

ESTIMATING ANIMAL ABUNDANCE WITH CAPTURE FREQUENCY DATA

ANNE CHAO, Institute of Applied Mathematics, National Tsing Hua University, Hsinchu, Taiwan, 30043, Republic of China

Abstract: I describe a new method for estimating the number of animals in a closed population with capture-recapture data under a heterogeneity model. The model is applied to capture frequency records of striped skunk (*Mephitis mephitis*), eastern chipmunk (*Tamias striatus*), eastern cottontail (*Sylvilagus floridanus*), and taxicab populations. I also report the results of a Monte Carlo simulation to assess the relative merits of the proposed moment estimator and the jackknife estimator. If many individuals are caught more than twice, the jackknife estimator is superior to the proposed moment estimator. However, when the mean capture probability is small, so that most captured animals are caught only 1–2 times in the samples, the moment estimator is usually less biased than the jackknife estimator. The mean coverage probability of the confidence interval associated with the moment estimator is also closer to the nominal level than that of the jackknife interval.

J. WILDL. MANAGE. 52(2):295–300

My study considers the problem of comparing estimators of population size in a multiple capture-recapture experiment under the model that incorporates heterogeneity of capture probabilities; i.e., individuals have different but constant probabilities of capture over trapping samples (Burnham and Overton 1978, 1979). The population is assumed to be closed so that there are no changes due to birth, death, emigration, or immigration during the sampling period. Refer to Cormack (1968, 1979), Otis et al. (1978), and Seber (1982, 1986) for a general review of various models and estimators currently in use.

Most of the previous methods for deriving population estimates generally require that all animals have the same capture probabilities. Edwards and Eberhardt (1967), Carothers (1973), Otis et al. (1978), and Burnham and Overton (1979) have examined the biases arising from a failure of this assumption and confirmed that heterogeneity can cause substantial bias in the estimators when assuming equal catchability. For example, Edwards and Eberhardt (1967) conducted a livetrapping experiment on a confined eastern cottontail population of known size. They found that the usual Schnabel estimates (Schnabel 1938, Schumacher and Eschmeyer 1943) were considerably lower than the true population size. A similar finding was reported by Carothers (1973), who conducted a unique study on the taxicab population, in Edinburgh, United Kingdom. Mares et al. (1981) estimated the abundance of eastern chipmunks using a capture-recapture technique and found that 3 commonly used estimators always underestimated the known abundance. Greenwood et al.

(1985) recently provided a data set from striped skunk populations of known sizes. They made numerical comparisons of 10 different estimators under several models and found all underestimated true population size. These data sets will be considered in the following sections.

Burnham and Overton (1978, 1979) introduced a heterogeneity model by assuming that the individual capture probability is a random sample from an arbitrary distribution function. They also derived the jackknife estimators (of several orders), which have been extensively studied by Otis et al. (1978) using many real and simulated data sets. Pollock and Otto (1983) subsequently supported the use of the jackknife estimator based on some numerical comparisons. Greenwood et al. (1985) also concluded that the jackknife estimator produced the best estimates of skunk abundance. In general, the jackknife estimator works well if most individuals are captured many times. However, if many animals have very small capture probabilities, the jackknife estimator will have a negative bias, as will every other known estimator (White et al. 1982:65).

Chao (1984, 1987) derived an alternative approach more appropriate for data sets with only low-order capture frequencies. In the following sections, the new moment estimator will be reviewed and applied to real data sets (Edwards and Eberhardt 1967, Carothers 1973, Mares et al. 1981, Greenwood et al. 1985). A Monte Carlo simulation study that suggests guidelines for choosing an estimator is also provided.

I thank K. P. Burnham, R. J. Greenwood, and G. C. White for valuable suggestions and com-

Table 1. Comparison of 4 population size estimators for striped skunks (Greenwood et al. 1985).

Capture frequencies	1977 (8 trap-nights)			1978 (4 trap-nights)		
	F	M	Both	F	M	Both
f_0	1	3	4	7	4	11
f_1	2		2	7	3	10
f_2	4	3	7	3	1	4
f_3	2	3	5	1		1
f_4	1	2	3	2		2
f_5		2	2			
S^a	9	10	19	13	4	17
Known N	10	13	23	20	8	28
Schnabel ^b	11.00 ± 3.65 ^c	10.91 ± 2.39	21.38 ± 3.79	14.82 ± 4.92	6.00 ± 29.13	20.33 ± 6.40
Jackknife ^d	10.75 ± 1.81	10.00 ± 0.00	20.75 ± 1.81	18.25 ± 3.03	6.25 ± 1.98	24.50 ± 3.62
Moment \hat{N}	9.50 ± 1.03	10.00 ± 0.00	19.29 ± 0.68	21.17 ± 8.28	8.50 ± 7.19	29.50 ± 10.68
Moment \tilde{N}	— ^e	—	—	21.56 ± 8.65	—	—

^a No. distinct individuals captured in samples.

^b Schnabel (1938).

^c Estimate ± SE.

^d Burnham and Overton (1978).

^e The condition $3f_1f_3 > f_2^2$ is not satisfied, so $\hat{N} = \tilde{N}$.

ments on this work. I also thank R. A. Lancia for suggesting some improvements in presentation. Part of this research was supported by the National Science Council of the Republic of China.

PROPOSED MOMENT ESTIMATOR

Assume that the true population size is N and there are t trapping samples. The heterogeneity model assumes that the j th animal has the capture probability p_j , which is constant over trapping samples, and p_1, p_2, \dots, p_N is a random sample from an arbitrary, unknown distribution function. Under this model, the capture frequencies (f_1, f_2, \dots, f_t), where f_k is the number of animals captured exactly k times in the t trapping samples, contains all the information for estimating the population size. For example, the jackknife estimator is a linear combination of capture frequencies.

When t is not too small (≥ 5) and most capture probabilities are relatively small, Chao (1987) demonstrated that if most animals are captured only 1–2 times in the samples, then a moment estimator of the population size is obtained as $\hat{N} = S + f_1^2/(2f_2)$, where S is the number of distinct animals captured in the t samples. Also, if f_1, f_2 , and f_3 further satisfy the conditions that $tf_1 > 2f_2$, $tf_2 > 3f_3$, and $3f_1f_3 > 2f_2^2$, then a modified moment estimator is $\tilde{N} = S + [f_1^2/(2f_2)][1 - 2f_2/(tf_1)]/[1 - 3f_3/(tf_2)]$. Otherwise, $\tilde{N} = \hat{N}$. All the derivation details are given in Chao (1987). It is clear that $\tilde{N} \geq \hat{N}$.

Both \hat{N} and \tilde{N} are referred to as moment estimators because they are obtained as lower

bounds of N subject to some moment constraints. Note that for \hat{N} or \tilde{N} , only function of the first 2 or 3 frequency counts are used to predict the number of uncaptured animals, f_0 . Intuitively, those animals with small capture probabilities are likely to be uncaptured (i.e., in the frequency f_0) or captured very few times. Thus, it seems reasonable that f_1, f_2 , and f_3 would contain most of the information about f_0 .

In practical applications, a confidence interval of the population size is more useful than a point estimate. To construct confidence intervals, the following variance estimators are used:

$$\begin{aligned} \widehat{\text{var}} \hat{N} &= f_2[0.25(f_1/f_2)^4 + (f_1/f_2)^3 \\ &\quad + 0.5(f_1/f_2)^2], \text{ and} \\ \widehat{\text{var}} \tilde{N} &= f_2[0.25A^2(f_1/f_2)^4 + A^2(f_1/f_2)^3 \\ &\quad + 0.5A(f_1/f_2)^2], \end{aligned}$$

where

$$A = [1 - 2f_2/(tf_1)]/[1 - 3f_3/(tf_2)].$$

Thus, a 95% confidence interval based on \hat{N} is obtained as $[S + (\hat{N} - S)/C, S + (\hat{N} - S)C]$ where $C = \exp\{1.96[\log(1 + (\widehat{\text{var}} \hat{N})/(\hat{N} - S)^2)]^{1/2}\}$. Similarly, one can construct confidence intervals based on \tilde{N} .

APPLICATIONS TO REAL DATA SETS

Striped Skunk Data

Greenwood et al. (1985) conducted a live-trapping capture–recapture experiment from a skunk population of known size. They calculated 10 estimators under various models and found that the (first order) jackknife estimator (Table

Table 2. Comparison of 6 population size estimators for eastern cottontails (Edwards and Eberhardt 1967).

Capture frequencies	1961 (18 trap-days)	1963 (8 trap-days)
f_0	59	61
f_1	43	36
f_2	16	15
f_3	8	13
f_4	6	3
f_5		1
f_6	2	1
f_7	1	
S^a	76	69
Known N	135	130
Geometric (1) ^b	136	151
Geometric (2) ^b	164	150
Schnabel ^c	85 ± 10.5 ^d	86 ± 11.0
Jackknife ^c	159 ± 21.9	130 ± 33.8
Moment \hat{N}	135 ± 24.0	112 ± 19.4
Moment \tilde{N}	136 ± 25.1	126 ± 25.3

^a No. distinct individuals captured in samples.
^b Edwards and Eberhardt (1967).
^c Schnabel (1938).
^d Estimate ± SE.
^e Burnham and Overton (1978).

Table 3. Comparison of 4 population size estimators for eastern chipmunks (Mares et al. 1981).

Capture frequencies	Data (13 trap-days)
f_0	10
f_1	14
f_2	13
f_3	18
f_4	12
f_5	7
f_6	5
f_7	1
f_8	1
S^a	72
Known N	82
Schnabel ^b	72.3 ± 5.75 ^c
Jackknife ^d	84.9 ± 4.88
Moment \hat{N}	79.5 ± 5.31
Moment \tilde{N}	81.5 ± 6.50

^a No. distinct individuals captured in samples.
^b Schnabel (1938).
^c Estimate ± SE.
^d Burnham and Overton (1978).

1) gave the best results of all methods evaluated based on 2 comparison criteria. From their data I calculated the moment estimates \hat{N} and \tilde{N} (Table 1). In this case, $\hat{N} = \tilde{N}$ except for females in 1978; therefore, the comparison will be mainly focused on \tilde{N} .

In 1977, a large portion (≥77%) of the skunks were captured in each data set and the jackknife estimator is accurate. Almost all estimation methods and the moment estimator agree favorably for these data. However, for 1978 the mean capture probability was relatively small and most skunks were caught 1–2 times. For 1978, the moment estimator appears reliable because its bias is relatively small, although the precision is low. The increase in standard error seems unavoidable because of the relatively low capture rates. All the other approaches are biased low.

Greenwood et al. (1985) compared estimators based on mean relative error and median standardized difference. The proposed \hat{N} has a smaller mean relative error (10%) and smaller median standardized difference (0.31) than the 14% and 0.93 values of the jackknife estimator. Thus, using these 2 criteria, the moment estimator is better than the jackknife estimator.

Eastern Cottontail Data

Edwards and Eberhardt (1967) used 2 data sets to illustrate the performance of various es-

timators. One data set was obtained in 1961 by livetrapping a confined eastern cottontail population for 18 consecutive days and the other was a field record of 8 trapping days obtained in 1963 (Table 2). Although the true population size is unknown for 1963 data, these authors conducted a drive-census and determined the estimated population size was 130 using the Lincoln-Petersen estimator.

Edwards and Eberhardt (1967) suggested fitting the capture frequencies by a geometric series and obtained 2 estimates for each data set by different estimation procedures (Table 2). Using their data, I compared \hat{N} , \tilde{N} , jackknife, and Schnabel estimates (Table 2).

As expected, the Schnabel estimator severely underestimates and the 2 geometric estimators consistently overestimate the true population size. This agrees with the finding of Carothers (1973) and Greenwood et al. (1985). As noted by Seber (1986:275) there is sufficient evidence from the literature to show that models such as geometric or negative binomial are inappropriate. The jackknife and the moment estimates seem reasonable for these data sets. For the 1961 data, Burnham and Overton (1978) suggested using the third-order jackknife estimate of 159. They further provided an improved, interpolated jackknife estimate of 142 but these estimates are worse than the moment estimate \hat{N} or \tilde{N} . For the 1963 data, the jackknife method works well, and the proposed \tilde{N} is biased low probably because the value of f_3 is relatively

