Estimating Population Size by Mark-resighting Counts: Theoretical Considerations and Empirical Results

La stima dell’abbondanza tramite marcatura e riavvistamento: considerazioni teoriche e risultati empirici

Marzia Marcheselli
Università di Siena, Dipartimento di Metodi Quantitativi, marcheselli@unisi.it

Caterina Pisani
Università di Siena, Dipartimento di Metodi Quantitativi, pisani4@unisi.it


Keywords: Mark-resighting counts, joint hypergeometric model, Minta-Mangel model.

1. Introduction

Mark-resighting methods have been proposed to estimate the abundance of closed populations on a study area. These procedures involve counting and identifying marked animals and recording the number of unmarked animals spotted during multiple subsequent resighting occasions. The advantage of mark-resighting with respect to capture-recapture procedures is that resightings are generally cheaper to acquire than physically capturing and handling the animals, especially when radiotelemetry or other tracking devices are at disposal. Thus, the overall field effort is decreased even if collected information is obviously reduced with respect to that obtained with capture-recapture methods. Accordingly the huge list of capture-recapture methodologies can not be adopted with resighting data and, at least to our knowledge, only two estimators of population abundance are available: the joint hypergeometric maximum likelihood estimator (JHE) (Bartmann et al., 1987, Neal et al., 1993) and the bootstrap estimator (MM) (Minta and Mangel, 1989). It is worth noting that the assumptions underlying these models are not well delineated in the literature. Moreover, in both cases a variance estimator is not available and confidence intervals are constructed directly from the likelihood function (Hudson, 1971) in the first case or from the bootstrapped distribution in the second. Program NOREMARK (White, 1996) has been developed to implement both the procedures. The aim of this paper is to outline the basic assumptions of the JHE and MM models, to propose some variance estimators and to investigate the robustness of the corresponding estimators under model mis-specifications.
2. JHE and MM estimators

Let $N$ denote the abundance of a closed population on a study area. The JHE estimate is the value of $N$ which maximizes

$$L(N,M,n_1,...,n_T|m_1,...,m_T) = \prod_{t=1}^{T} \frac{\binom{M}{m_t} \binom{N-M}{n_t-m_t}}{\binom{N}{n_t}}$$

where $M$ is the number of marked individuals in the population while $n_t$ and $m_t$ are the overall number of animals and the number of marked animals sighted at the $t$-th occasion ($t = 1, ..., T$). Despite Seber (1982, p. 59) and White and Garrott (1990, p.262), the joint hypergeometric model only holds under the very unreliable assumption that at the $t$-th occasion a fixed number of $n_t$ animals are selected from the population by simple random sampling without replacement and that sighting occasions are independent. Alternatively, Minta and Mangel (1989) suggest a bootstrap estimator of $U = N - M$ based on the sighting distribution of marked animals. If $n = n_1 + ... + n_T$ and $m = m_1 + ... + m_T$ and if $p_x = M_s/M$ ($x = 0, 1, ..., T$) where $M_s$ denotes the number of marked animals spotted $x$ times, then a sequence of iid random variables are generated from $p_0, p_1, ..., p_T$ until the total number of sightings equals or exceeds $u = n - m$. If the resulting waiting time, say $w$, is viewed as the number of unmarked animal necessary to generate $u$ resightings, then $B$ bootstrap samples may be generated and the mode of the bootstrap distribution may be taken as the estimate of $U$. In the NOREMARK software, White (1993) proposes a modified procedure which considers only bootstrap samples where the number of sightings exactly equals $u$ (note that in this case $w$ can take finite values with probability less than one). However, apart from the fact that there is no reason for performing a bootstrap procedure since the probability distribution of $w$ may be straightforwardly obtained, it is worth noting that in the Minta-Mangel procedure the role of the parameter and the random variable is reversed. Accordingly, the procedure cannot be accepted from an inferential point of view. Moreover, the assumptions underlying the procedure are not well delineated by the authors, even if it is at once apparent that the procedure makes sense only if the numbers of sightings for each animal, say $X_1, ..., X_N$, constitute a set of iid random variables from a (unknown) distribution $\pi_0, \pi_1, ..., \pi_T$. Indeed, in this case, given any subset $L$ of $L$ animals, the random vector $(L_0, ..., L_T)$, where $L_s$ denotes the number of animals sighted $x$ times, is multinomial with parameters $L$ and $\pi_0, \pi_1, ..., \pi_T$. Accordingly, $p_0, p_1, ..., p_T$ constitute the maximum likelihood estimates of $\pi_0, \pi_1, ..., \pi_T$ obtained from the set $M$ of the $M$ marked animals. Thus, since $u = \sum_{i \in U} X_i$ where $U$ denotes the set of the $U$ unmarked animals, then the profile likelihood of $u$ at $p_0, p_1, ..., p_T$ turns out to be $L(U, p_0, p_1, ..., p_T | u) = a_{u,U}$ where $a_{u,U}$ denotes the
coefficient of degrees $u$ in the polynomial \( \left( \sum_{x=0}^{T} s^x p_x \right)^U \). Thus, the profile estimate of $U$ may be straightforwardly obtained by numerical methods and the Hudson confidence interval may be obtained from the profile likelihood function. Even if the bootstrap and the profile procedures usually give rise to very similar estimates, in some cases the resulting estimates may be very different (see e.g. the case $T = 1$).

3. Variance estimation

Let $\hat{N}$ be the JHE estimator of $N$ based on $T$ resighting occasions and $\hat{N}_t$ be the well-known Lincoln-Petersen estimate obtained at the $t$-th occasion. Since, it seems quite natural to assume that $V(\hat{N}) \leq V(\hat{N}_t)$, then it may be straightforwardly proved that the sample variance of the $\hat{N}_t$, $(t = 1, \ldots, T)$ constitutes a conservative estimator of $V(\hat{N})$. Moreover, denote by $\tilde{U}$ the MM estimator of $U$ and by $\hat{\mu}$ and $\hat{\sigma}^2$ the expectation and variance with respect to $p_0, p_1, \ldots, p_T$. Owing to the central limit theorem, for $U$ sufficiently large it is apparent that $\tilde{U} \equiv u/\hat{\mu}$ in such a way that

\[
V(\tilde{U}) \approx \frac{(U + M)U}{M} \frac{\sigma^2}{\hat{\mu}^2}
\]

where $\mu$ and $\sigma^2$ represent expectation and variance with respect to $\pi_0, \pi_1, \ldots, \pi_T$. Thus, an asymptotically unbiased estimator of $V(\tilde{U})$ is simply obtained replacing $U$ by $\tilde{U}$ and $\mu$ and $\sigma^2$ by $\hat{\mu}$ and $\hat{\sigma}^2$. Obviously the procedure is likely to underestimate the variance when the $X_i$’s are positively correlated.

4. A simulation study

A simulation study was performed in order to investigate the performance of the JHE and MM estimators when the model assumptions are lacking. To this purpose, a population of $N = 100,500$ units was considered on a square region of size 1 and $T = 10,20$ resighting occasions were supposed. The number of marked animal was set to $M = 0.1N$ or $M = 0.3N$ and the population was randomly partitioned into $NG$ groups, where $NG$ was obtained by adding 1 to a binomial random variable with parameters $N - 1$ and $\gamma = \{E(NG) - 1\}/(N - 1)$, where $\gamma$ is such that $E(NG) = 2.10$. Group sizes were obtained by adding 1 to any component of a multinomial random vector with parameters $N - NG$ and $1/NG, \ldots, 1/NG$. The marked animals were randomly assigned to groups. Moreover in order to consider group movements, at the first occasion groups were randomly located on the square region, while in the $T - 1$ subsequent occasions, group movements were simulated by randomly changing the
spatial location inside a circle of radius $\rho = 0.1, 0.7$ centered at the previous location. Finally, at any occasion, a line transect census was simulated by using 4 transects of length 1 evenly placed on the lower side of the study area. Accordingly, the perpendicular distance of any group to the nearest transect was bounded by $1/8$. As to the sighting process, a group of size $g$ at distance $y$ from the transect was spotted with probability $g(y) = 1 - (8y)\beta^{-1}$ where $\beta$ was determined in such a way that the final probability of spotting a group was 0.3 or 0.7. When a group was spotted, all the animals in the group were supposed to be spotted.

For each of the 64 scenarios resulting from combining the values of $N, M, E(NG), T, \rho, \beta$, 10000 replications were performed. Results of the simulation motivate the following conclusion: (i) the lack of assumptions heavily impacts on the variance of the estimators rather then on their expectations; indeed the bias is usually negligible while the increase of variance gives rise to very low coverage for the Hudson confidence intervals; (ii) the conservative variance estimator proposed for the JHE model turns out to be effective, giving rise to confidence intervals with coverage very near to their nominal level; (iii) as expected, the asymptotic variance estimator proposed for the MM gives rise to a marked underestimation and poor coverage; however owing to the high levels of correlation between the JHE and MM estimates, the conservative estimator of variance proposed for the JHE model may be suitably adopted also for estimating the variance in the MM model.

References


