.09*	0.10	0.10
	0.50	0.49	0.50	0.50	0.50
	0.80	0.80	0.80	0.80	0.80
	1	1.00	1.00	1.00	1.00
	1.2	1.20	1.20	1.20	1.20
	2	2.03	2.02	2.01	2.00
0.25 : 0.75	10	10.8	10.6	10.3	10.1
	0.80	0.77	0.79	0.79	0.80
	1	0.97	0.99	0.99	1.00
0.75 : 0.25	1.2	1.17	1.19	1.19	1.19
	0.80	0.82	0.81	0.80	0.80
	1	1.03	1.01	1.01	1.00
0.10 : 0.90	1.2	1.24	1.22	1.21	1.21
	0.80	0.72*	0.75*	0.78	0.79
	1	0.91*	0.95*	0.98	0.99
0.90 : 0.10	1.2	1.10*	1.15	1.18	1.19
	0.80	0.87*	0.83	0.81	0.80
	1	1.10*	1.05	1.02	1.01
	1.2	1.32*	1.27*	1.23	1.21

*The expected number of individuals in the sample for one of the stocks is 10.0 or fewer.

Table 5. Error rates of 95% confidence intervals for \hat{w}_i as a function of the initial frequencies of two stocks (p), the relative fitness of stock 1 (w_1), and the number of fish sampled after selection (N).

$p_1 : p_2$	$w_1 : w_2$	N			
		50	100	200	400
0.5 : 0.5	0.10	0.06	0.06	0.05	0.05
	0.50	0.05	0.06	0.05	0.05
	0.80	0.05	0.04	0.05	0.05
	1	0.06	0.06	0.06	0.05
	1.2	0.05	0.04	0.05	0.05
	2	0.05	0.06	0.05	0.05
0.25 : 0.75	10	0.05	0.06	0.05	0.05
	0.80	0.06	0.05	0.05	0.05
	1	0.05	0.05	0.05	0.05
0.75 : 0.25	1.2	0.06	0.06	0.05	0.05
	0.80	0.04	0.05	0.05	0.05
	1	0.05	0.05	0.05	0.05
0.10 : 0.90	1.2	0.05	0.05	0.05	0.05
	0.80	0.02*	0.05*	0.05	0.05
	1	0.05*	0.04*	0.05	0.05
0.90 : 0.10	1.2	0.05*	0.04	0.06	0.04
	0.80	0.06*	0.05	0.05	0.05
	1	0.05*	0.04	0.05	0.05
	1.2	0.02*	0.05*	0.04	0.05

*The expected number of individuals in the sample for one of the stocks is 10.0 or fewer.

Table 6. Geometric mean multiplicative error (GMME) of estimates of \hat{w}_i as a function of the initial frequencies of two stocks (p), the relative fitness of stock 1 (w_1), and the number of fish sampled after selection (N).

$p_1 : p_2$	$w_1 : w_2$	N			
		50	100	200	400
0.5 : 0.5	0.10	1.52	1.34	1.22	1.15
	0.50	1.27	1.19	1.13	1.08
	0.80	1.26	1.17	1.12	1.08
	1	1.26	1.17	1.12	1.08
	1.2	1.26	1.18	1.12	1.08
	2	1.27	1.18	1.13	1.09
0.25 : 0.75	10	1.52	1.34	1.22	1.15
	0.80	1.33	1.22	1.15	1.10
	1	1.31	1.20	1.14	1.10
0.75 : 0.25	1.2	1.29	1.20	1.13	1.09
	0.80	1.29	1.19	1.13	1.09
	1	1.31	1.20	1.13	1.10
0.10 : 0.90	1.2	1.33	1.22	1.15	1.10
	0.80	1.53*	1.36*	1.24	1.16
	1	1.49*	1.32*	1.21	1.14
0.90 : 0.10	1.2	1.46*	1.29	1.19	1.13
	0.80	1.44*	1.28	1.19	1.13
	1	1.49*	1.32	1.21	1.14
	1.2	1.54*	1.36*	1.23	1.16

*The expected number of individuals in the sample for one of the stocks is 10.0 or fewer.

The favorable results presented here support the development of maximum likelihood estimates of relative fitness (and confidence intervals for those estimates) for more complex or realistic study designs. For example, the study designs reviewed here assumed that the initial proportions of each stock were known at the beginning of the experiment. This may not always be the case. Manly (1972) has provided formulae for maximum likelihood estimates of relative fitness, along with estimates of the standard error of the estimates, but the confidence intervals are likely to be inaccurate. Alternatively, the method could be extended to account for uncertainty in stock identification.

Acknowledgements

We thank Mike Ford and an anonymous reviewer for suggestions that improved the manuscript. MLT was supported by NSF grant DEB-0415932.

References

- Casella, G., and Berger, R.L. 2001. *Statistical inference*. 2nd ed. Brooks Cole, Pacific Grove, Calif.
- Garant, D., Dodson, J.J., and Bernatchez, L. 2003. Differential reproductive success and heritability of alternative reproductive tactics in wild Atlantic salmon (*Salmo salar* L.). *Evolution*, **57**: 1133–1141.
- Haldane, J.B.S. 1956. The estimation and significance of the logarithm of a ratio of frequencies. *Ann. Hum. Genet.* **20**: 309–311.
- Hedrick, P.W. 2000. *Genetics of populations*. Jones and Bartlett, Boston.
- Hendry, P.A., Letcher, B.H., and Gries, G. 2003. Estimating natural selection acting on stream dwelling Atlantic salmon: implications for the restoration of extirpated populations. *Conserv. Biol.* **17**: 795–805.
- Hinrichsen, R.A. 2003. The power of experiments for estimating relative reproductive success of hatchery-born spawners. *Can. J. Fish. Aquat. Sci.* **60**: 864–872.
- Kalbfleisch, J.D., and Sprott, D.A. 1970. Application of likelihood methods to models involving large numbers of parameters. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **32**: 175–208.
- Kalinowski, S.T. 2004. Relative fitness analysis (1.0). Available at www.montana.edu/kalinowski.
- Manly, B.F.J. 1972. Estimating selective values from field data. *Biometrics*, **28**: 1115–1125.
- Manly, B.F.J. 1985. *The statistics of natural selection on animal populations*. Chapman and Hall, New York.
- McCullagh, P., and Nelder, J.A. 1989. *Generalized linear models*. 2nd ed. Chapman and Hall, New York.
- McLean, J.E., Bentzen, P., and Quinn, T.P. 2004. Differential reproductive success of sympatric, naturally spawning hatchery and wild steelhead, *Oncorhynchus mykiss*. *Environ. Biol. Fishes*, **69**: 359–369.
- Meeker, W.Q., and Escobar, L.A. 1995. Teaching about approximate confidence-regions based on maximum-likelihood estimation. *Am. Stat.* **49**: 48–53.
- Morgan, M.T., and Conner, J.K. 2001. Using genetic markers to directly estimate male selection gradients. *Evolution*, **55**: 272–281.
- Pawitan, Y. 2000. A reminder of the fallibility of the Wald statistic: likelihood explanation. *Am. Stat.* **54**: 54–56.
- Smouse, P.E., Meagher, T.R., and Kobak, C.J. 1999. Parentage analysis in *Chamaelirium luteum* (L.) Gray (Liliaceae): why do some males have higher reproductive contributions? *J. Evol. Biol.* **12**: 1069–1077.
- Sokal, R.R., and Rohlf, F.J. 1997. *Biometry*. 3rd ed. W.H. Freeman and Co., New York.
- Venzon, D.J., and Moolgavkar, S.H. 1988. A method for computing profile-likelihood-based confidence-intervals. *J. R. Stat. Soc. Ser. C Appl. Stat.* **37**: 87–94.