International Council for the Exploration of the Sea

ICES C.M. 1993/D:51
ref. K

Index-Removal Estimators of Population Size which Incorporate Information on Sampling Gear Selectivity

John M. Hoenig, Earl G. Dawe and Xucai Xu
Department of Fisheries and Oceans, P.O. Box 5667, St. John's, Nfld A1C 5X1, Canada

Index-removal estimation is one of the basic approaches to estimating the size of animal populations. The logic is quite simple: if catch rate (catch per unit of sampling effort) is proportional to animal abundance, and if a known removal causes the catch rate to decline by a specified proportion P, then the removal is equal to 100P % of the population. For example, if the catch rate is 10 before the removal of 300 animals, and is 7 after the removal is made, then we calculate that the removal of 300 animals resulted in a loss of (10 - 7)/10 = 3/10 of the population. Thus, the population size (before the removal) must have been 1000 animals. Also, the catchability coefficient, q, would be 10/1000 = .01. An assumption of the method is that all animals have the same probability of capture. Clearly, this is not the case if the sampling gear is size-selective. The bias can be minimized by making separate, independent estimates for each size-class of animal. However, in general, we know that the catchability of larger animals is greater than that of smaller animals. Therefore, we can achieve greater statistical efficiency if we utilize information on catchability in the estimation procedure. We propose that one might wish to first compute separate estimates for each size class. Then, if the estimated catchability coefficients show an increasing trend with size class, one could estimate the size-class-specific population sizes with the constraints that the estimated catchability coefficients must be monotonically increasing with size class. One could also assume that the catchability coefficients must be a specified function of size such as a logistic function.

Index-removal estimation is one of the basic approaches to estimating the size of animal populations. The logic is quite simple: if catch rate (catch per unit of sampling effort) is proportional to animal abundance, and if a known removal causes the catch rate to decline by a specified proportion P, then the removal is equal to 100P % of the population. For example, if the catch rate is 10 before the removal of 300 animals, and is 7 after the removal is made, then we calculate that the removal of 300 animals resulted in a loss of (10 - 7)/10 = 3/10 of the population. Thus, the population size (before the removal) must have been 1000 animals. More formally, if $E(c_1)$ and $E(c_2)$ are the expected values of the observed catch rates before and after the removal, respectively, and if R is the number of animals removed, then the population size is given by

$$N = \frac{E(c_1)\,R}{E(c_1) - E(c_2)}.$$

Furthermore, we can calculate the catchability coefficient by dividing the initial catch rate by the estimated initial population size, i.e., catchability coefficient = 10/1000 = 0.01. The catchability coefficient is the fraction of the population taken by one randomly placed unit of sampling effort when the fraction taken is small (e.g., less than 2% - see Ricker 1975).

This approach is well known in the wildlife literature (Petrides 1949; Eberhardt 1982; Seber 1982; Roseberry and Woolfe 1991) but has received little attention in the fisheries literature (Dawe et al. 1993). Seber (1982) and Routledge (1989) discuss the statistical theory in detail. In particular, Routledge (1989) generalized the approach to include J removals and J + 1 surveys.

For the simplest case described above, the assumptions of the method are that: 1) the population is closed except for the removals which are known exactly, and 2) all animals have the same probability of capture which does not change from survey to survey. It is easily verified that heterogeneity of capture probabilities can introduce bias. Suppose, for example, that the population is composed of 500 males and 500 females, that males have a catchability coefficient of 0.01 whereas females have a catchability coefficient of 0.005 (i.e., half that of the males), and that 300 males and 100 females are removed from the population between the time of the two surveys. In the first survey we would expect to catch 0.01 x 500 = 5 males if one randomly placed unit of sampling effort is expended. In the second survey we would expect to catch 0.01 x (500 - 300) = 2 males. Thus, the calculated size of the initial population of males would be

$$N = \frac{5}{5 - 2} \times 300 = 500$$

which is what we want. Similarly, the size of the female population would be calculated to be 500, as desired. However, suppose that one did not realize that males have a different catchability coefficient than females and one calculated the size of the total population from combined data on males and females. In the first survey, one would expect to catch 7.5 animals (5 males + 2.5 females) with one randomly placed unit of sampling effort. In the second survey, one would expect to catch 2 + 2 = 4 animals. Consequently, the calculated population size would be

$$\frac{7.5}{7.5 - 4} \times 400 = 857$$

instead of the actual value of 1000. Note that the heterogeneity of capture probabilities is a problem because the removal was selective with respect to capture probabilities (i.e., proportionately more of the males were removed than of the females).

The problem of heterogeneity can be minimized by making separate estimates for various subsets of the population. For example, separate estimates could be made for males and for females or for different size groups of animals. However, when information is available on the relative catchability of different groups, this information can be incorporated in the estimation procedure to increase the statistical efficiency of the estimator. For example, one may have good reason to believe that male crabs are more catchable than females in a trap survey because of differential behaviour or differential body size among the sexes. Similarly, one may believe that large crabs are more catchable by traps than small crabs. In these cases, one may wish to introduce order restrictions in the estimation procedure to ensure that the estimated catchabilities are consistent with the available information on relative catchabilities of the various groups.

In this paper, we consider a suite of four models which vary in the amount of information assumed about the relative catchabilities of the different groups. The simplest approach is to make separate, independent estimates for each group. If qualitative information is available about the relative catchabilities of the groups then one can introduce order restrictions for the catchability coefficients. One might also assume a functional relationship for the way catchability coefficients vary with a covariate. In particular, the catchability coefficient might be a logistic function of body size. In this case, we would estimate the parameters of the functional relationship rather than the catchability coefficients for each size group. Finally, it may happen that a sampling gear selectivity curve is available from some other study. In this case, the parameter estimates of the selectivity curve can be incorporated directly in the population estimation procedure.

We begin by assuming that the catches per unit of sampling effort in the surveys follow Poisson or multinomial distributions. This is consistent with previous treatments of the subject in the wildlife literature. We then discuss briefly the possibility of assuming catch rate follows a normal distribution as suggested by Routledge (1989). The normal distribution would appear to be a more reasonable model for many fishery applications.

Four models when catches follow Poisson distributions

We assume that the expected value of the total number of animals caught when one unit of sampling effort is expended at each of f randomly selected locations is given by

$$E(C) = qfN = \lambda \text{ (say)}$$

where E(.) denotes expected value of the quantity in parentheses, q is the catchability coefficient, f is the number of units of sampling effort, and N is the population size. Thus, catch is assumed proportional to sampling effort and to abundance. This assumption is justified if the sampling is with replacement (animals are released unharmed after being caught) or the fraction of the population caught is negligible so that the population size N does not change due to the random sampling. Furthermore, we assume that the total number of animals caught during the survey, C, follows a Poisson distribution with parameter λ, i.e., $C \sim P(\lambda)$. Thus, the probability density function for the number of animals caught is

$$f(C) = \frac{\lambda^C e^{-\lambda}}{C!} = \frac{(qfN)^C e^{-qfn}}{C!} .$$

If R animals are removed from the population, the abundance becomes N - R and, under the assumption that expected value of the catch is proportional to abundance, the expected value of the catch becomes E(C) = qf(N-R). If we assume that the distribution of catch remains Poisson then we can write the likelihood, Λ, for obtaining a series of catches $\{C_1, C_2, \ldots C_J\}$ from J surveys having respective sampling efforts $\{f_1, f_2, \ldots f_J\}$ as

$$\Lambda = \prod_{j=1}^{J} \frac{\lambda_j^{C_j} e^{-\lambda_j}}{C_j!} = \frac{(qf_jN_j)^{C_j} e^{-qf_jN_j}}{C_j!} . \quad (1)$$

where λ_j is the Poisson parameter for the jth random survey and N_j is the number of animals in the population just before the jth survey. Thus, N_1 is equal to the original population N, and

$$N_j = N - \sum_{k=1}^{j-1} R_k \quad \text{for } j \geq 2 \quad (2)$$

where R_k is the number of animals removed from the population after the kth survey. Here, we have treated the removals R_k as known, fixed values.

There are two unknowns, q and N, in equation (1). When the data consist of two surveys and one removal, the estimates which maximize the likelihood are given by

$$\hat{N} = \frac{C_1/f_1 \; R}{C_1/f_1 - C_2/f_2} = \frac{c_1 \; R}{c_1 - c_2}$$

and

$$\hat{q} = \frac{C_1/f_1}{\hat{N}} = \frac{c_1}{\hat{N}} = \frac{c_1 - c_2}{R}$$

($c_j = C_j/f_j$ is the catch rate in the jth survey for j = 1,2). When there are more than two surveys and one removal, the estimates must be found numerically (Routledge 1989).

Method 1: independent estimates by size group

Suppose we have reason to believe that the I subgroups in the population have different catchabilities, q_i. Suppose, further, that we believe the catches of the various subgroups are independent Poisson random variables. This assumption is made explicitly for some change-in-ratio estimation models (see Seber 1982). Then the likelihood (1) can be generalized by the introduction of an index i denoting subgroup-specific population sizes and catchabilities. Thus, the likelihood becomes

$$\Lambda = \prod_{i=1}^{I} \prod_{j=1}^{J} \frac{\lambda_{ij}^{C_{ij}} e^{-\lambda_{ij}}}{C_{ij}!} = \frac{(q_if_jN_{ij})^{C_{ij}} e^{-q_if_{ij}N_{ij}}}{C_{ij}!} . \quad (3)$$

Here, the N_{ij} are defined in a manner analogous to equation (2). That is, N_{i1} is the original number of animals in the population in subgroup i and

$$N_{ij} = N_{i1} - \sum_{k=1}^{j-1} R_{ik} \quad \text{for } j \geq 2 \qquad (4)$$

where R_{ik} is the number of animals removed from the ith subgroup of the population after the kth survey. Here, we have again treated the removals R_{ik} as known, fixed values.

Equation (3) has 2I unknowns: I initial abundances and I catchability coefficients. It is easily verified that maximizing (4) with respect to the 2I unknowns is equivalent to maximizing (1) separately for each subgroup in the population.

Method 2: introducing order restrictions for the catchabilities

Suppose we have good reason to believe that $q_1 < q_2 < ... q_I$ and we wish to introduce these order restrictions into the estimation procedure. Let $q_o = 0$ and let

$$q_i = q_{i-1} + \delta_i^2, \quad \text{for } i = 1, 2, .. I$$

and substitute these definitions into the likelihood (3). We now have I initial abundances and I values of δ_i^2 to estimate. Note that, regardless of the value of the estimate $\hat{\delta}_i$, the value of $\hat{\delta}_i^2$ must be non-negative and, thus, the estimate of q_i must be greater than or equal to the estimate of q_{i-1}.

Method 3: introducing a functional relationship for catchability

Often, the catchability of an animal will vary with the animal's body size. For example, the chances of a fish escaping through the meshes of a trawl generally decreases as the size of the fish increases. In contrast, the ability of some animals to avoid sampling gear may increase as the animals increase in age or size. In fisheries work, it is common to model the selectivity of fishing gear as a logistic function of body size. Thus, the proportion of the animals of length l that is retained by the gear, p(l), can be described by

$$p(l) = \frac{1}{1 + \exp(-\alpha(l - l_o))} \qquad (5)$$

where α is a shape parameter and l_o is a location parameter. The catchability of animals in the ith size group would then be proportional to the selectivity for that group:

$$q_i = \beta\, p(l) \qquad (6)$$

where β is an additional parameter relating the catchability to the selectivity of the gear. Equation (6) can be substituted for the q_i in equation (3). In this case, we estimate the parameters α and l_o of the logistic curve and the scaling parameter β, instead of the I catchabilities, q_i.

Method 4: using independent estimates of gear selectivity

It is often the case that the gear selectivity can be estimated by comparing the catches from two sampling gears with different mesh sizes. If gear selectivity parameter estimates are available from an independent study then one need only estimate the initial population sizes, N_{i1}, and the scaling parameter, β, of the logistic curve.

Discussion

The methods described here can be used as part of a general model building strategy. One can start by looking at separate estimates of catchability by size group. If these show a general trend or pattern that is consistent with expectation based on knowledge of the biology of the species and the characteristics of the sampling gear, then the estimates might be smoothed somewhat by imposing order restrictions. One might also estimate the parameters of a selectivity-with-size model. However, this would require sufficient contrast in the data, i.e., a sufficient range of sizes in the data. One could also use assumed selectivity parameters if these are available from an external study. A likelihood ratio test could be used to test if selectivity parameters estimated by the index-removal method are significantly different from assumed values from an external study.

Literature cited

Dawe, E.G., J.M. Hoenig and X. Xu. 1993. Change-in-ratio and index-removal methods for population assessment and their application to snow crab (*Chionoecetes opilio*). Can. J. Fish. Aquat. Sci. 50 (in press).

Eberhardt, L.L. 1982. Calibrating an index using removal data. J. Wildl. Manage. 46:734-740.

Petrides, G.A. 1949. View points on the analysis of open season sex and age ratios. Trans. N. Amer. Wildl. Conf. 14:391-410.

Ricker, W.E. 1975. Computation and Interpretation of Biological Statistics of Fish Populations. Bull. Fish. Res. Bd. Can. 191.

Roseberry J.L. and A. Woolf. 1991. A comparative evaluation of techniques for analyzing white-tailed deer harvest. Wildl. Monogr. 117:1-59.

Routledge, R. 1989. The removal method for estimating natural populations: incorporating auxiliary information. Biometrics 45:111-121.

Seber, G.A.F. 1982. The Estimation of Animal Abundance and Related Parameters, 2nd edition. Macmillan, New York.