ENCYCLOPEDIA OF STATISTICAL SCIENCES

UPDATE VOLUME 2

© 1998

A WILEY-INTERSCIENCE PUBLICATION

John Wiley & Sons, Inc.

NEW YORK • CHICHESTER • WEINHEIM • BRISBANE • SINGAPORE • TORONTO

[35] Tanner, M. A. and Wong, W. H. (1987). The calculation of posterior distributions by data augmentation. *J. Amer. Statist. Ass.*, **82**, 528–550.

(DATA AUGMENTATION
GIBBS SAMPLING
INCOMPLETE DATA
MISSING DATA, TYPES OF)

DONALD B. RUBIN
NATHANIEL SCHENKER

INDEX-REMOVAL METHODS

Index-removal methods can be used to estimate the size of an animal population by making use of the change in an index of population size caused by a known (or estimated) removal. The value of the index is assumed to be proportional to the population size. Suppose, for example, that the catch per unit of sampling effort is 30 (e.g., 30 animals per net haul) before 2000 animals are removed and is 10 after the animals are removed. Then we infer that the removal of 2000 animals reduced the abundance to one-third of its original value, because the index of abundance was reduced to a third. Consequently, there must have been 2000 × 3/2 = 3000 animals initially. (This requires several assumptions discussed later.) The method was apparently first proposed by Petrides [14] and is closely related to *catch-effort removal* methods and *change-in-ratio** methods. Open-population versions with multiple age groups are an example of *catch-at-age* analysis. (Open-population models allow for mortality and emigration, recruitment and immigration.)

CLOSED-POPULATION MODELS

Two Indices, One Removal

Assume that

1. the population is closed (i.e., constant) except for the removal(s),
2. all animals have the same probability of capture in the surveys, and this probability does not vary from survey to survey,
3. sampling is with replacement, or the fraction of the population taken in the surveys is negligible.

Consider first the case where there are two surveys and one removal. Let $E(\hat{c}_i)$ be the expected catch per unit of sampling effort at time i, N_0 be the initial population size (before the removal), and R be the removal (between surveys). Then the ratio of the expected values of the catch rates is equal to the ratio of the population sizes before and after the removal, so

$$\frac{E(\hat{c}_1)}{E(\hat{c}_2)} = \frac{qN_0}{q(N_0 - R)} = \frac{N_0}{N_0 - R},$$

where q, the *catchability coefficient*, is the constant of proportionality relating expected catch rate to population size. Thus, q is the fraction of the population captured by one randomly placed unit of sampling effort, provided this fraction is small. An estimate of the initial population size can be obtained by the method of moments* as

$$\hat{N}_0 = \hat{R}\frac{\hat{c}_1}{\hat{c}_1 - \hat{c}_2}, \quad (1)$$

where $\hat{c}_i$ is the estimated catch per unit effort in survey i, and $\hat{R}$ is the removal (possibly estimated). This is also the maximum likelihood estimate (MLE) when the catch rates (and removal, if estimated) are MLEs. Also, the catchability coefficient can be estimated by

$$\hat{q} = \hat{c}_1/\hat{N}_0.$$

Another parameter of interest is the *exploitation rate*, u, the fraction of the population harvested. This can be estimated [9] by

$$\hat{u} = \frac{\hat{R}}{\hat{N}_0} = \frac{\hat{c}_1 - \hat{c}_2}{\hat{c}_1}.$$

Note that the removal need not be known to estimate the exploitation rate.

The estimated variances of $\hat{N}_0$ and $\hat{q}$ can be obtained by the Taylor's series (delta) method (*see* STATISTICAL DIFFERENTIALS) as [3]

$$\hat{V}(\hat{N}_0) = \frac{\hat{N}_0^2\,\hat{V}(\hat{c}_2) + (\hat{N}_0 - \hat{R})^2\hat{V}(\hat{c}_1)}{(\hat{c}_1 - \hat{c}_2)^2} + \frac{\hat{c}_1^2\,\hat{V}(\hat{R})}{(\hat{c}_1 - \hat{c}_2)^2} - \frac{2\hat{N}_0(\hat{N}_0 - \hat{R})\widehat{\mathrm{Cov}}(\hat{c}_1, \hat{c}_2)}{(\hat{c}_1 - \hat{c}_2)^2},$$

$$\hat{V}(\hat{q}) = \frac{\hat{V}(\hat{c}_1) + \hat{V}(\hat{c}_2)}{\hat{R}^2} + \frac{(\hat{c}_1 - \hat{c}_2)^2\hat{V}(\hat{R})}{\hat{R}^4} - \frac{2\,\widehat{\mathrm{Cov}}(\hat{c}_1, \hat{c}_2)}{\hat{R}^2}$$

where the variances and covariance of the indices depend on the survey design, and the variance of the estimated removal depends on the particular sampling design. These are equivalent to the estimates obtained by inverting the Fisher information* matrix if the observations are simple random samples from distributions in the exponential family* (Chen et al. [3]). The covariance term is zero if sampling stations are rerandomized for the second survey. Chen et al. point out that when the spatial pattern is persistent over time (areas of high abundance tend to stay high, and areas of low abundance stay low), then the covariance of the two catch rates will be positive if the same stations are occupied in both surveys; thus, the covariance term will reduce the variance.

For example, they found [3] that the correlation of trap catches of legal-sized snow crabs at 53 sampling locations in a Newfoundland bay before the commercial fishery with the catches at the same stations after the commercial fishery four weeks later was +0.72. The catch rate declined from 61.3 crabs per trap before the fishery to 44.5 crabs per trap after the fishery, which removed 0.58×10^6 crabs. By (1), they estimated the population size before the fishery to be 2.1008×10^6 crabs. The estimated variances were $\hat{V}(\hat{c}_1) = 18.178$, $\hat{V}(\hat{c}_2) = 13.002$, and $\widehat{\mathrm{Cov}}(\hat{c}_1, \hat{c}_2) = 11.121$. The standard error of the population size, calculated without the covariance term, was 0.31×10^6; inclusion of the covariance term reduced the standard error by 47%.

If the population is not closed (assumption 1), then the change in catch rates between the two sampling times will not reflect the change in the original (initial) population. For example, if recruitment occurs between sampling times, then the decline in catch rate will be underestimated and the initial population size will be overestimated. The assumption of equal probability of capture for all animals in the survey fails when catchability varies with individual characteristics such as animal size. This causes a bias in the estimates if the removal is selective with respect to size. One way to avoid this is to make separate estimates by size group, sex, or other factor that may affect catchability. If the sampling fraction is substantial and sampling is without replacement, then assumption 3 fails and the situation becomes a removal estimation scheme.

Extensions and Relationships

Routledge [15] generalized the method to allow for I removals and $I + 1$ surveys ($I \geqslant 1$). He presented two likelihood functions. One assumes each survey catch is a Poisson random variable with parameter λ_i proportional to the (remaining) population size. The other assumes the catch in survey i is normally distributed with mean and variance proportional to population size. Routledge studied a laboratory population of cockroaches and used the number of scats (fecal pellets) left on squares of filter paper as the index of abundance.

Another possibility for generalizing the method is to allow for multiple types or classes of animal in the population, e.g., different age classes. Udevitz and Pollock [16] showed that equation (1) provides the maximum likelihood estimators for each age class if assumptions 1 through 3 hold for each age class separately. Chen et al. [3] suggest imposing order restrictions, or incorporating a model of catchability, when catchability appears related to the size of the animals.

The index-removal method is similar to the removal method. The latter makes use of the decline in catch rate in the harvest over time as the population is depleted. For example, one might set a large net or trap three nights in a row, and empty it each day (the animals are not released back into the population). Then one would expect the first catch to be the highest and the third the lowest, due to the successive depletion of the population. For two (observed) removals R_1 and R_2, the MLE is [18]

$$\hat{N} = R_1 \frac{R_1}{R_1 - R_2} = \frac{R_1^2}{R_1 - R_2},$$

of the same form as (1). The difference in the methods is that the monitoring and depletion of the population are separate activities in the index-removal method, whereas in the removal method they take place simultaneously.

Routledge [15] recognized that one could collect data for both index-removal and removal estimation in the same study and showed how the likelihoods for each method could be combined. Dawe et al. [7] showed that index-removal estimation could be combined with change-in-ratio estimation. The latter divides the population into two distinct types of animal, say x and y, and makes use of the change in population composition due to a known selective harvest. Chen [2] combined all three methods in one likelihood framework. The removals are modeled unconditionally as multinomial; the relative abundances are Poisson, conditional on the removals; and the proportion that is type x is binomial, conditional on the index.

OPEN POPULATION MODELS

Open population models for index-removal estimation are based on a structural model (first-order difference equation*) which assumes that the catch is taken in a short period of time during which there is no natural mortality. Thus,

$$N_{t+1} = (N_t - C_t + P_t)S_n, \qquad t = 1, \ldots, T, \tag{2}$$

where N_t is the number of catchable or recruited animals present at the start of year t, C_t is the catch of recruited animals in year t, P_t is the number of prerecruits at the start of year t (i.e., animals that will be recruited at the start of year $t + 1$), and S_n is the probability of surviving natural (nonfishing or nonhunting) mortality during the year. There is an implicit assumption in (2) that the catch occurs at the beginning of the year; the structural model can be modified to allow for catch occurring at other times of the year (Collie and Sissenwine [5]). The catches and the natural mortality are generally assumed known, although in theory it is possible to estimate the natural mortality.

The population abundance N_t and the abundance P_t of prerecruits are unknown. It is assumed that indices of abundance (n_t and p_t, respectively) can be obtained for N_t and P_t by means of a survey. (For example, the mean number of recruited animals per net haul might be an index for N_t.) Thus, it is assumed that the expected values of the indices are

$$E(p_t) = q_p P_t, \tag{3}$$

$$E(n_t) = q_n N_t, \tag{4}$$

where q_p and q_n are catchability coefficients (assumed constant) for prerecruits and recruited animals, respectively. Combining (3) and (4) with (2) gives

$$E(n_{t+1}) = E\left[\left(n_t - q_n C_t + \frac{q_n}{q_p} p_t\right) S_n\right].$$

This model can be fitted using ordinary least squares. However, current practice is to allow for separate errors in the indices of abundance as well as in the structural model. This is an example of a measurement error model in classical statistics (see, e.g., Fuller [10]).

To estimate the parameters, let

$$n_{t+1} = \left(n_t - q_n C_t + \frac{q_n}{q_p} p_t\right) S_n e^{\varepsilon_t}, \tag{5}$$

$$n_t = n_t^* e^{\kappa_t}, \tag{6}$$

$$p_t = p_t^* e^{\delta_t}. \tag{7}$$

The equations (6) and (7) relate the observed index values to the true values (denoted with an asterisk). Here, ε_t, κ_t, and δ_t are i.i.d. random errors distributed normally with mean zero. There are $2T + 1$ parameters to be estimated: n_t^* for each year, p_t^* for each year except the last, q_n, and q_p. However, in practice, the ratio q_n/q_p is usually either set equal to unity (Collie and Sissenwise [5]) or to an assumed value (Conser [6]). Parameters are estimated by minimizing the following sum of squares:

$$SS = \lambda_\varepsilon \sum_{t=2}^{T} e_t^2 + \sum_{t=1}^{T} k_t^2 + \lambda_\delta \sum_{t=1}^{T-1} d_t^2,$$

where e_t, k_t, and d_t are the residuals from (5), (6), and (7), respectively, and λ_ε and λ_δ are relative weights for the different sources of error. These weights are in practice chosen somewhat arbitrarily, and sensitivity analysis is used to determine if the choice of weights is critical. There are 3T-2 residuals, so when $2T$ parameters are estimated, there are T-2 degrees of freedom. Collie and Kruse [4] have fitted the model assuming errors in the indices but no process error in the structural model, in order to reduce the number of parameters estimated.

Collie and Sissenwine [5] presented a similar model for use when the survey data and catches can be divided into age groups. Walters and Collie [17] simplified the age-structured model by assuming a single type of error and casting the model in the form of a linear regression. Open population index-removal models with multiple age groups are a simple approach to a general class of problem known as *catch-at-age analysis* [12, 13].

Collie and Sissenwine [5] and Conser [6] trace the origin of models based on (2), (3) and (4) to Allen's development [1] of an open-population removal estimator, which Allen viewed as a generalization of DeLury's work [8]. For this reason, the open-population index-removal models are sometimes called modified deLury estimators. However, these equations follow more closely the earlier work of Leslie and Davis [11]. DeLury, like Leslie and Davis, developed a closed-population removal model, but DeLury's was a continuous model that utilized catch and fishing effort data, whereas Leslie and Davis' model was discrete and required catch and catch-rate data. Both models used the same fishing or hunting activity to monitor and deplete the population, unlike Petrides' model [14], which separates these activities. For this reason, the open-population models discussed here may best be considered as extensions of Petrides' index-removal estimation scheme.

Open-population index-removal methods are becoming prominent in fisheries research [6, 4]. This appears to be a topic deserving of theoretical development. For example, current models treat the (commercial) catches as fixed. However, these could be treated as random variables with expectation a function of the population at the beginning of the year and the fishing effort during the year. This would result in an open population analogue of Routledge's combined removal–index-removal estimator [15].

References

[1] Allen, K. R. (1966). Some methods for estimating exploited populations. *J. Fish. Res. Bd. Can.*, **23**, 1553–1574.

[2] Chen, C.-L. (1995). New Developments in Change-in-Ratio Estimation of Population Size. Ph.D. dissertation, North Carolina State University, Raleigh.

[3] Chen, C.-L., Hoenig, J. M., Dawe, E. G, Brownie, C., and Pollock, K. H. (1997). New developments in change-in-ratio and index-removal methods, with application to snow crab (*Chionoecetes opilio*). *Can. Spec. Pub. Fish. Aquat. Sci.*, **125**, in press.

[4] Collie, J. S. and Kruse, G. H. (1997). Estimating king crab abundance from commercial catch and research survey data. *Can. Spec. Pub. Fish. Aquat. Sci.*, **125**, in press.

[5] Collie, J. S. and Sissenwine, M. P. (1983). Estimating population size from relative abundance data measured with error. *Can. J. Fish. Aquat. Sci.*, **40**, 1871–1879.

[6] Conser, R. J. (1994). Stock Assessment Methods Designed to Support Fishery Management Decisions in Data-Limited Environments: Development and Application. Ph.D. dissertation, University of Washington. Seattle.

[7] Dawe, E. G., Hoenig, J. M., and Xu, X. (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**, 1467–1476.

[8] DeLury, D. B. (1947). On the estimation of biological populations. *Biometrics*, **3**, 145–167.

[9] Eberhardt, L. L. (1982). Calibrating an index by using removal data. *J. Wildlife Manage.*, **46**, 734–740.

[10] Fuller, W. A. (1987). *Measurement Error Models*. Wiley, New York.

[11] Leslie, P. H. and Davis, D. H. S. (1939). An attempt to determine the absolute number of rats on a given area. *J. Animal Ecol.*, **8**, 94–113.

[12] Megrey, B. A. (1989). Review and comparison of age-structured stock assessment models from theoretical and applied points of view. *Amer. Fish. Soc. Symp.*, **6**, 8–48.

[13] Mohn, R. K. and Cook, R. (1993). *Introduction to Sequential Population Analysis*, Sci. Council Studies No. 17. Northw. Atl. Fish. Org., Dartmouth, Nova Scotia, Canada.

[14] Petrides, G. A. (1949). Viewpoints on the analysis of open season sex and age ratios. *Trans. N. A. Wildlife Conf.*, **14**, 391–410.

[15] Routledge, R. D. (1989). The removal method for estimating natural populations: incorporating auxiliary information. *Biometrics*, **45**, 111–121.

[16] Udevitz, M. S. and Pollock, K. H. (1995). Using effort information with change-in-ratio data for population estimation. *Biometrics*, **51**, 471–481.

[17] Walters, C. J. and Collie, J. S. (1989). An experimental strategy for groundfish management in the face of large uncertainty about stock size and production. *Can. Spec. Pub. Fish. Aquat. Sci.*, **108**, 13–25.

[18] Zippin, C. (1956). An evaluation of the removal method of estimating animal populations. *Biometrics*, **12**, 163–169.

(CAPTURE-RECAPTURE
CHANGE-IN-RATIO ESTIMATORS
DISTANCE SAMPLING
MULTIVARIATE RATIO ESTIMATORS
RATIO ESTIMATORS
WILDLIFE SAMPLING)

JOHN M. HOENIG
KENNETH H. POLLOCK

CONCLUSION

The CIR method has not been used very much in practice; however, we expect this will change in the future. Recent theoretical developments have made the models more realistic. Also, the current emphasis in development of modern methods of estimating abundance of animal populations is the combination of more than one approach to allow model checking and increased precision. CIR is thus one easy-to-apply method that could be combined with index-removal, removal, catch-effort, or capture–recapture methods.

References

[1] Chapman, D. G. (1955). Population estimation based on change of composition caused by selective removal. *Biometrika*, **42**, 279–290.

[2] Chapman, D. G. and Murphy, G. I. (1965). Estimates of mortality and population from survey-removal records. *Biometrics*, **21**, 921–935.

[3] Chen, C. L. (1995). Change-in-Ratio Methods. Ph.D. thesis, Department of Statistics, North Carolina State University, Raleigh, N.C.

[4] Conner, M. C., Lancia, R. A., and Pollock, K. H. (1986). Precision of the change-in-ratio technique for deer population management. *J. Wildlife Manage.*, **50**, 125–129.

[5] Dawe, E. G., Hoenig, J. M., and Xu, X. (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**, 1467–1476.

[6] Eberhardt, L. L. (1982). Calibrating an index by using removal data. *J. Wildlife Manage.*, **46**, 734–740.

[7] Hanson, W. R. (1963). Calculation of productivity, survival and abundance of selected vertebrates from sex and age ratios. *Wildlife Monogr.*, **9**, 1–60.

[8] Heimbuch, D. G. and Hoenig, J. M. (1989). Change-in-ratio estimates for habitat usage and relative population size. *Biometrics*, **45**, 439–451.

[9] Hoenig, J. M., Pepin, P., and Lawing, W. D. (1990). Estimating relative survival rate for two groups of larval fishes from field data: do older larvae survive better than young? *Fishery Bull. U.S.*, **88**, 485–491.

[10] Kelker, G. H. (1940). Estimating deer populations by a differential hunting loss in the sexes. *Proc. Utah Acad. Sci. Arts and Lett.*, **17**, 6–69.

[11] Lancia, R. A., Nichols, J. D., and Pollock, K. H. (1994). Estimating the number of animals in wildlife populations. In *Research and Management Techniques for Wildlife and Habitats*, T. A. Bookout, ed., 5th ed. Wildlife Society, Bethesda, Md. pp. 215–253.

[12] Laplace, P. S. (1786). Sur les naissances, les mariages et les morts. In *Histoire de l'Académie Royale des Sciences, Aimée 1783*, Paris, p. 693.

[13] Otis, D. L. (1980). An extension of the change-in-ratio method. *Biometrics*, **36**, 141–147.

[14] Paulik, G. J. and Robson, D. S. (1969). Statistical calculations for change-in-ratio estimators of population parameters. *J. Wildlife Mange.*, **33**, 1–27.

[15] Petrides, G. A. (1949). Viewpoints on the analysis of open season sex and age ratios. *Trans. North Amer. Wildlife Conf.*, **14**, 391–410.

[16] Pollock, K. H. (1991). Modeling capture, recapture, and removal statistics for estimation of demographic parameters for fish and wildlife populations: past, present and future. *J. Amer. Statist. Ass.*, **86**, 225–238.

[17] Pollock, K. H., Lancia, R. A., Conner, M. C., and Wood, B. L. (1985). A new change-in-ratio procedure robust to unequal catchability of types of animal. *Biometrics*, **41**, 653–662.

[18] Rupp, R. S. (1966). Generalized equation for the ratio method of estimating population abundance. *J. Wildlife Manage.*, **30**, 523–526.

[19] Seber, G. A. F. (1982). *The Estimation of Animal Abundance and Related Parameters*, 2nd ed. Macmillan, New York.

[20] Udevitz, M. S. (1989). Change-in-Ratio Estimators for Estimating the Size of Closed Populations. Ph.D. thesis, Department of Statistics, North Carolina State University.

[21] Udevitz, M. S. and Pollock, K. H. (1991). Change-in-ratio estimates for populations with more than two subclasses. *Biometrics*, **47**, 1531–1546.

[22] Udevitz, M. S. and Pollock, K. H. (1995). Using effort information with change-in-ratio data for population estimation. *Biometrics*, **51**, 000–000.

(ADAPTIVE SAMPLING
CAPTURE–RECAPTURE
DISTANCE SAMPLING
TRANSECT METHODS
WILDLIFE SAMPLING)

KENNETH H. POLLOCK
JOHN M. HOENIG

CHANGE-OVER TRIALS *See* CHANGE-OVER DESIGNS; CROSSOVER TRIALS