

STATISTICAL PROBLEMS IN DYNAMICS OF EXPLOITED FISHERIES POPULATIONS

D. G. CHAPMAN
UNIVERSITY OF WASHINGTON

1. Introduction

The many postwar conferences on allocation of fishing rights on the high seas between various nations and on the conservation of marine resources have focused attention on the dynamics of marine populations and particularly those exploited by man. This interest has brought forth numerous models of various degrees of sophistication pertaining to the dynamics of such populations. The forerunner of these perhaps was the model of Baranov [2]; the most extensive modern study is that due to Beverton and Holt [5]. The emphasis of all of these studies has been in determining the optimum level of exploitation by man, that is, to determine what is the maximum sustainable catch. Since catch is the result of fishing effort, initial population size, growth rate, natural mortality, and certain other parameters, the over-all structure can be examined only if there is information to estimate such parameters. This information comes partly from samples obtained or experiments made by the research biologists and partly from data obtained in connection with the exploitation. In fact, in general it is feasible to study a large population only if it is exploited. Consequently a statistical study of the dynamics of a large population is necessarily based on such information, that is, catch data, effort data, returns of tags from the fishery, and so on.

Such data may be used to estimate the parameters of the various processes that go to make up the total yield and the changes in these parameters in response to manipulations of man. That is, one can attempt to study the structure of the several processes in the small. Alternatively it is possible to consider only the end result of the total catch and try to relate this to changes in effort or other manipulations of man.

Beverton and Holt in the paper cited used the first approach and studied a number of models in detail; however, it was not their aim to give a complete statistical treatment. They have used a number of special methods; some methods suggested have been used uncritically. Often no formulas have been given for variances of the estimates nor confidence intervals for the parameters. It is the aim of this paper, therefore, to formulate some of the statistical problems associated with the dynamics of exploited fish populations, to provide solutions to some of them, and to note problems which are still open.

Almost invariably the statistical problems have been formulated in terms of deterministic models. Where the studies treated large populations far from any "boundaries" it may have been reasonable to assume that the differences between deterministic and stochastic models are second-order effects. However, the formulation here, except for the growth process, will be as a stochastic process. It is also unsatisfactory to study a population as if it were isolated from the other species in its environment; usually, we know too little of the prey and predator species to treat them specifically and can only consider them indirectly through their effects on the exploited population.

2. Mortality rates

The deterministic theory of fishing is based on the equation, for yield in weight,

$$(2.1) \quad \frac{dY_w}{dt} = qf(t)N(t)w(t),$$

where $Y_w(t)$ = catch (by weight) of the fishery in time $(0, t)$,

$f(t)$ = effort applied at time t ,

$N(t)$ = population size in numbers at time t ,

$w(t)$ = average weight of fish at time t ,

and q is a parameter usually defined as the fraction of the population taken by one unit of effort. In probability terms it can be defined as the probability that a unit of effort captures any specified fish. While we do not use the deterministic equation (2.1) or its integrated form, we merely point out that a study of yield certainly depends on the processes $N(t)$, $w(t)$, the parameter q , and the initial conditions. We first study $N(t)$.

We assume that the population has initial size N_0 and is subject to the two death processes with instantaneous transition probabilities M , and $qf(t)$, that is, the differential equation for the probability of n individuals surviving at time t is

$$(2.2) \quad P'_N(t) = [M + f(t)][(N + 1)P_{N+1} - NP_N(t)].$$

As is well known, this has the solution

$$(2.3) \quad P_N(t) = \binom{N_0}{N} \exp \{-N_0[Mt + qF(t)]\} (1 - \exp \{-[Mt + qF(t)]\})^{N_0 - N},$$

where $F(t) = \int_0^t f(u) du$, that is, $F(t)$ is the accumulated effort up to time t .

Further,

$$(2.4) \quad E[N(t)] = N_0 \exp \{-[Mt + qF(t)]\}$$

and

$$(2.5) \quad \sigma^2[N(t)] = N_0 \exp \{-[Mt + qF(t)]\} (1 - \exp \{-[Mt + qF(t)]\}).$$

Equation (2.4) is essentially the same as the usual deterministic equation. In the deterministic theory of fishing a similar differential equation leads to a catch equation. Hidden in this approach is the assumption that $f(t) = f$, that is, that

fishing goes on at a continuous level of effort at all times. This is unrealistic for most fisheries. We avoid making this assumption in general but rather approximate $f(t)$ by a step function.

Thus let $C(t)$ = catch (in numbers) in interval $(0, t)$ and let

$$(2.6) \quad f(t) = f_i, \quad t_{i-1} \leq t < t_i; i = 1, 2, \dots,$$

where $t_0 = 0$, and $\lim_{i \rightarrow \infty} t_i = \infty$, and we write $t_i - t_{i-1} = \Delta_i$ so that

$$(2.7) \quad F(t_i) = \sum_{j=1}^i f_j \Delta_j = F_i,$$

say. During the i th interval the expected mortality is

$$(2.8) \quad N_0 \{ \exp [-(Mt_{i-1} + qF_{i-1})] - \exp [-(Mt_i + qF_i)] \}$$

and hence the expected catch in this interval is

$$(2.9) \quad \frac{qf_i}{M + qf_i} N_0 \exp [-(Mt_{i-1} + qF_{i-1})] \{1 - \exp [-(M + qf_i)\Delta_i]\}.$$

Such a step function approximation to $f(t)$ may be the most that can be achieved with actual data. However, for theoretical purposes we may regard this as a Riemann sum and with the usual limiting operation find that

$$(2.10) \quad E[C(t)] = \int_0^t N_0 q f(u) \exp \{ -[Mu + qF(u)] \} du.$$

In the special case that $f(t) = f$, then

$$(2.11) \quad E[C(\infty)] = \frac{qf}{M + qf} N_0.$$

Because of the important role that annual data play in fishery statistics we note some formulas pertaining to $C(1)$. If

$$(2.12) \quad f_i = \begin{cases} f, & 0 \leq t \leq \tau, \\ 0, & \tau \leq t \leq 1, \end{cases}$$

then

$$(2.13) \quad E[C(1)] = \frac{qfN_0}{M + qf} \{1 - \exp [-(M + qf)\tau]\}$$

and

$$(2.14) \quad E[N(1)] = N_0 \exp [-(M + qf\tau)]$$

so that it is natural to identify $qf\tau$ as the fishing mortality rate.

For simplicity in later work we write $qf\tau = F$ and note that when fishing is at a constant level, that is, $f(t) = f$, then $\tau = 1$ so that $F = qf$ in this simple case.

A further specialization is the case of a short intensive fishery which may be realized by letting $\tau \rightarrow 0$, $f \rightarrow \infty$, with $0 < F < \infty$, hence

$$(2.15) \quad E[C(1)] = N_0(1 - e^{-F}).$$

2.1. Tagging estimates. Consider now the estimation of the mortality rates M , F (or M and the catchability coefficient q). In some cases only $M + F$ are

estimable; we then set $M + F = Z$, the total instantaneous mortality rate. The subpopulation of initial size N_0 may be a *tagged* population and the estimation of population parameters in this case has been well studied for a sequence of discrete point samples. These results are applicable to short intensive fisheries of the type above, formula (2.13). In this case the total survival rate and the total population size can be estimated (see Chapman [6]); it follows easily that the mortality components F , M may be also estimated. Less attention has been paid to the case where sampling of the population and tag recovery is based on a continuous fishery. For a fishery that is both continuous and at a constant level, Gulland [13] has given the maximum likelihood estimates of M and F and noted that they are biased. That no minimum variance unbiased estimate of F exists may be shown following the methods of Barankin [1]. But an essentially unbiased estimate may be found for this case where $f(t)$ is constant. Suppose there are n recaptures of the N_0 tagged animals at times t_i , where $i = 1, 2, \dots, n$. As noted above, the probability of recapture is $qf/(M + qf)$, that is, $F/(M + F)$ or F/Z . Now assume that the recaptures are binomially distributed,

$$(2.16) \quad P\{n\} = \binom{N_0}{n} \left(\frac{F}{M + F} \right)^n \left(\frac{M}{M + F} \right)^{N_0 - n}.$$

Further, given $n \geq 1$, the random variables t_1, t_2, \dots, t_n have an exponential distribution, that is, their density function has the form

$$(2.17) \quad f(t) = \begin{cases} 0, & -\infty < t \leq 0, \\ (F + M) \exp [-(F + M)t], & 0 < t < \infty. \end{cases}$$

Define $T = \sum_{i=1}^n t_i$. For $n > 1$ there exists a minimum variance estimate of $F + M$ as a function of T , namely, $(n - 1)T^{-1}$, but no such estimate exists for $n = 1$ and only trivially for $n = 0$. Also, n/N_0 is a minimum variance estimate of $F/F + M$. Hence, neglecting $P\{n = 1\}$, that is, a single recapture,

$$(2.18) \quad \hat{F} = \begin{cases} \frac{n(n-1)}{N_0 T}, & n \geq 2, \\ 0, & \text{otherwise,} \end{cases}$$

is almost unbiased. Also,

$$(2.19) \quad \sigma^2(\hat{F}) = \frac{F(F + 2m)}{N_0} + \frac{2(F + M)}{N_0^2} + O\left(\frac{1}{N_0^3}\right).$$

Further, the corresponding estimate of M is

$$(2.20) \quad \hat{M} = \frac{(N_0 - n)(n - 1)}{N_0 T},$$

with

$$(2.21) \quad \sigma^2(\hat{M}) = \frac{M(F^2 + FM + M^2)}{N_0} + O\left(\frac{1}{N_0^3}\right).$$

For the more general case this method of attack is complicated by the unpleasant form both of the ultimate probability of recapture and of the conditional

distribution of the t_i . The expected number of recaptures is given by (2.10) while the conditional distribution of the t_i is

$$(2.22) \quad h(t) = \frac{qf(t) \exp \{-[Mt + qF(t)]\}}{\int_0^\infty qf(u) \exp \{-[Mu + qF(u)]\} du}$$

A general method of estimating M and q based on the recaptures t_1, t_2, \dots, t_n (but not on N_0) would be to obtain maximum likelihood estimates from this conditional density. The equations to be solved are

$$(2.23) \quad nE(u) = T, \quad nE[F(u)] = F_T,$$

where $F_T = \sum_{i=1}^n F(t_i)$ and where the expectations are functions of M and q . In case $f(t)$ is approximated by the step function (2.6), then $E(u)$ is the ratio of

$$(2.24) \quad \sum_{i=1}^\infty \frac{f_i}{(M + qf_i)^2} \{[(M + qf_{i-1})t_{i-1} + 1] \exp [-(M + qf_{i-1})t_{i-1}] - [(M + qf_i)t_i + 1] \exp [-(M + qf_i)t_i]\} \exp [-q(F_{i-1} - t_{i-1}f_i)]$$

divided by

$$(2.25) \quad \sum_{i=1}^\infty \frac{f_i}{M + qf_i} \exp [-(Mt_{i-1} + qF_{i-1})] \{1 - \exp [-(M + qf_i)\Delta_i]\}.$$

In practice the infinite sums reduce to finite sums but even if there are only a small number of terms, estimation of M and q appears prohibitive without a high-speed computer. As an alternative, we consider two approximate methods depending on whether the Δ_i , the intervals of constant effort, are small or large. Which method is usable will also depend on the available data. Neither of these methods depends on knowledge of N_0 ; we comment on this later.

If the total number of recaptures is small so that the dependence between successive recaptures may be neglected, and if the recaptures are grouped by interval of uniform effort ($t_{i-1} \leq t < t_i$), then

$$(2.26) \quad E\left(\frac{n_i}{f_i}\right) = \frac{N_0q}{M + qf_i} \{1 - \exp [-(M + qf_i)(t_i - t_{i-1})]\} \exp \left[-(Mt_{i-1} + q \sum_{j=1}^{i-1} f_j \Delta_j) \right].$$

Now if

$$(2.27) \quad \log \frac{1 - \exp [-(M + qf_i)(t_i - t_{i-1})]}{M + qf_i}$$

may be approximated by

$$(2.28) \quad \log (t_i - t_{i-1}) - \frac{(M + qf_i)(t_i - t_{i-1})}{2},$$

which is reasonable if $(M + qf_i)(t_i - t_{i-1})$ is small, then q and M may be estimated by regressing $\log n_i/f_i$ on t_i and F_i . The variances of the estimates of M and q are found as usual in multiple regression analysis. As is frequently found

in data of this type, the logarithmic transformation may lead to approximate homoscedasticity.

If the periods Δ_i are longer and if recapture times are recorded, then these provide information on $M + qf_i$ for each period Δ_i . The earlier analysis is not directly applicable since the periods of observation are truncated. For each period (t_{i-1}, t_i) of length Δ_i , the conditional distribution of the times of recapture measured from the start of the period has the form of the truncated exponential

$$(2.29) \quad h(t) = \frac{Z_i e^{-Z_i t}}{1 - e^{-Z_i \Delta_i}},$$

where we have written $M + qf_i = Z_i$ and, as noted, $\Delta_i = t_i - t_{i-1}$.

The maximum likelihood estimate of the scale parameter of the truncated exponential was studied by Deemer and Votaw [10]. The estimate of Z_i is the solution of the equation

$$(2.30) \quad \frac{T_i}{n_i \Delta_i} = \frac{1}{Z_i \Delta_i} - \frac{1}{e^{Z_i \Delta_i} - 1},$$

when $T_i < n_i \Delta_i / 2$ and 0 otherwise. Here $T_i = \sum_{i=1}^{n_i} t'_i$, the t'_i being times of recapture in the interval Δ_i measured from t_{i-1} and n_i being the number of recaptures. Deemer and Votaw provided a table [10] to assist in the solution of (2.26). The large-sample variance of Z_i is

$$(2.31) \quad \sigma^2(\hat{Z}_i) = \frac{1}{n_i} [Z_i^{-2} - \Delta_i^2 e^{-Z_i \Delta_i} (1 - e^{-Z_i \Delta_i})^{-2}]^{-1}.$$

Finally then, since $Z_i = M + qf_i$, estimates of M and q are again found by weighted regression analysis, the estimated weights of the \hat{Z}_i being found with the help of (2.31).

In both of these methods separate estimation of M and q is possible if and only if $f_i \neq f$ for all i , that is, effort varies. Also, in general the greater the variation the smaller the variances of estimate of M and q .

A statistically complete treatment of the estimation of M and q (or M and F) based on N_0 , n , and $t_1 \cdots t_n$ for a continuous fishery with varying fishing intensities remains an open question. A number of attempts that have been made are reviewed and criticized by Beverton and Holt ([5], pp. 185–191). While there is a theoretical gap here, from a practical point of view it may be advantageous not to have to use information on N_0 , for there may be initial tag loss and mortality caused by the tagging procedure which reduces N_0 from the indicated value. Further, there may be a time lapse before the tagged fish are fully "vulnerable" (systematic errors of type C in Ricker's terminology [20], p. 122).

There are in addition other problems associated with tagging experiments, for example, continuing tag loss or differential tagging mortality, incomplete reporting or return of tags (Ricker's systematic type A and B errors). In studies of large marine populations there are also the more acute problems of placing a sufficient number of tags in a large and widely scattered population and of having them dispersed randomly throughout the population. While this point

cannot be discussed here in detail, it is important to point out that a tagging experiment cannot be analyzed completely without taking into account movements in the population and the degree of mixing. For example, it should be asked: is the population composed of a collection of subpopulations which do not mix at all, or a single homogeneous freely mixing population, or is the actual situation somewhere between these two extremes? Much qualitative work has been done upon migration based on information obtained from tagging experiments. A statistical study of tag recoveries in a stratified population was given by Chapman and Junge [7]; these results apply to point samples from a population with $M = 0$. Extensions to the general case with samples obtained by a continuous fishery are necessary.

Rather than a purely statistical approach which may involve a formidable tagging program, models may be formulated which utilize biological information on the movements of the fish. Skellam [24] has shown the possibility of using a diffusion model in a related problem, namely, the dispersion of a population into new areas and in connection with certain other ecological models. Beverton and Holt, in their work already referred to [5], suggest some tentative models which are deterministic. However, if the results are to be applied to tagging experiments, then a stochastic model may be necessary, for whereas the population as a whole may number in the millions, the tag returns classified by subarea and by year will number from two or three to a few score. For these reasons estimates from age composition and effort data have been sought that are not based on a tagging process. Leslie and Davis [15] and independently DeLury [11] showed how effort data could be used to estimate the population size (and hence q or F) for populations with $M = 0$. For the more general case where $M \neq 0$ it is necessary to consider two possibilities, the fishery operating at a constant level vs. a variable-level fishery. Only in the latter case is separate estimation of M and q possible. In the former case estimation of $Z = F + M$, the total mortality rate, is possible. We consider this case first.

For the case of constant effort and a short intensive fishery, methods developed by Chapman and Robson [8] for the analysis of a single catch curve are applicable. We assume that the short intensive fishery operates annually at a similar effort level. Now consider any year class and code as zero that age at which the fish are fully available to the fishery. Let X = age of capture (coded). The assumption of constant fishing and natural mortalities implies a constant probability of survival and thus leads to the following frequency function for X ,

$$(2.32) \quad f(x) = (1 - S)S^x,$$

where $S = 1 - e^{-z}$. Then, following Chapman and Robson, the minimum variance unbiased estimate of S is

$$(2.33) \quad \hat{S} = \bar{X} \left(1 + \bar{X} - \frac{1}{n} \right)^{-1}$$

with

$$(2.34) \quad \sigma^2(\hat{S}) = \frac{S(1-S)^2}{n}$$

Extension of this to truncated samples are given in the paper cited; also given is a nearly unbiased estimate of $F + M$.

If the fishery is continuous at a constant level and continuous catch data is recorded, then times of capture t_1, t_2, \dots, t_n from a specified year class yield a minimum variance unbiased estimate of Z , namely,

$$(2.35) \quad \hat{Z} = (n-1)T^{-1}, \quad \sigma^2(\hat{Z}) = \frac{Z^2}{n-2}$$

With data obtained annually or continuously, the combining of results from several year classes presents no problem. The relevant statistic to be calculated is T , the total time until capture measured from zero. Estimation of Z from incomplete (truncated) data follows exactly the formulas given earlier (2.30), (2.31).

Turning to the more general case where there are variations in effort, M and q can both be estimated. This possibility of utilizing variations in efforts to estimate M and q separately from age and effort data was apparently first recognized by Silliman [23], who considered variations in total mortality from a high level of effort to a low level. Widrig [28] and independently Beverton and Holt [5] extended the methods as follows. Let C_i = catch in year i ; assume $f_i = f$ in year i ; from (2.9)

$$(2.36) \quad E(C_i|N_i) = \frac{qf_i N_i}{M + qf_i} \{1 - \exp[-(M + qf_i)]\},$$

$$(2.37) \quad E(N_{i+1}|N_i) = N_i \exp[-(M + qf_i)],$$

whence they write

$$(2.38) \quad \log \frac{C_i}{f_i} - \log \frac{C_{i+1}}{f_{i+1}} = \log \frac{\{1 - \exp[-(M + qf_i)]\} (M + qf_{i+1})}{\{1 - \exp[-(M + qf_{i+1})]\} (M + qf_i)} + M + qf_i.$$

The nonlinear term on the right side of (2.38) may be negligible; in particular it is, to a first approximation,

$$(2.39) \quad \frac{q}{2}(f_{i+1} - f_i) \left(1 - \frac{M}{6}\right).$$

The method proposed is to regress $y_i = \log C_i/f_i - \log C_{i+1}/f_{i+1}$ on f_i by least squares (and in addition Beverton and Holt propose an iterative procedure to take account of the nonlinear term). Ordinary regression methods are insufficient for the y_i are correlated, since they are defined as successive differences. In addition there is the correlation of the series of catches from the same population. If the second of these effects and the nonlinear term are neglected, as is perhaps reasonable for large populations and small changes of effort, and if it is assumed that the $\log C_i$ are normally distributed with expectations of C_i given by (2.36) and common variances σ^2 , then the y_i with $i = 1, 2, \dots, r-1$ have

a simple moment matrix with a well-known and easily computed inverse. This inverse is proportional to

$$(2.40) \quad \begin{bmatrix} r-1 & r-2 & r-3 & \cdots & 3 & 2 & 1 \\ r-2 & 2(r-2) & 2(r-3) & \cdots & 6 & 4 & 2 \\ r-3 & 2(r-3) & 3(r-3) & \cdots & 9 & 6 & 3 \\ \cdot & \cdot & \cdot & \cdots & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot & \cdot & \cdot \\ 3 & 6 & 9 & \cdots & 3(r-3) & 2(r-3) & r-3 \\ 2 & 4 & 6 & \cdots & 2(r-3) & 2(r-2) & r-2 \\ 1 & 2 & 3 & \cdots & r-3 & r-2 & r-1 \end{bmatrix}$$

so that least squares estimates of the parameters may be calculated. If (2.39) is not negligible, then not only is a tedious iteration required but this weighting is not optimum.

These problems may be avoided by noting that either of the regression methods given earlier can be applied to captures from any specified year class. These regression equations were

$$(2.41) \quad \log \frac{n_i}{f_i} = \log(N_0 q) - M \frac{(t_i + t_{i-1})}{2} - q \frac{F(t_i) + F(t_{i-1})}{2},$$

where the intervals Δ_i are small, and

$$(2.42) \quad \hat{Z}_i = M + qf_i,$$

where the \hat{Z}_i were based on times of recaptures within the intervals Δ_i . That the (unknown) N_0 does not enter into these equations is now essential. Actual data of this kind will be grouped into broad classes and to this extent the variances given by (2.27) underestimate the true variance. We re-emphasize that the estimation of M and q is improved as variations in effort increase.

Throughout this treatment we have followed the usual assumption in fisheries of a constant natural mortality rate M . If M is not constant mortality rates could be determined only from a tagging experiment (see, for example, Darroch [9]), though this would require a large tagging experiment in a widely dispersed marine population.

3. Growth rates

The growth rate of the population plays a role in the determination of the optimum exploitation: exploitation should be designed to let escape those age classes in which growth exceeds natural mortality but exploit as rapidly as possible those ages in which the situation is reversed. The earliest attempts to resolve this problem were based on empirical growth curves. But another factor in the optimum level of exploitation is the response of growth rate to change in the population level. It does not seem possible to incorporate such responses into a purely empirical curve. Thus there has been a search for simple growth curves

which are not entirely artificial. The most commonly used growth curve is that due to von Bertalanffy [4], which starts from the basic equation

$$(3.1) \quad \frac{dw}{dt} = Hs - kw,$$

where w = weight, s = surface, H , k are parameters, and proceeds further from the assumption that

$$(3.2) \quad s = pl', \quad w = ql^3,$$

where p , q are parameters, l is length, and w , s , and l are all functions of time.

However, it is known for many fish that w is not proportional to length cubed but some other power of l . It is also questionable both that synthesis of mass is directly proportional to surface or that surface area is directly proportional to length squared. A more general starting equation than (3.1) is

$$(3.3) \quad \frac{dw}{dt} = Hw^\alpha - kw$$

with

$$(3.4) \quad w = ql^\beta.$$

Then we have

$$(3.5) \quad \frac{dl}{dt} = El^\beta - Kl,$$

where

$$(3.6) \quad E = \frac{Hq^{\alpha-1}}{\beta}, \quad K = \frac{k}{\beta}, \quad \delta = 1 - \beta(1 - \alpha).$$

Integrated, this yields

$$(3.7) \quad l_i^{1-\delta} = \frac{E}{K} - \frac{1}{K} \exp[-K(1-\delta)(t+C)],$$

with C the arbitrary constant of integration.

If at $t = 0$ we set $l = l_0$ (not necessarily zero) and further note that as $t \rightarrow \infty$ then $l \rightarrow (E/K)^{1/(1-\delta)}$, the equation becomes

$$(3.8) \quad l_i^{1-\delta} = l_\infty^{1-\delta} - (l_\infty^{1-\delta} - l_0^{1-\delta}) \exp[-K(1-\delta)t].$$

Finally, write $L_t = l_i^{1-\delta}$; $K' = K(1-\delta)$ and the equation reduces to the von Bertalanffy equation

$$(3.9) \quad L_t = L_\infty - (L_\infty - L_0)e^{-K't},$$

but unless $\delta = 0$ the units involved are of course different. It may be noted that integrating (3.3) as an equation in w would lead to an equation of similar form to (3.8) in w .

Consider first the case where δ is known, that is, is assumed to be zero as classically has been the case or is estimated separately from weight-length relationships, together with the additional assumption that $\alpha = 2/3$. Then (3.9) is meaningful with L_t the observed random variable and $K' = K$ so that we drop the prime for now. The estimation of the parameters of a nonlinear regres-

sion equation of the form of (3.9) has been the subject of an extensive literature in recent years. In particular, Stevens [25] has given a procedure to determine maximum likelihood estimates of the three parameters, assuming that the dependent variables are uncorrelated and have equal variances. Stevens' method involves some heavy computation, though it is well adapted to high-speed computers and hence should not be avoided on this account. Patterson [18] has given a simpler procedure and he and Finney [19], [12] have discussed in two separate articles its efficiency and the efficiency of the classical autoregressive estimation of the parameter K in the exponent. The latter is based on the fact that equation (3.9) can be transformed to

$$(3.10) \quad L_{t+1} = L_{\infty}(1 - e^{-K}) + e^{-K}L_t.$$

While this transformation is the one that has been used in recent work in fisheries (see Beverton and Holt [5], pp. 279-288, and Walford [27]), two other variations of this seem more promising. Let

$$(3.11) \quad L_{t+1} - L_t = \Delta_t L, \quad 1 - e^{-K} = \kappa.$$

Then

$$(3.12) \quad \Delta_t L = \kappa(L_{\infty} - L_t)$$

and

$$(3.13) \quad \Delta_t L = \Delta_{t-1} L(1 - \kappa).$$

The latter equation was given by Willers [29]. The use of (3.12) would depend on the assumption that $\Delta_t L$ is distributed independently of L_t . This may be reasonable but it is also probable that the growth increments $\Delta_t L$ have variances which are functions of t , so that some weighting factor should be introduced. This and questions connected with the use of (3.13) to estimate K and L need further investigation. The use of a weighted regression to give parameter estimates based on (3.13) would lead to simple confidence interval estimates for these parameters.

It is reasonable to think of these equations as applying to an individual fish, each fish having its own parameters L_{∞} , L_0 . We are really interested in $E(L_{\infty})$, $E(\kappa)$ over the whole population. But the linear regression applied to all the data leads us to weighted estimates of these parameters. Data on lengths of fish at several ages is obtained through the technique known as back-calculation, making use of the annual rings laid down on scales, otoliths, and so on. It has been found that there is usually some simple relationship (often one of direct proportionality) between scale radii and body length. In the simplest case, κ could be estimated directly from scale-radius increments. For the more general case, with δ of (3.8) unknown, an iterative procedure can be set up to choose $\hat{\delta}$ as that estimate of δ which minimizes the residual variance of $\Delta_t^{1-\delta} L$ when regressed on $l_t^{-\delta}(\Delta_t^{1-\delta} L = l_{t+1}^{\delta} - l_t^{\delta})$. This leaves open two important problems: the variance of $\hat{\delta}$ so obtained and the effect estimation of δ has on the estimated variances of \hat{L}_{∞} and $\hat{\kappa}$. When high-speed computing equipment is available this

procedure can be programmed to yield $\hat{\delta}$, \hat{L}_∞ , $\hat{\kappa}$ for each fish measured and hence variances of estimates of $E(\delta)$, $E(L_\infty)$, $E(\kappa)$ are obtained in the usual elementary manner.

The extended form of the von Bertalanffy equation and the procedure of estimation given above both bear some resemblance to a formulation of Parker and Larkin [17]. However, they start with the simpler equation $dw/dt = kw^z$, where k and z are parameters, rather than (3.1) and are led to an equation of the form $l_{i+1} = \alpha + l_i^z$ with α , z parameters.

The value of any of these parametric treatments is in giving meaning to the parameters. Thus Beverton and Holt, following von Bertalanffy, assume K is directly related to the rate of catabolism of the fish and is thus essentially constant. This implies that where growth rate is density dependent, if equation (3.9) is an adequate formulation of the growth pattern, then L_∞ is the parameter that changes with population size. The determination of whether this is so and if so what the relationship is between L_∞ and population size follows standard procedures. Studies of this are limited at the moment chiefly by the availability of data.

4. Sampling problems

The problems associated with acquisition of the data necessary to estimate population parameters have not been considered so far. As noted, the primary sampling of an exploited fish population is done by the commercial fishery. It is easy to realize that the commercial fishery is hardly interested in proper design to obtain a probability sample. The research agency must determine what information it is possible to obtain from the fishery and what additional information must be obtained by sampling the commercial catch (which is then a subsampling with respect to the original population) and what must be obtained by independent experiments.

The first problem is that the population may not be well defined: are there isolated segments of the population not fished? While this problem can be side-stepped by defining the population as those elements available to the fishery, it is not clear that with this definition the population remains the same from year to year. Widrig [28] in particular has raised this question with respect to the California sardine population. The problems associated with movement of the fish have been referred to earlier. Even if the fishery moves rapidly with respect to the movement of the fish so that the sampling of the fishery is adequate, there remains the question of whether the probability of a fish being sampled by the fishery is the same for each individual in the population. In fact, in general it is known this is not so. One type of selectivity found in most fisheries is based on size. The commercial fishery fixes its gear (in type or place) as required by law or dictated by economic considerations to avoid the smallest fish. There may also be selection against the largest sizes (if, for example, they avoid the gear). In some cases there may be a question of whether or not there is selection. More

frequently in fisheries the existence of selection is well known and the problem is the estimation of the "selective curve." This needs to be made more precise.

Let X, Y be random variables with probability density functions $r^{-1}S(x)f(x)$ and $f(x)$ respectively, where $0 \leq S(x) \leq 1$,

$$(4.1) \quad r = \int_{-\infty}^{\infty} s(x)f(x) dx, \quad F(y) = \int_{-\infty}^{\infty} f(t) dt$$

is the length distribution in the population while $S(x) =$ probability that an individual of length x is sampled. $S(x)$ is the "selection function." In general, $0 \leq S(x) \leq 1$ but it is clear that it may be redefined so that $\sup S(x) = 1$, $-\infty < x < \infty$. We do so.

A few forms of $S(x)$ are of interest as suggested by actual situations. The simplest is

$$(4.2) \quad S(x) = \begin{cases} 0, & x < a, \\ 1, & x > a, \end{cases}$$

which results in a truncated distribution. This idealized situation rarely occurs and more usually $S(x)$ has the same shape as a symmetrical cumulative distribution. This led the earliest workers to assume $S(x)$ has the integrated normal form, that is,

$$(4.3) \quad S(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp \left[-\frac{(t - \mu_S)^2}{2\sigma_S^2} \right] dt,$$

with parameters μ_S, σ_S^2 .

If it is further assumed that $f(x)$ is known and the data comes in groups as will usually be the case, then the estimation problem is closely related to those of probit or logit analysis.

Let $p_i =$ relative frequency of fish in length class i ,

$\xi_i =$ midpoint of class i ,

$N =$ total population size,

$\rho =$ probability that individual is available to unselective gear,

$x_i =$ number caught in length class i , with $i = 1, 2, \dots, r$.

It is reasonable to assume that the x_i have a Poisson distribution with

$$(4.4) \quad E(x_i) = N\rho_i \Phi \left(\frac{\xi_i - \mu_S}{\sigma_S} \right),$$

and since $N\rho$ are inseparable we write instead $N\rho = \gamma$. The unknown parameters are γ, μ_S, σ_S . It is feasible to obtain R.B.A.N. estimates of these iteratively. Given any estimate of γ (a simple starting estimate is $X_r p_r^{-1}$) then estimation of μ_S, σ_S proceeds along the lines outlined by Taylor [26], that is, minimizing

$$(4.5) \quad \sum_{i=1}^r \frac{1}{x_i} \left[\Phi^{-1} \left(\frac{x_i}{p_i} \right) - \frac{\xi_i - \mu_S}{\sigma_S} \right]^2 \left[\Phi' \left(\frac{x_i}{p_i} \right) \right]^2 p_i^2.$$

Then, if μ_S, σ_S are estimated, it is possible to estimate γ by minimizing

$$(4.6) \quad \sum_{i=1}^r \frac{1}{x_i} \left[x_i - \gamma p_i \Phi \left(\frac{\xi_i - \mu_S}{\sigma_S} \right) \right]^2.$$

The minimizing parameter values are solutions of linear equations only.

If $f(x)$ is not known but observations on Y are also available (in fisheries terms this means that it is also possible to take a nonselective sample), then we have an additional set of variables Y_i with $i = 1, 2, \dots, r$, which may be assumed to have a Poisson distribution with $E(Y_i) = N\rho'p_i$. Denote $N\rho' = \gamma$ and $\rho/\rho' = \lambda$. As $\gamma, \gamma' \rightarrow \infty$,

$$(4.7) \quad \frac{\frac{x_i}{y_i} - \lambda \Phi \left(\frac{\xi_i - \mu_S}{\sigma_S} \right)}{\frac{x_i}{y_i^2} + \frac{x_i^2}{y_i^3}}$$

is asymptotically $N(0, 1)$ and a similar estimation procedure may be used. It is not correct to say the resulting estimates are R.B.A.N. for we start with the random variable X_i/Y_i rather than (X_i, Y_i) . If there are no observations on Y and $f(x)$ is either completely unknown or assumed to have some parametric form with some unknown parameters, then the problem of estimating μ_S, σ_S , and $f(x)$ or the parameters of $f(x)$ is completely open.

In fisheries we may have none of these situations but instead may have catches by two or more nets, each with its own selectivity function. Or alternatively, an attempt may be made to build a model relating the selectivity to the physical characteristics of the fish and of the catching device (compare Baranov [3], McCombie and Fry [16]). Also, other forms may be assumed for the selection curve. There may be selection in both directions; in this case, if some symmetry is apparent in the selection function, it is customary to assume that the selection function has the same form as a normal density. On the other hand, the selection in the left and right tails may be due to different reasons so that it is not reasonable to assume this symmetry. Only partial solutions have been given in these cases and in general these are statistically incomplete.

5. Maximum sustainable productivity

The problems so far considered relate essentially to static situations, that is, the estimation of the parameters is treated for a given population level. Next must be determined which parameter values are responsive to changes in population level and what those changes are. This requires that data to make satisfactory estimates of the parameters be available over a range of values of the population levels and, as in all regression studies, the larger the range the more satisfactory the results obtained. The North Sea plaice population has been studied carefully for a long time and the cessation of fishing that was imposed by World War II gave rise to the large variations in exploitation and hence in population level that are required to explore such possibilities. Beverton and Holt have used the available data to make such exploration in considerable

detail. However, as noted, their aim was to indicate the range of possible effects rather than to yield precise intervals for the estimates.

In the absence of such detailed information, a number of studies have been made based on catch and effort data alone, that is, the processes have been studied in the large. The aim of such studies has been to relate equilibrium yield (the catch sustainable by the population without changing the population level) to population size. Perhaps the most careful of such studies is that due to Schaefer [21], [22]. He assumes a logistic growth law so that the equilibrium population change for any year is related to population level from catch and catch per unit of effort data. All of this involves a number of important assumptions which he notes but also some additional assumptions on the validity of effort data over a period of years. Nevertheless, this gross approach may be more useful at present to the management of the fishery, which must try to assess maximum sustainable productivity (catch) of the fishery and which has only such data available to it.

More interesting perhaps statistically is the establishment of a stochastic estimation procedure to determine this maximum, following the procedure of Kiefer and Wolfowitz [14] and several subsequent authors. In fact, many regulatory schemes do now proceed with the aim of working toward the maximum sustainable productivity. It might be useful to formulate explicit stochastic estimation procedures which must incorporate side conditions of "continuity" of behavior, that is, the regulation must be reasonably stable, and which also take into account the lags in the system.

REFERENCES

- [1] E. W. BARANKIN, "Locally best unbiased estimates," *Ann. Math. Statist.*, Vol. 20 (1949), pp. 477-501.
- [2] T. I. BARANOV, "On the question of the biological basis of fisheries," *Rep. Div. Fish Management and Scientific Study of the Fishing Industry*, Vol. I (1918), pp. 81-128.
- [3] ———, *The Theory and Assessment of Fishing Gear*, Moscow, Fishchempromisdat, 1948.
- [4] L. VON BERTALANFFY, "A quantitative study of organic growth (Inquiries on growth laws. II)," *Human Biology*, Vol. 10 (1938), pp. 181-213.
- [5] R. J. H. BEVERTON and S. J. HOLT, *On the Dynamics of Exploited Fish Populations*, Fish. Invest. Series II, Vol. XIX, Ministry of Agriculture, Fisheries and Food, Great Britain, London, Her Majesty's Stationery Office, 1957.
- [6] D. G. CHAPMAN, "The estimation of biological populations," *Ann. Math. Statist.*, Vol. 25 (1954), pp. 1-15.
- [7] D. G. CHAPMAN and C. J. JUNGE, "The estimation of the size of a stratified animal population," *Ann. Math. Statist.*, Vol. 27 (1956), pp. 375-389.
- [8] D. G. CHAPMAN and D. S. ROBSON, "The analysis of a catch curve," *Biometrics*, Vol. 16 (1960), pp. 354-368.
- [9] J. N. DARROCH, "The multiple recapture census. II. Estimation when there is immigration or death," *Biometrika*, Vol. 46 (1959), pp. 336-351.
- [10] W. L. DEEMER and D. V. VOTAW, JR., "Estimation of parameters of truncated or censored exponential distributions," *Ann. Math. Statist.*, Vol. 26 (1955), pp. 498-504.
- [11] D. B. DELURY, "On the estimation of biological populations," *Biometrics*, Vol. 3 (1947), pp. 145-167.

- [12] D. J. FINNEY, "The efficiency of alternative estimators for an asymptotic regression equation," *Biometrika*, Vol. 45 (1958), pp. 370-388.
- [13] J. A. GULLAND, "On the estimation of population parameters from marked members," *Biometrika*, Vol. 42 (1955), pp. 269-270.
- [14] J. KIEFER and J. WOLFOWITZ, "Stochastic estimation of the maximum of a regression function," *Ann. Math. Statist.*, Vol. 23 (1952), pp. 462-467.
- [15] P. N. LESLIE and D. H. S. DAVIS, "An attempt to determine the absolute number of rats in a given area," *J. Anim. Ecol.*, Vol. 8 (1939), pp. 94-113.
- [16] A. M. McCOMBIE and F. E. J. FRY, "Selectivity of gill nets for lake whitefish *coregonus clupeaformis*," *Trans. Amer. Fish. Soc.*, Vol. 89 (1960), pp. 176-184.
- [17] R. A. PARKER and P. A. LARKIN, "A concept of growth in fishes," *J. Fish. Res. Bd. Canada*, Vol. 16 (1959), pp. 721-745.
- [18] H. D. PATTERSON, "A simple method for fitting an asymptotic regression curve," *Biometrics*, Vol. 12 (1956), pp. 323-329.
- [19] ———, "The use of auto-regression in fitting an exponential curve," *Biometrika*, Vol. 45 (1958), pp. 389-400.
- [20] W. E. RICKER, *Handbook of Computations for Biological Statistics of Fish Populations*, Bull. 119, Fish. Res. Bd. Canada, Ottawa, 1958.
- [21] M. B. SCHAEFER, "Some aspects of the dynamics of populations important to the management of the commercial marine fisheries," *Inter-Amer. Trop. Tuna Comm.*, Bull. 1 (1954), pp. 27-56.
- [22] ———, "A study of the dynamics of the fishery for yellowfin tuna in the Eastern Tropical Pacific Ocean," *Inter-Amer. Trop. Tuna Comm.*, Bull. 2 (1957), pp. 247-285.
- [23] R. P. SILLIMAN, "Studies on the Pacific pilchard or sardine (*Sardinops caerulea*) 5: A method of computing mortalities and replacements," *U.S. Fish and Wildlife Service Spec. Sci. Rep.*, No. 24 (1943).
- [24] J. G. SKELLAM, "Random dispersal in theoretical populations," *Biometrika*, Vol. 38 (1951), pp. 196-218.
- [25] W. L. STEVENS, "Asymptotic regression," *Biometrics*, Vol. 7 (1951), pp. 247-267.
- [26] W. F. TAYLOR, "Distance functions and regular best asymptotically normal estimates," *Ann. Math. Statist.*, Vol. 24 (1953), pp. 85-92.
- [27] L. A. WALFORD, "A new graphic method of describing the growth of animals," *Biol. Bull.*, Vol. 90 (1946), pp. 141-147.
- [28] T. M. WIDRIG, "Method of estimating fish populations with application to Pacific sardine," *U.S. Fish and Wildlife Service, Fish. Bull.*, Vol. 56 (1954), pp. 141-166.
- [29] F. A. WILLERS, *Practical Analysis*, New York, Dover, 1948.