New Zealand Fisheries Assessment Research Document 90/4

A maximum likelihood stock reduction method

R. I. C. C. Francis

MAF Fisheries Greta Point
P O Box 297
Wellington

August 1990

MAF Fisheries, N.Z. Ministry of Agriculture and Fisheries

This series documents the scientific basis for stock assessments and fisheries management advice in New Zealand. It addresses the issues of the day in the current legislative context and in the time frames required. The documents it contains are not intended as definitive statements on the subjects addressed but rather as progress reports on ongoing investigations.

INTRODUCTION

Stock reduction methods are designed to estimate the biomass of a fish population using

- a catch history,
- a series of biomass indices (e.g., trawl survey relative abundance estimates, or catch per unit effort indices), and
- information on the productivity of the species.

Also estimated is the 'catchability', q, which is the constant of proportionality between the biomass indices and absolute biomass (i.e. if O is a biomass index, then O/q is an estimate of absolute biomass).

Sissenwine (1988) presented a quick and simple stock reduction method which requires minimal information on productivity. His method assumes that the stock productivity (expressed as an instantaneous rate, P) is constant over the period of the biomass indices. Generally the value of P used with this method would be 'guestimated' using some rule of thumb relating it to the natural mortality of the stock (with, possibly, some modification according to inferences based on observed changes in fish size in the population).

The method presented here differs from Sissenwine's in two major ways. First, productivity is not assumed constant but is (implicitly) calculated separately for each year. Second, the present method is statistically based and so allows the estimation of a probability distribution (and thus, confidence limits) for the estimated biomass. (A third apparent difference is that the current method produces biomass estimates for the whole history of the fishery, whereas Sissenwine's method, as published, estimates only the initial and final biomasses. Mike Sissenwine assures me, however, that his method may be adapted to estimate biomass at intermediate years).

This method was devised to solve a specific stock reduction problem (Francis & Robertson 1990) and its formulation (and the title of this paper) reflects that origin. A referee (Mike Sissenwine) has pointed out that the formulation is unnecessarily restrictive, and that the approach could easily be extended to solve the more general problem of estimating population parameters from a series of biomass indices. That is, it need not be restricted to the 'stock reduction problem' in which the stock is reduced from a virgin state. Such extensions will be considered in a subsequent publication.

ASSUMPTIONS AND DATA REQUIREMENTS

We assume that the dynamics of the population are well described by an age structured model (like the biological part of the model described by Mace & Doonan, 1988). This requires the following life history parameters:

- natural mortality, M,
- von Bertalanffy growth parameters, L_{inf}, k, t_0,
- length–weight relationship parameters, a, b,
- the 'steepness', Δ, for the Beverton and Holt stock–recruitment relationship (Mace & Doonan, 1988), and
- ages at maturity and recruitment, A_r, A_m.

The other data required are:

- a complete catch history for the fishery,
- biomass indices, O_i, for at least three times in the history of the fishery.

We assume that each biomass index is normally distributed with expected value equal to q times the true biomass, and that all indices have the same coefficient of variation, c. For trawl survey indices q is the product of the vulnerability, and the areal and vertical availabilities (in the terminology of Francis, 1989). (Further implications of these assumptions, and some ways of varying them, are discussed below).

The parameters estimated are q, c, and B_0. From B_0 the age structured model may be used to calculate biomass estimates for the entire history of the fishery.

If an independent estimate of the catchability, q, is available the procedure may be modified to assume this value and estimate just c and B_0.

THE MAXIMUM LIKELIHOOD APPROACH TO ESTIMATION

For the purpose of estimating q, c, and B_0 the data are treated in two different ways. First, we assume that the life history parameters and catch history are known exactly. (The effect of uncertainty in these parameters is best incorporated after the stock reduction analysis in a sensitivity analysis). Second, we assume that the biomass indices are observed with error. That is, they are random variables – if replicate observations had been made at the same time, different indices would have been obtained.

The assumptions in the previous section allow us to calculate, for given values of q, c, and B_0, what the likelihood is that the observed biomass indices would occur. The principle of maximum likelihood estimation is simply (as the name applies) that the best estimates of q, c, and B_0 are those that make the observed biomass indices most likely.

ESTIMATION FORMULAE

For each biomass index O_i, let E_i be the true population biomass at the time O_i was estimated. Note that, given the catch history and the above life history parameters, the E_i may be calculated using the age structured model and depend only on B_0. The above assumptions imply that O_i is normally distributed with mean qE_i and standard deviation cqE_i, for all i.

Under these assumptions the log–likelihood of observing the biomass indices, O_i, is given by:

$$\lambda = -0.5n\log(2\pi) - n\log(c) - n\log(q) - \Sigma_i\log(E_i)$$
$$- 0.5c^{-2}\, \Sigma_i(O_i/(qE_i)-1)^2$$

where n is the number of biomass indices. (For given q, c, and B_0, the bigger λ is, the more likely it is that the observed O_i would occur.)

Using standard mathematical methods for maximisation (see Appendix) it is straightforward to show that the maximum likelihood estimates, $\hat{q}$ and $\hat{c}$, are given by

$$\hat{q} = (1/n)\Sigma_i(O_i/E_i), \text{ and} \qquad (1)$$

$$\hat{c}^2 = (1/n)\Sigma_i(O_i/(\hat{q}E_i)-1)^2, \qquad (2)$$

and that the maximum likelihood estimate, $\hat{B}_0$, is that which maximizes

$$\lambda = -\, n\log(\hat{c}) - n\log(\hat{q}) - \Sigma_i \log(E_i). \quad (3)$$

If an independent estimate of q is available then equation (1) is ignored and this value of q is used in place of $\hat{q}$ in (2) and (3).

To obtain an unbiassed estimate of c, the n in (2) should be replaced by (n–2). [This is analagous to the estimation of a sample variance: the maximum likelihood estimate (which is biassed) is:

$$[\Sigma_i(X_i - \bar{X})^2]/N$$

but the standard estimator is

$$[\Sigma_i(X_i - \bar{X})^2]/(N-1),$$

and this is unbiassed.]

THE ESTIMATION ALGORITHM

The method proceeds according to the following steps:

1. Choose a trial value of B_0.
2. Given B_0, calculate the biomass values, E_i, corresponding to each biomass index, O_i, using the catch history and the age–structured model.
3. Calculate (using equation (1) above) the best value of the catchability, q, for this B_0; that is, the value of q that gives the best match between the calculated biomass values (from step 2) and the sequence of biomass indices divided by q.
4. Calculate (using (2) and (3)) the log–likelihood, λ, for these values of B_0 and q. (Remember, the bigger λ is, the better the fit is.)
5. Repeat steps 1 to 4 with a range of trial B_0 values to find the value of B_0 that maximizes.

An example will illustrate how the method is applied.

AN EXAMPLE

Consider a fish stock with the following life history parameters:

M	=	0.1	a	=	8.55 x 10–3
L_{inf}	=	100	b	=	2.88
k	=	0.2	Δ	=	0.95
t_0	=	0	A_r	=	A_m = 4

and the following catch history:

Year	1	2	3	4	5	6	7	8
Catch	200	1000	2000	2200	1000	2500	2800	3500

Suppose trawl surveys were carried out in the middle of the fishing season in years 4, 6, 7, and 8, with results

Year	4	6	7	8
Biomass index, O_i,	6467	3471	4290	2654

The results of step 2, the calculation of annual biomass values from B_0, are illustrated (for three different trial values of B_0) in Fig. 1. (Since the trawl surveys were carried out in the middle of the fishing season, mid-season biomasses are calculated in the age structured model.)

For B_0 = 20000 t the calculated mid-season biomasses (for the trawl survey years) are

Year	4	6	7	8
Calculated biomass, E_i,	15802	12849	10498	7670

Fig. 2 shows the effect of q on the comparison of the calculated mid-season biomass values, E_i, with the estimated values, O_i/q. It is clear from this figure that the best value of q for B_0 = 20000 t is between 0.3 and 0.4. In step 3 we use equation (1) to calculate

$$\hat{q} = (6467/15802 + 3471/12849 + 4290/10498 + 2654/7670)/4$$
$$= 0.359$$

For step 4 we use equations (2) and (3) to calculate

$$\hat{c}^2 = [\ (6467/(0.359*15802)-1)^2 + (3471/(0.359*12849)-1)^2 + (4290/(0.359*10498)-1)^2 + (2654/(0.359*7670)-1)^2]/4 = .0253$$

so $\hat{c} = 0.159$,

and $\lambda = -4\log(0.159) - 4\log(0.359) - (\log(15802) + \log(12849) + \log(10498) + \log(7670)) = -25.88$

Finally, to ilustrate step 5, Fig. 3 shows the relationship between the trial B_0 values and the log-likelihood. The best estimate of B_0 is 19100 t. For this value of B_0 the calculated biomass values for the trawl survey years are

Year	4	6	7	8
Calculated biomass, E_i,	14900	11945	9587	6739

From these and equations (1) and (2) we can calculate $\hat{q}$ = 0.391, $\hat{c}$ = 0.157. An unbiassed estimate of c is 0.222.

CONFIDENCE INTERVAL FOR B_0

Suppose B_{0est} is the maximum likelihood estimate derived above. A series of simulations may be used to estimate a confidence interval for B_{0est}, given a value of c. What these simulations achieve is to answer the question, for each of a series of trial B_0 values, 'If this were the true B_0, how likely is it that we would get an estimate of virgin biomass > B_{0est}?'.

The procedure is as follows:

1. Pick a trial value of B_0.
2. Calculate the biomasses, E_i, using the age structured model with this value of B_0.
3. Generate simulated biomass indices, O_i, (assuming O_i is normal with mean E_i and coefficient of variation c).
4. Calculate the maximum likelihood estimate B_0 for the simulated O_i.
5. Repeat steps 3 and 4 one hundred times and calculate the proportion, p, of the B_0 that are > B_{0est}.
6. Repeat steps 1 to 6 for a range of trial B_0 values.
7. Graph p against the trial B_0 values and draw a smooth line through the points.
8. The 95% confidence interval for B_{0est} is the range of trial B_0 values on the graph for which $0.025 < p < 0.975$.

Fig. 4 shows the graph derived by this procedure using the data from the above example.

MORE ABOUT THE ASSUMPTIONS

A major assumption in stock reduction methods is that the catchability, q, is the same for all the biomass indices. This assumption would be invalid if, for example, the biomass indices were average catch per boat–day and the boats used had significantly increased in fishing power over the period of the indices. For trawl survey indices the assumption would be invalid if there was a trend over time in areal availability (or vulnerability, or vertical availability).

The assumption that the biomass indices all have the same coefficient of variation, c, is felt to be the most plausible of three alternative assumptions. Because the biomass indices used in stock reduction analyses tend to cover a great range in size the simplest assumption – that all the biomass indices have the same variance – is considered untenable. (If there is not much variation in the size of the indices then there is little difference between assuming constant variance or constant coefficient of variation). The third alternative – using estimated variances – was not used because these are not generally available for the catch per unit effort indices and the variance estimates for the trawl survey indices are not usually sufficiently reliable.

If trawl survey indices are used, and there is a substantial between survey variation in the number of trawl stations, then it would be reasonable to give more weight to surveys with more stations. This may be done by assuming that the coefficient of variation of O_i is, say,

$$(100/m_i)^{0.5}c ,$$

where m_i is the number of stations in the ith survey. With this assumption c becomes the coefficient of variation of a 'standard' trawl survey using 100 stations, and the variance of survey indices is assumed to be inversely proportional to the number of stations. The modified formulae are

$$\lambda = -0.5n\log(2\pi) - n\log(c) - n\log(q) - \Sigma_i \log(E_i/W_i)$$

$$- 0.5c^{-2}\, \Sigma_i[W_i(O_i/(qE_i)-1)]^2 \;,$$

$$\hat{q} = \Sigma_i[W_i^2(O_i/E_i)]/(\Sigma_i W_i^2), \qquad (1')$$

$$\hat{c}^2 = (1/n)\Sigma_i[W_i(O_i/(\hat{q}E_i)-1)]^2, \text{ and} \qquad (2')$$

$$\lambda = -\, n\log(\hat{c}) - n\log(\hat{q}) - \Sigma_i \log(E_i/W_i), \qquad (3')$$

where the W_i are the weights $(m_i/100)^{0.5}$.

REFERENCES

Francis, R. I. C. C. 1989: A standard approach to biomass estimation from bottom trawl surveys. New Zealand Fisheries Assessment Research Document 89/3.

Francis, R. I. C. C. and Robertson, D. A. 1990: Assessment of the Chatham Rise (QMA 3B) orange roughy fishery for the 1989/90 and 90/91 seasons. New Zealand Fisheries Assessment Research Document 90/3.

Mace, P. M. and Doonan, I. J. 1988: A generalised simulation model for fish population dynamics. New Zealand Fisheries Assessment Research Document 88/4.

Sissenwine, M. P. 1988: A quick method of estimating absolute biomass from a time series of relative biomass and catch data. New Zealand Fisheries Assessment Research Document 88/3.

APPENDIX: Derivation of Maximum Likelihood Estimators

It is desired to maximise the log–likelihood, λ, given by

$$\lambda = -0.5n\log(2\pi) - n\log(c) - n\log(q) - \Sigma_i\log(E_i)$$

$$- 0.5c^{-2}\Sigma_i(O_i/(qE_i)-1)^2 \quad \text{(A1)}$$

Differentiating by c and q respectively, we get

$$\delta\lambda/\delta c = -n/c + c^{-3}\Sigma_i(O_i/(qE_i)-1)^2 \text{ and}$$

$$\delta\lambda/\delta q = -n/q -(cq)^{-2}\Sigma_i(O_i/(qE_i)-1)(O_i/E_i) \ .$$

The maximum likelihood estimates $\hat{c}$ and $\hat{q}$ occur when these derivatives are set equal to zero, i.e.,

$$\hat{c}^2 = (1/n)\Sigma_i(O_i/(\hat{q}E_i)-1)^2 \qquad \text{(A2)} \qquad \text{and}$$

$$n\hat{c}^2\hat{q} = \Sigma_i(O_i/(\hat{q}E_i)-1)(O_i/E_i) \ . \qquad \text{(A3)}$$

Substituting for $\hat{c}^2$ from (A2) in (A3) we get

$$\hat{q}\Sigma_i(O_i/(\hat{q}E_i)-1)^2 = \Sigma_i(O_i/(\hat{q}E_i)-1)(O_i/E_i) \ ,$$

which, when expanded gives

$$(1/\hat{q})\Sigma_i(O_i^2/E_i^2) -2\Sigma_i(O_i/E_i) + \hat{q}n = (1/\hat{q})\Sigma_i(O_i^2/E_i^2) - \Sigma_i(O_i/E_i)$$

so

$$\hat{q} = (1/n)\Sigma_i(O_i/E_i) \quad \text{(A4).}$$

Substituting for $\hat{c}$ (from (A2)) and for $\hat{q}$ (from (A4)) in (A1)

$$\lambda = -0.5n\log(2\pi) - n\log(\hat{c}) - n\log(\hat{q}) - \Sigma_i\log(E_i) - n/2$$

so, ignoring constant terms, the maximum likelihood estimate, $\hat{B}_0$, is that which maximises

$$\lambda = - n\log(\hat{c}) - n\log(\hat{q}) - \Sigma_i\log(E_i)$$

where $\hat{c}$ and $\hat{q}$ are as given by (A2) and (A4).

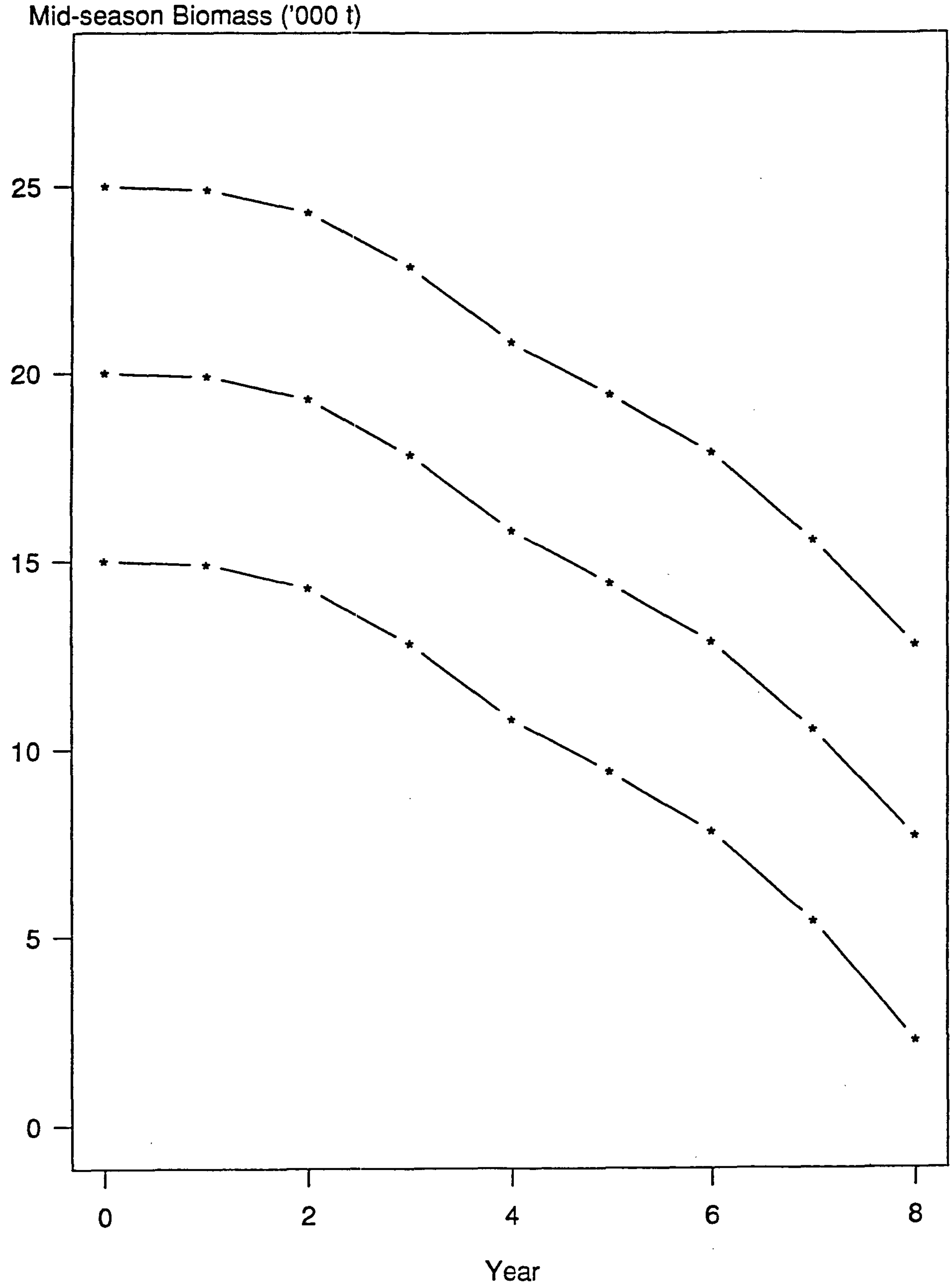

Fig. 1. Mid–season biomass estimates, E_i, calculated for three trial values of B_0: 15000, 20000, and 25000 t.

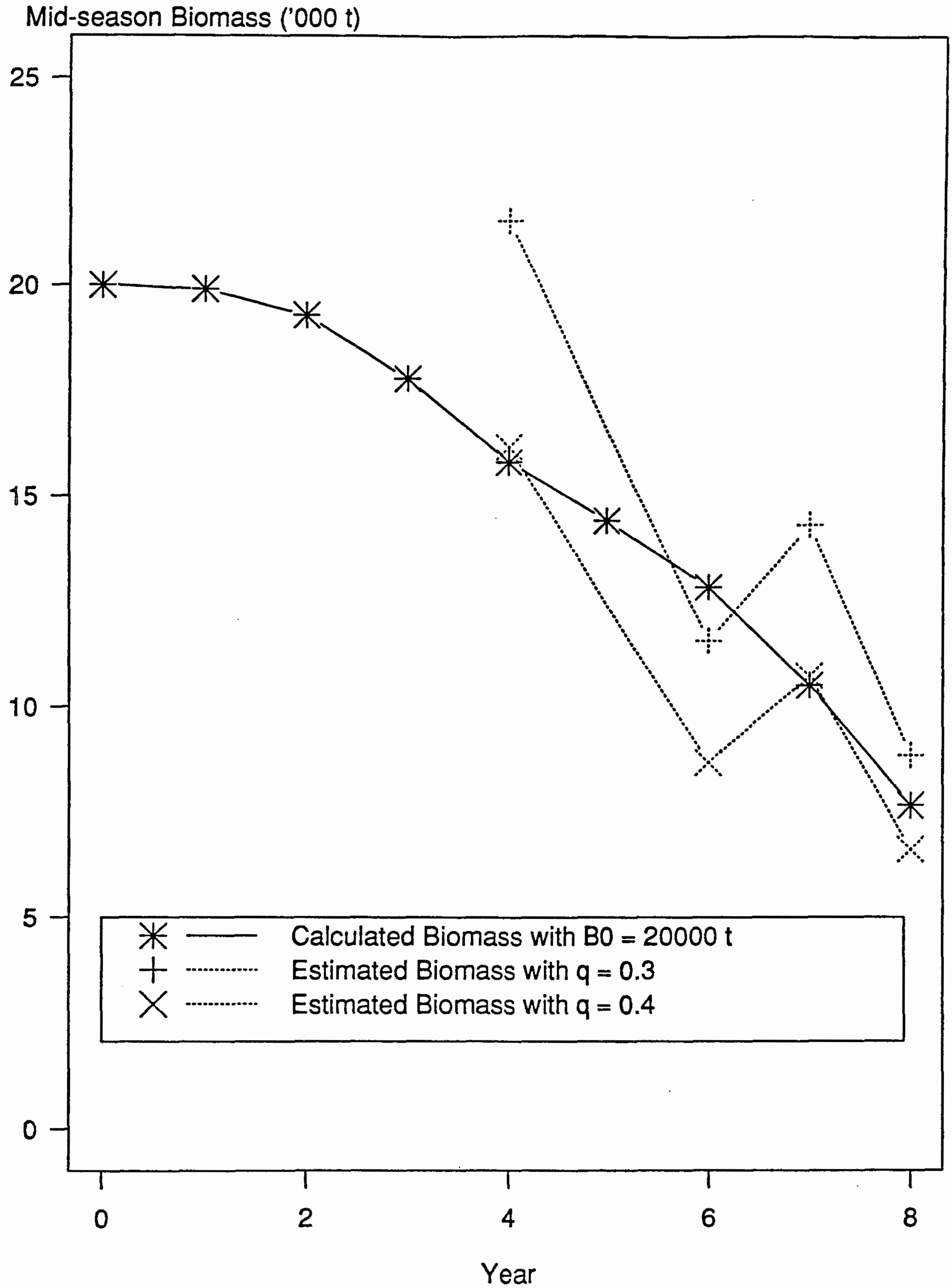

Fig. 2. Comparison between calculated biomass, E_i, and estimated biomass, O_i/q, for trial B_0 = 20000 t and q = 0.3 and 0.4 .

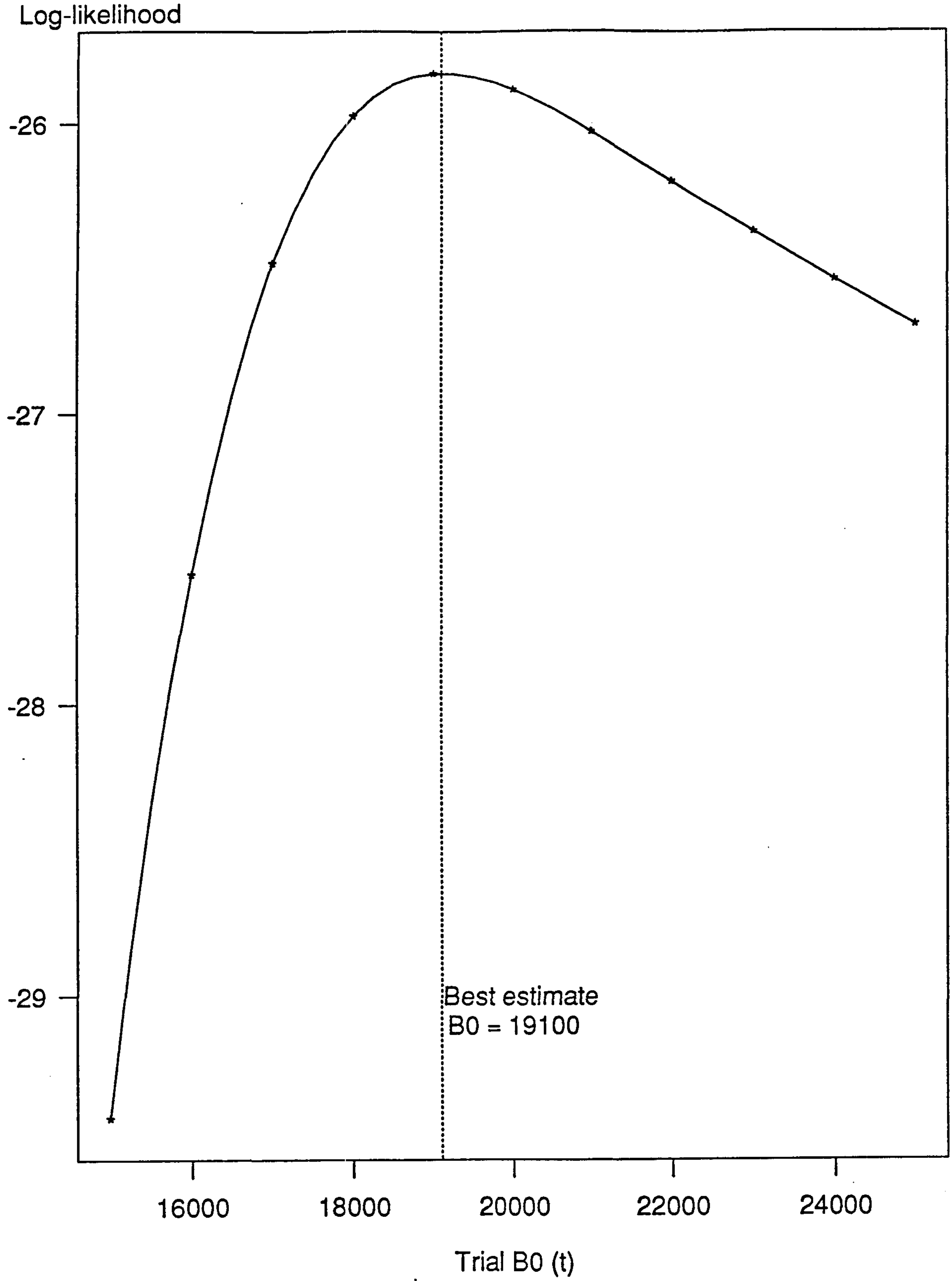

Fig. 3. The relationship between log–likelihood, λ, and trial values of B_0. The best estimate of B_0 is that which maximizes the log–likelihood.

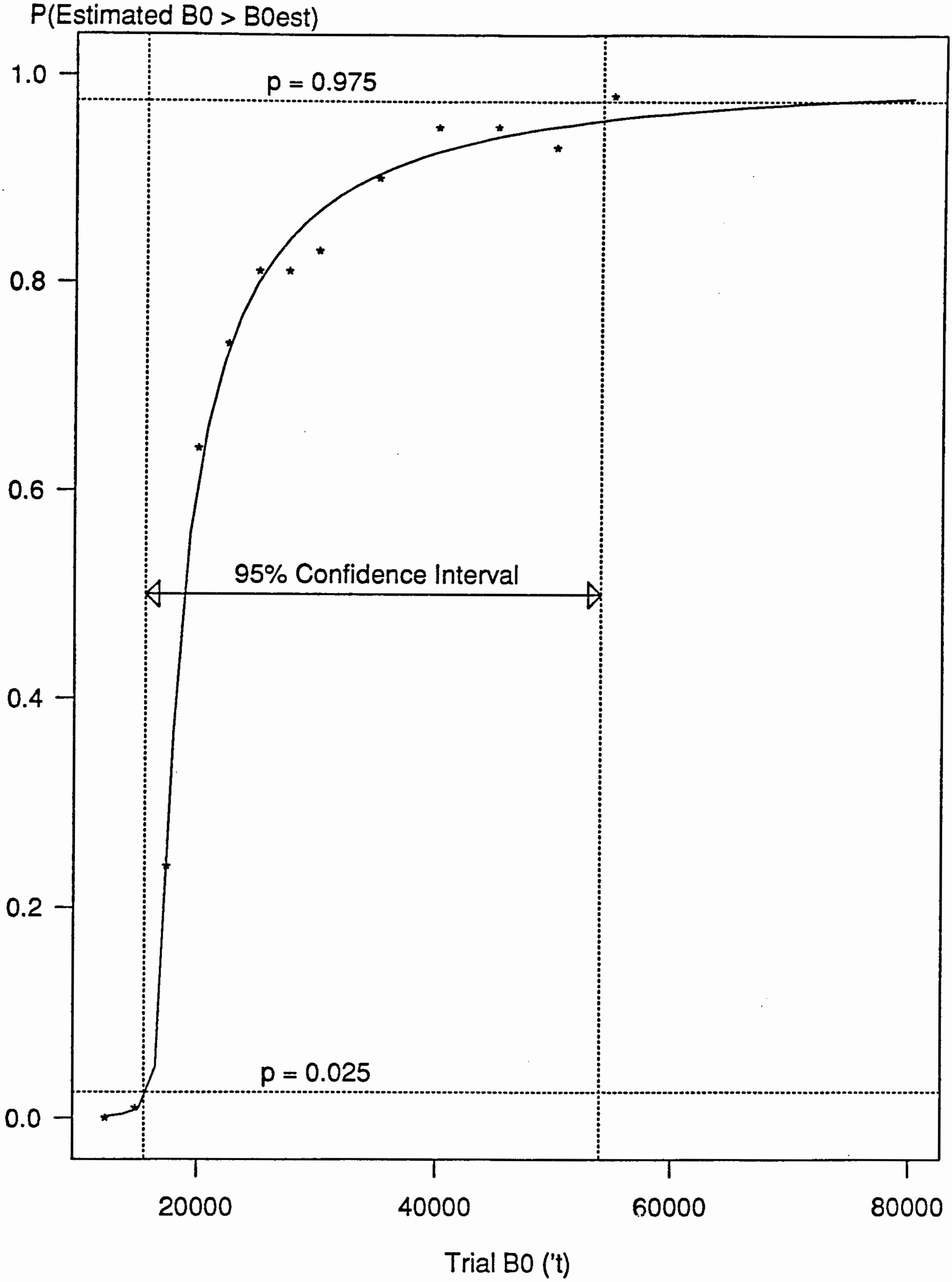

Fig. 4. Illustration of the calculation of a 95% confidence interval for B_0 using the data from the example.