AISM (2007) 59: 603–615
DOI 10.1007/s10463-006-0065-5

Cibele Queiroz da-Silva

Asymptotics for a population size estimator of a partially uncatchable population

Received: 18 October 2004 / Revised: 7 December 2005 /
Published online: 22 August 2006

Abstract We present the asymptotic distribution for an estimator of the population size for the case of s partially catchable populations. Our approach is useful for capture–recapture studies with photo-identification data where part of the population does not have any distinctive characteristic which allows unique identification of the individuals. This work represents an extension of Theorem 4 in Sanathanan (1972, *The Annals of Mathematical Statistics, 43*, 142–152).

Keywords Capture–recapture · Closed uncatchable populations · Asymptotics

1 Introduction

Capture–recapture estimators of the size of a closed population require that a number of assumptions be satisfied if the estimates are to be unbiased. These assumptions are discussed in detail in Seber (1982). One of the main assumptions require homogeneity of animal behaviour with respect to catchability. The bias in the estimation of abundance resulting from heterogeneous capture probabilities is generally negative, i.e., the population size estimates tend to underestimate the true population size. In many instances characteristics of the individuals being captured are responsible for such heterogeneity. For example, older individuals may be easier to catch than younger ones, or females be easier to catch than males, etc. The bias resulting from lack of homogeneity can be minimized either by stratification of the data in order to obtain abundance estimates for more homogeneous sub-populations (Sekar and Deming, 1949; Rivest et al., 1995) or by modeling the capture probabilities as a function of variables related to the process (Alho, 1990; Huggins, 1989, 1991). In the first case, the sub-populations formed by classifying the

C. Q. da-Silva
Departamento de Estatística, IE/UnB,
70910-900 Brasília DF, Brazil
E-mail: cibeleqs@unb.br

animals according to variables influencing their capture probabilities are, actually, post-strata. Analyses on strata of the population can always be done, provided the attribute characteristic has been recorded and the sample sizes within strata are not too small.

This paper deals with the problem of estimating the size N of a closed population containing uncatchable individuals in a post-stratified photo-id Capture–recapture study. Besides that we evaluate the uncertainty of such an estimator. The framework that we consider is one in which the whole population is composed of s partially catchable demographically closed sub-populations. Our approach is useful for Capture–recapture studies where photo-id data are collected and part of the population does not have any distinctive characteristic which allows unique identification of the individuals (in such a case those individuals are then considered *uncatchable*). Estimation of bowhead whale (*Balaena mysticetus*) abundance using photo-id data is an example of such a problem. In the case of the bowhead whale it is not possible/practical to attach an artificial mark to the captured individuals, but the acquired natural marks throughout their lives are useful to allow the analyst to distinguish individuals. Contrary to the notion of a marked individual in Capture–recapture studies, a *marked bowhead* means that it has acquired natural marks enough to make reidentification possible. Since part of the population never acquire any natural marks, classical Capture–recapture estimators are not adequate to estimate the size of the whole population composed of naturally *marked* and *unmarked* individuals. For the bowhead whale, classes of maturity (mature/immature) which are related to the size of the animal, is a characteristic which allows post-stratification of the individuals. During the bowhead whale migration period, mature and immature whales tend to travel in different groups (age segregation). The earliest whales tend to be small, and the later migrants are mostly adults (mature whales). Such behaviour causes heterogeneity in capture probabilities which has to be dealt with adequate methods in the estimation of population abundance.

For such a problem, evaluation of uncertainty of abundance estimators has been tackled so far either through the use of bootstrap methods (da Silva et al., 2000) or through confidence regions (Schweder, 2003). da Silva et al. (2000) used parametric bootstrap methods to draw inferences to the bowhead whale population size. Their likelihood expressions do not belong to a regular family of distributions, preventing that variances of the estimators could be obtained via the standard large sample theory of maximum likelihood estimators. Besides that, other kinds of approximations were too complicated due to some covariance terms involved in the calculations. Schweder (2003) used confidence distribution to provide confidence intervals for the stratified by maturity classes bowhead whale population size. In that work the author also faces the problem of estimating uncertainty for abundance. As it can be noticed from Table 3 of Schweder (2003), estimated values for the standard errors for abundance are not reported. According to the author, uncertainty estimation was affected by bias in the population size estimates and skewness in confidence distribution.

In this paper we derive, as an extension of Sanathanan's result, an asymptotic distribution for the population size estimators of s partially catchable populations. Sanathanan (1972) derived an asymptotic theory for estimating the number of trials of a multinomial distribution from an incomplete observation of the total cells. Estimation of population size when a Capture–recapture experiment is undertaken is an example of such a problem, since cell totals are observed only for the cases where individuals are captured at least once over the sampling experiment.

In Sect. 2 we introduce some notation. In Sect. 3 we define a conditional likelihood based on good photos which incorporates information about the uncatchable part of the population. In Sect. 4 we present and prove a theorem for the asymptotic distribution of an estimator for the size of a population composed of s partially catchable closed sub-populations. In Sects. 5 to 7 we describe a simulation study and a model meant to compare the estimated standard errors for $\hat{N}$ using a derived asymptotic expression with those based on a parametric bootstrap method. In Sect. 8 we present the results and conclusions.

2 Notation

Quality of photos and extent of natural marks of an animal are important variables in our model formulation. A capture essentially means that a good quality photo of an individual was taken. In this case, if a natural mark (scars, white pigmentation patterns, etc) is found then the individual is considered marked. We now introduce some notation. Let s be the number of populations being considered, then for $t = 1, \ldots, s$,

- N_t^u : the total number of unmarked (uncatchable) individuals in population t.
- N_t^m : the total number of marked (catchable) individuals in population t.
- $N = \sum_{t=1}^{s} N_t$, with $N_t = N_t^m + N_t^u$: the total number of individuals in population.
- $\mathbf{N^m} = (N_1^m, \ldots, N_s^m)$: the vector of population sizes of marked individuals.
- $\Theta = (\theta_1, \ldots, \theta_r)$: the vector of independent parameters; $r < t(l-1)$.
- $\Psi = (\psi_1, \ldots, \psi_s)$: where $\psi_t = N_t^m / N_t$.
- X_a^t : the number of good photos of the individuals in population t at occasion a, $a = 1, \ldots, A$, where good photos are those from which the identification of the individuals is possible.
- x_a^t : the number of good photos of marked individuals in population t at occasion a, $a = 1, \ldots, A$.
- Ω: the set with 2^A elements where each element is a sequence of A binary components.
- n_{ti} : the total number of marked individuals in population t with capture history i, where i is a label for an element of Ω, with $i = 1, \ldots, l$.
- p_{ti} : the probability of an individual in population t having capture history i, where i is a label for an element of Ω, with $i = 1, \ldots, l$.
- n_t: the number of different individuals in population t that were captured over the experiment.

Let $(n_{t1}, \ldots, n_{tl})$ be distributed according to the multinomial law $M(N_t^m; p_{t1}, \ldots, p_{tl})$, with $p_{tl} = 1 - \sum_{i=1}^{l-1} p_{ti}$, with $p_{ti}(\Theta) = f_{ti}(\Theta)$, $i = 1, \ldots, l$ where f_{ti} are known functions. For example, $f_{ti}(\Theta)$ may be a logistic function.

Sanathanan (1972) showed that expressing the capture histories in terms of f_{ti} leads to estimability of population size. In the next section we present a conditional likelihood for the model which we are proposing and some definitions.

3 A conditional likelihood based on good photos

The model we are going to discuss involves a combination of s multinomial models factorized in the same fashion described by Sanathanan (1972) and s binomial models. The multinomial models account for the marked (catchable) part of the population while the binomial ones incorporate, through the number of good photos of unmarked individuals, information about the uncatchable part of the population. Next we present a conditional likelihood function based on good photos. The conditional likelihood of $(\mathbf{N^m}, \Psi, \Theta)$, given $\{X_a^t\}$ is

$$\begin{aligned}\mathcal{L} &= L\left(\mathbf{N^m}, \Psi, \Theta\right) = P\left(\{n_{t1}, \ldots, n_{tl}\}, \{x_a^t\} \mid \{X_a^t\}, \mathbf{N^m}, \Psi, \Theta\right)\\ &= P\left(\{n_{t1}, \ldots, n_{tl}\} \mid \mathbf{N^m}, \Theta\right) P\left(\{x_a^t\} \mid \{X_a^t\}, \Psi\right)\\ &= \prod_{t=1}^{s} \frac{N_t^m!}{(N_t^m - n_t)! \prod_{i=1}^{l-1} n_{ti}!} [p_{tl}(\Theta)]^{N_t^m - n_t} \prod_{i=1}^{l-1} [p_{ti}(\Theta)]^{n_{ti}}\\ &\quad \times \prod_{t=1}^{s}\prod_{a=1}^{A} \binom{X_a^t}{x_a^t} \psi_t^{x_a^t} (1-\psi_t)^{X_a^t - x_a^t}, \end{aligned} \tag{1}$$

where $p_{ti}(\Theta) = f_{ti}(\Theta)$, $i = 1, \ldots, l$, $n_{tl} = N_t^m - n_t$, and $n_t = \sum_{j=1}^{l-1} n_{tj}$. Thus, if $A = 3$, there are $l = 2^3 = 8$ possible capture histories: (1,1,1),(0,1,1),. . .,(0,0,0). For $t = 1, \ldots, s$, $N_t^m - n_t$ is the number of individuals in the population with capture history (0, 0, 0).

Let $\mathcal{L} = L(\mathbf{N^m}, \Theta)\, L(\Psi)$. According to Sanathanan (1972), $L(\mathbf{N^m}, \Theta)$ can be written as $L(\mathbf{N^m}, \Theta) = \prod_{t=1}^{s} L_{t1}\left(N_t^m, p_{tl}(\Theta)\right) L_{t2}(\Theta)$. Thus, let us write $\mathcal{L}$ as

$$\mathcal{L} = \prod_{t=1}^{s} L_{t1}\left(N_t^m, p_{tl}(\Theta)\right) L_{t2}(\Theta) L_{t3}(\psi_t) = L_1 \times L_2 \times L_3, \tag{2}$$

where

$$\begin{aligned} L_{t1}(N_t^m, p_{tl}(\Theta)) &= (N_t^m!/(n_t!(N_t^m - n_t)!))[1 - p_{tl}(\Theta)]^{n_t}[p_{tl}(\Theta)]^{N_t^m - n_t},\\ L_{t2}(\Theta) &= (n_t!/(n_t1! \ldots n_{tl-1}!))[q_{t1}(\Theta)]^{n_{t1}} \cdots [q_{tl-1}(\Theta)]^{n_{tl-1}} \quad \text{and}\\ L_{t3}(\psi) &= \prod_{a=1}^{A} \binom{X_a^t}{x_a^t} \psi_t^{x_a^t}(1-\psi_t)^{X_a^t - x_a^t}, \end{aligned}$$

with $q_{ti}(\Theta) = p_{ti}(\Theta)/(1 - p_{tl}(\Theta))$, $i = 1, \ldots, l-1$.

For $t = 1, \ldots, s$, $\hat{\psi}_t = \left(\sum_{a=1}^{A} x_a^t / \sum_{a=1}^{A} X_a^t\right)$, with $\hat{\psi}_t \rightarrow_{\text{a.s.}} \psi_t$. Following the same lines of Lemma 1 in Sanathanan (1972), for any p_t,

$$\hat{N}_t^m = n_t/(1-\hat{p}_{tl}), \qquad \hat{N}_t = n_t/(\hat{\psi}_t(1-\hat{p}_{tl})) = \hat{N}_t^m/\hat{\psi}_t \quad \text{and} \quad \hat{N} = \sum_{t=1}^{s} \hat{N}_t. \quad (3)$$

An unconditional MLE of N is obtained when there exists $\hat{N}_t^m$ of N_t^m and $\hat{\psi}_t$ of ψ_t for $t = 1, \ldots, s$, which simultaneously maximize $\mathcal{L}$ over all admissible values of $(\mathbf{N^m}, \Psi, \Theta)$. A conditional MLE of N, $\hat{N}_c$, is obtained when we find $\hat{N}_t^m$ maximizing $L_{t1}(N_t^m, \hat{p}_{tc})$ where $\hat{p}_{tc} = p_{tl}(\hat{\Theta}_c)$ and $\hat{\Theta}_c$ is the value of Θ maximizing $L_{t2}(\Theta)$.

In the following section we enunciate and prove a theorem which is an extention of Theorem 4 by Sanathanan (1972). Such theorem incorporates the uncatchable part of the population making possible that asymptoptic properties of the MLE's be evaluated for the whole population size estimator of N.

4 Case of s partially catchable populations

Let $\mathbf{N}_o^m$, Ψ_o, Θ_o, and $\mathbf{N}_o$, respectively, be the true values of $\mathbf{N}^m$, Ψ, Θ, and $\mathbf{N}$. For $t = 1, \ldots, s$, and $i = 1, \ldots, l$, let $p_{ti}(\Theta)$ be denoted by p_{ti}^o when $\Theta = \Theta_o$, and denoted by $\hat{p}_{ti}$ when $\Theta = \hat{\Theta}$. Similarly let the partial derivatives of $p_{ti}(\Theta)$ with respect to θ_j be denoted by $p_{ti,j}^o$ when $\Theta = \Theta_o$, and denoted by $\hat{p}_{ti,j}$ when $\Theta = \hat{\Theta}$. Let $L_{1j} = \partial \log L_1/\partial\theta_j$, $L_{2j} = \partial \log L_2/\partial\theta_j$, and $\hat{L}_j = \hat{L}_{1j} + \hat{L}_{2j}$.

Theorem 1 *Let the $p_{ti}(\Theta)$'s admit first order partial derivatives which are continuous at every admissible value Θ. Let $\hat{\mathbf{N}}^m = (\hat{N}_1^m, \ldots, \hat{N}_s^m)$, $\hat{\Psi} = (\hat{\psi}_1, \ldots, \hat{\psi}_s)$, $\hat{\Theta} = (\hat{\theta}_1, \ldots, \hat{\theta}_r)$, and $\hat{\mathbf{N}} = (\hat{N}_1, \ldots, \hat{N}_s)$ be the estimates of $\mathbf{N}_o^m$, Ψ_o, Θ_o, and $\mathbf{N}_o$ respectively such that*

(i) $\hat{\Theta} \rightarrow_{\text{a.s.}} \Theta_o$

(ii) $\hat{\Psi} \rightarrow_{\text{a.s.}} \Psi_o$

(iii) $\left\langle (N_{ot}^m)^{-1/2}\left(\hat{N}_t^m - n_t/(1-\hat{p}_{tl})\right)\right\rangle \rightarrow_{\text{a.s.}} 0$

(iv) $\left\langle (N_{ot}^m)^{-1/2}\left(\hat{N}_t - n_t/(\hat{\psi}_t(1-\hat{p}_{tl}))\right)\right\rangle \rightarrow_{\text{a.s.}} 0$

(v) $(N_T^m)^{-1/2}\hat{L}_j \rightarrow_{\text{a.s.}} 0, \quad j = 1, \ldots, r$

where $\langle e_t \rangle$ denotes the vector $(e_1, \ldots, e_s)$, and $N_T^m = \sum_{t=1}^{s} N_{ot}^m$. Let $\bar{\Sigma}^{-1} = (\sigma^{i,j})$ be the $(r+s) \times (r+s)$ matrix given by

$$\sigma^{j,h} = \sum_{t=1}^{s} c_t \sum_{i=1}^{l} \left[p_{ti}^o\right]^{-1} p_{ti,j}^o p_{ti,h}^o, \qquad j = 1, \ldots, r;\ h = 1, \ldots, r,$$

$$\sigma^{j,r+t} = -(c_t)^{1/2}\left[p_{tl}^o\right]^{-1} p_{tl,j}^o, \qquad j = 1, \ldots, r;\ t = 1, \ldots, s,$$

$$\sigma^{r+t,r+u} = \delta_{tu}\left[p_{tl}^o\right]^{-1}(1-p_{tl}^o), \qquad t = 1, \ldots, s;\ u = 1, \ldots, s,$$

where $\delta_{tu} = 1_{\{t=z\}}$. *Assume that*

$$\lim_{N_T^m \to \infty} \frac{N_{ot}^m}{N_T^m} = c_t, 0 < c_t < 1, \sum_{t=1}^{s} c_t = 1.$$

Then,

$$U^t = \left((N_T^m)^{1/2}(\hat{\Theta} - \Theta_o), (N_{o1}^m)^{-1/2}(\hat{N}_1 - N_{o1}), \ldots, (N_{os}^m)^{-1/2}(\hat{N}_s - N_{os}) \right) \quad (4)$$

is asymptotically $N(0, \bar{\Sigma})$. *Now assume*

$$\lim_{N_{ot}^m \to \infty} \lim_{\sum_{a=1}^{A} X_a^t \to \infty} \frac{1}{A} \sum_{a=1}^{A} \frac{x_a^t}{N_{ot}^m} = b_t,$$

then the asymptotic variance of $\hat{N}$ *is*

$$Var(\hat{N}) = \sum_{t=1}^{s} \frac{N_{ot}^m}{\psi_{ot}^2} \left[\sigma_{N_{ot}^m}^2 + \frac{1 - \psi_{ot}}{A b_t} \right], \quad (5)$$

where $\sigma_{N_{ot}^m}^2 = \tilde{\sigma}^{r+t,r+t}, \quad t = 1, \ldots, s$, *with* $\bar{\Sigma} = (\tilde{\sigma}^{i,j})$.

Proof Equation (4) follows directly from Theorem 4 by Sanathanan (1972). Equation (5) follows from the next development.
Since $N_t = N_t^m / \psi_t$, let

$$H_t = \left(N_{ot}^m\right)^{-1/2} (\hat{N}_t - N_{ot}) = \left(N_{ot}^m\right)^{-1/2} \left(\frac{\hat{N}_t^m}{\hat{\psi}_t} - \frac{N_{ot}^m}{\psi_{ot}} \right)$$
$$= \left(N_{ot}^m\right)^{-1/2} \left(\frac{\hat{N}_t^m - N_{ot}^m}{\hat{\psi}_{ot}} \right) + \left(N_{ot}^m\right)^{1/2} \left(\frac{1}{\hat{\psi}_t} - \frac{1}{\psi_{ot}} \right). \quad (6)$$

From equation (6), and since from condition $(ii)\, \hat{\psi}_t \to_{\text{a.s.}} \psi_{ot}$, and from expression (4), $\left(N_{ot}^m\right)^{-1/2} \left(\hat{N}_t^m - N_{ot}^m\right) \to_d \sigma_{N_{ot}^m} Z_1$, where $Z_1 \sim N(0, 1)$,

$$\left(N_{ot}^m\right)^{-1/2} \left(\frac{\hat{N}_t^m - N_{ot}^m}{\hat{\psi}_{ot}} \right) = \frac{\left(N_{ot}^m\right)^{-1/2} \left(\hat{N}_t^m - N_{ot}^m\right)}{\hat{\psi}_{ot}} \to_d \frac{\sigma_{N_{ot}^m} Z_1}{\psi_{ot}}.$$

Also from equation (6),

$$\left(N_{ot}^m\right)^{1/2}\left(\frac{1}{\hat{\psi}_t}-\frac{1}{\psi_{ot}}\right)=\frac{\left(N_{ot}^m\right)^{1/2}}{\left(\sum_{a=1}^A X_a^t\right)^{1/2}}\left(\sum_{a=1}^A X_a^t\right)^{1/2}\left(\frac{1}{\hat{\psi}_t}-\frac{1}{\psi_{ot}}\right)$$
$$=\left(\frac{\sum_{a=1}^A x_a^t}{\sum_{a=1}^A X_a^t}\right)^{1/2}\left(\frac{N_{ot}^m}{\sum_{a=1}^A x_a^t}\right)^{1/2}\left(\sum_{a=1}^A X_a^t\right)^{1/2}$$
$$\times\left(\frac{1}{\hat{\psi}_t}-\frac{1}{\psi_{ot}}\right). \tag{7}$$

In equation (7), from condition (ii), $\left(\sum_{a=1}^A x_a^t/\sum_{a=1}^A X_a^t\right)^{1/2} \rightarrow_{\text{a.s.}} (\psi_{ot})^{1/2}$, while from the delta method, $\left(\sum_{a=1}^A X_a^t\right)^{1/2}\left(\frac{1}{\hat{\psi}_t}-\frac{1}{\psi_{ot}}\right) \rightarrow_d -\frac{1}{(\psi_{ot})^2}(\psi_{ot}(1-\psi_{ot}))^{1/2}Z_2$, where $Z_2 \sim N(0,1)$. Besides that,

$$\lim_{N_{ot}^m\rightarrow\infty}\lim_{\sum_{a=1}^A X_a^t\rightarrow\infty}\left(\frac{N_{ot}^m}{\sum_{a=1}^A x_a^t}\right)^{1/2} \rightarrow (Ab_t)^{-1/2}.$$

Therefore,

$$\left(N_{ot}^m\right)^{1/2}\left(\frac{1}{\hat{\psi}_t}-\frac{1}{\psi_{ot}}\right)\rightarrow_d -\left(\frac{(\psi_{ot})^{-2}(1-\psi_{ot})}{Ab_t}\right)^{1/2} Z_2.$$

Thus, the asymptotic variance of H_t is given by

$$\operatorname{Var}\left(\frac{\sigma_{N_{ot}^m}Z_1}{\psi_{ot}}-\left(\frac{1-\psi_{ot}}{\psi_{ot}^2 Ab_t}\right)^{1/2} Z_2\right)=\frac{\sigma_{N_{ot}^m}^2}{\psi_{ot}^2}+\frac{1-\psi_{ot}}{\psi_{ot}^2 Ab_t}-\rho\frac{2\sigma_{N_{ot}^m}^2}{\psi_{ot}}\left(\frac{1-\psi_{ot}}{Ab_t}\right)^{1/2}, \tag{8}$$

where ρ is the correlation between Z_1 and Z_2, i.e., the correlation between $\hat{\psi}_t$ and $\hat{N}_t^m$. However, according to the likelihood equations (1) and (2), and the derived estimators of ϕ_t and N_t^m (see end of Sect. 3), this correlation is zero. Since we are assuming the sub-populations are independent, this completes the proof of the Theorem.

In the next section we describe a simulation experiment aimed to generate the data that was used to exemplify and evaluate the methods discussed so far.

5 Simulation experiment

For the simulations we considered partially uncatchable closed populations including two subgroups: the mature and immature individuals. For simplicity the Capture–recapture data included only 4 (four) sampling occasions. At a given time an individual is considered captured when it is photographed and it has at least one good quality photo.

Population size and catchability effects on the estimates of the asymptotic variance of $\hat{N}$ were taken into account in the simulations by considering the four *cases*: large N and two levels of capture probabilities (high and low) and not so large N and the same levels of capture probabilities.

The simulated data contain the following information: (1) four binary vectors that together indicate the capture histories of the captured individuals. For each time, 1 indicates a capture and 0 otherwise; (2) four columns indicating the number of good photos taken from each of the captured individuals in each of the sampling occasions; (3) one binary vector indicating whether or not the individual is naturally marked; (4) one binary vector indicating whether or not the individual is mature. (5) four columns indicating the sampling effort. Those values were similar for individuals captured at a given time and belonging to the same maturity class.

The following aspects were considered in the simulations:

(a) Population sizes: In order to evaluate the impact of the population size on the performance of the estimated asymptotic standard errors, we considered hypothetical populations of sizes $N = 15,000$ and $50,000$.

(b) Marked population and maturity: we fixed at 30% the percentage of unmarked individuals and at 60% the percentage of mature individuals in the population. For the unmarked population we fixed at 70% the percentage of immature individuals. Such procedure results in four categories of individuals: *marked mature*, *marked immature*, *unmarked mature* and *unmarked immature*. According to the stablished percentages, a marked individual is much more likely to be mature than an unmarked one. For the simulations we considered the percentages above to randomly assign each individual in the population to one of the four categories. The same percentages were used for each of the population sizes described in item (a).

(c) Shooting probabilities: high $(0.10, 0.15, 0.20, 0.24)$ and low $(0.05, 0.075, 0.10, 0.12)$ shooting probabilities for each of the hypothetical populations sizes were considered. For each sampling occasion we used the respective pre-assigned probabilities to randomly select the individuals in the population to be photographed. We did not use extremely high such proportions since those cases are of very limited practical interest. For the shooting probabilities we did not consider differences among groups of individuals characterized either by the extent of markings or maturity class. We assumed that based on the sampling protocol of bowhead whales which are photographed regardless their size (related to maturity) or markings.

(d) Good photos: In order to simulate the number of photos taken from each photographed individual we supposed that this variable follows a zero truncated Poisson distribution. The parameter of the Poisson distribution, ξ, was fixed according to some bowhead whale data. We wanted a value ξ that after truncation gives an average of 1.541 photos/whale. A value of $\xi = 0.938$ satisfies this criterion. We assumed that the probability of getting a good photo (fixed at 0.8) in occasion j is the same for marked and unmarked individuals. That is a reasonable assumption since it is in agreement with the analyst's procedure protocol to attach rank quality to a photo. The analyst gives a grade to each photo based only on photo quality, not on the individual's extent of marks. The number of good photos taken from a given photographed individual was generated according to a binomial distribution with parameters given by the total number of photos taken from that individual and $p = 0.8$.

(e) Effort data: For the simulations, effort data for each maturity class (our covariate) has been fixed as the proportion of marked captured individuals in each sampling occasion. We adopted such procedure in order to guarantee a very good description of the effort weights at each time. That would assure less biased population size estimates.

6 Modelling capture probabilities

For exemplifying the methods discussed in the previous sections, the model conceived for the probability of capturing an individual in population t at sampling occasion a, λ_{ta}, was a logistic one being described by

$$\lambda_{ta} = \frac{\exp(\theta_o + \theta_1 f_{ta})}{1 + \exp(\theta_o + \theta_1 f_{ta})}, \tag{9}$$

where f_{ta} represents the sampling effort for population t at sampling occasion a. In our case we have two sub-populations, the mature and the immature individuals.

Notice that the capture history probabilities $p'_{ti}s$ are written as a funtion of the $\lambda'_{ta}s$.

Considering equations (2) and (9), the log-likelihood function relating only the parameters N_t^m and (θ_o, θ_1) is given by expression (10) ahead. The following additional notation was used:

- y_{tj} is the number of individuals from population t that were captured only at time j, for $t = 1, 2$, and $j = 1, \ldots, 4$.
- $y_{t,mj}$ is the number of individuals from population t that were captured only at times m and j, for $t = 1, 2$, and $m < j \in \{1, \ldots, 4\}$. Notation is analogous for $y_{t,mjv}$ for $m < j < v \in \{1, \ldots, 4\}$, and $y_{t,1234}$.

Also let $d_{t1} = \sum_{i=1}^{4} y_{ti}$; $d_{t2} = \sum_{m=1}^{3} \sum_{j=m+1}^{4} y_{t,mj}$; $d_{t3} = \sum_{m=1}^{2} \sum_{j=m+1}^{3} \sum_{v=j+1}^{4} y_{t,mjv}$; $d_{t4} = y_{t,1234}$; $d_5 = \sum_{t=1}^{2} \sum_{i=1}^{4} y_{ti} f_{ti}$; $d_6 = \sum_{t=1}^{2} \sum_{m=1}^{3} \sum_{j=m+1}^{4} y_{t,mj}(f_{tm} + f_{tj})$;

$$d_7 = \sum_{t=1}^{2} \sum_{m=1}^{2} \sum_{j=m+1}^{3} \sum_{v=j+1}^{4} y_{t,mjv}(f_{tm} + f_{tj} + f_{tv}) \quad \text{and}$$

$$d_8 = \sum_{t=1}^{2} y_{t,1234} \sum_{i=1}^{4} f_{ti}.$$

Thus, the log-likelihood function relating only the parameters N_t^m and (θ_o, θ_1) is given by

$$\ell = -\sum_{t=1}^{2}\sum_{a=1}^{4} \log\left(1 + \exp\left(\theta_o + \theta_1 f_{ta}\right)\right) \left[(N_t^m - n_t) + \sum_{i=1}^{4} d_{ti} \right]$$
$$+\theta_o \sum_{t=1}^{2}\sum_{i=1}^{4} i d_{ti} + \theta_1 \sum_{k=5}^{8} d_k. \quad (10)$$

Notice that $n_t = y_{t1} + \cdots + y_{t4} + \cdots + y_{t,1234}$.

7 Variance estimation via bootstrap

In this section we describe a bootstrap procedure aimed to estimate uncertainty about $\hat{N}$. These estimates will be compared (in Sect. 8) to the respective asymptotic standard deviations estimated using expression (5).

We follow Buckland (1980) and others in using the parametric bootstrap to estimate standard error. In the parametric bootstrap setting we draw B samples of size n from the distribution $\hat{F}_{\text{par}}$, an estimate of F derived from a parametric model for the data. Where parameters were needed to specify the distribution, estimates of these parameters computed from the original data were used. The choice between nonparametric and parametric bootstrap in Capture–recapture is addressed by Buckland and Garthwaite (1991). They note that even though the nonparametric bootstrap is more widely used and more familiar than the parametric bootstrap, the latter allows us to choose which underlying distribution model to assume for the data.

For the model presented in Sects. 3 and 6, a parametric bootstrap approach for estimating variance of $\hat{N}$ involves the following steps:

1. Obtain the "original data" by running the data simulation program once. Consider as fixed the obtained sampling effort and the total number of good photos at each samppling occasion, $\{X_a^t\}$.
2. Using the data obtained in step 1 estimate the parameters $\{N_t^m\}$, θ_o, θ_1, $\{\psi_t\}$, and then $\{\lambda_{ta}\}$ and $\{p_{ti}\}$.
3. Considering expression (1) and the estimates $\{\hat{p}_{ti}\}$ and $\{\hat{N}_t^m\}$, simulate the number n_{ti} of individuals in population t with capture history i, and the number n_t of different individuals in population t that were captured over the experiment, i.e., draw $(n_{t1}^*, \ldots, n_{tl}^*)$ from the multinomial law $M(\hat{N}_t^m; \hat{p}_{t1}, \ldots, \hat{p}_{tl})$. Calculate the estimate N_t^{m*} (and the p_{ti}^*'s) from the bootstrap sample.
4. Simulate the number x_a^t of good photos for group t at time a from a Binomial $(X_a^t, \hat{\psi}_t)$ distribution. Obtain the estimates $\{\psi_t^*\}$ and N^*.
5. Repeat steps 3–5 B times and calculate de standard deviation of sample $N_1^*, \ldots,$ N_B^*.

In the above steps, * denotes data or an estimate from the bootstrap sample.

The determination of the number B of bootstrap replications depends on the application. Efron (1981) suggests that bootstrap estimates of standard error usually have relatively little bias, and very seldom more than $B = 200$ replications are needed for estimating a standard error. Many more replications are needed to obtain a good estimate of bias or for construction of confidence intervals. Buckland and Garthwaite (1991) advocate that for a 95% confidence interval, $B = 1,000$ should be satisfactory, whereas $B = 200$ is likely to be inadequate. In this work, for standard error estimation of $\hat{N}$, we used $B = 300$ bootstrap replications.

All the estimation and bootstrap procedures were performed using codes written in FORTRAN. Even so, the bootstrap procedure is very time consuming, taking about 30 min. for each of the 400 original simulated samples for estimating variance of $\hat{N}$. In this work we used a 512 Mb RAM memory 1.1 GHz pentium IV processor.

8 Results and conclusions

Using the procedure described in Sect. 5, one hundred simulated data sets for each of the four cases (two different population sizes, $N = 15,000$ and $50,000$, and two levels of capture probabilities, high and low) were produced. For each case of the 400 simulated samples we performed maximum likelihood estimation for the total number of individuals in the population, N, using expressions (3), (9) and (10). We also estimated the standard error of $\hat{N}$ using both expression (5) and the parametric bootstrap procedure described in Sect. 7.

Notation $s.d._{\text{Asy.}}(\hat{N})$ and $s.d._{\text{Boot.}}(\hat{N})$ in Table 1 stand, respectively, for the asymptotic and bootstrap standard deviations of $\hat{N}$.

For each case Table 1 summarizes some results for the estimated values of N and the estimated standard deviations of $\hat{N}$ based on the asymptotic and bootstrap approaches. As we can observe from lines 1, 4, 7 and 10 of Table 1, the average/median values of $\hat{N}$ are very close to the true N, especially for large values of capture probabilities, as one could suppose. For $N = 50,000$ we observe that the values of each column of the summary statistics of $s.d._{\text{Asy.}}(\hat{N})$ and $s.d._{\text{Boot.}}(\hat{N})$

Table 1 Descriptive statistics for $\hat{N}$ and estimated standard errors of $\hat{N}$ based on the asymptotic and bootstrap approaches — one hundred simulated samples were used for each of the four cases

Case	Estimate	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
	$\hat{N}$	43,550	47,800	48,850	49,030	50,450	54,080
$N = 50,000$; small $p_i's$	$s.d._{\text{Asy.}}(\hat{N})$	1,351	1,552	1,598	1,612	1,688	1,875
	$s.d._{\text{Boot.}}(\hat{N})$	1,296	1,501	1,565	1,566	1,640	1,815
	$\hat{N}$	47,270	49,320	49,910	49,960	50,650	52,110
$N = 50,000$; large $p_i's$	$s.d._{\text{Asy.}}(\hat{N})$	771	818	836	837	855	901
	$s.d._{\text{Boot.}}(\hat{N})$	724	790	820	821	856	922
	$\hat{N}$	12,680	14,270	14,760	14,780	15,250	17,950
$N = 15,000$; small $p_i's$	$s.d._{\text{Asy.}}(\hat{N})$	709	839	888	891	930	1,207
	$s.d._{\text{Boot.}}(\hat{N})$	733	917	989	983	1,031	1,319
	$\hat{N}$	13,680	14,700	15,040	15,060	15,400	16,180
$N = 15,000$; large $p_i's$	$s.d._{\text{Asy.}}(\hat{N})$	399	444	462	462	477	514
	$s.d._{\text{Boot.}}(\hat{N})$	432	486	508	510	529	608

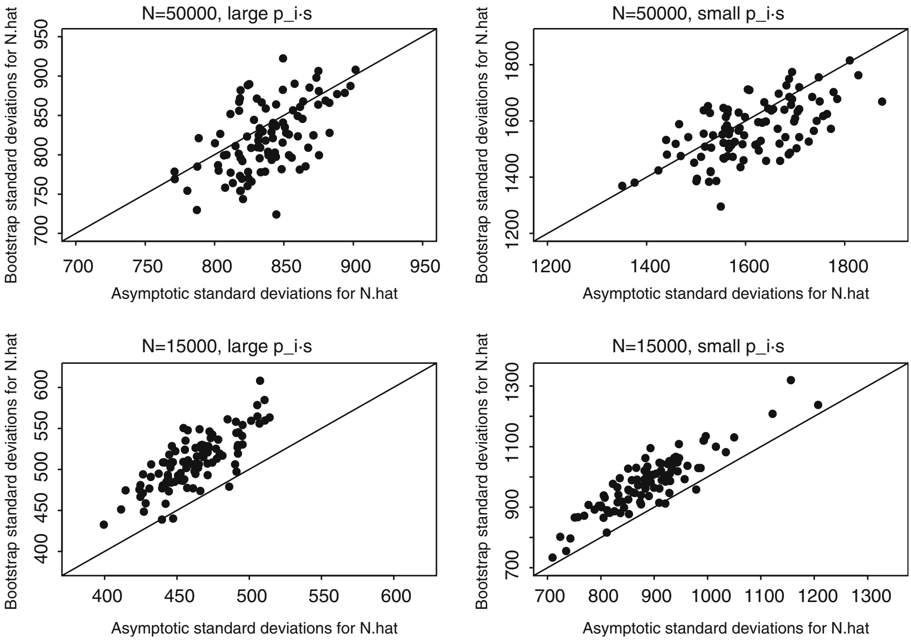

Fig. 1 Comparison between the asymptotic and bootstrap standard errors of $\hat{N}$

are very close. When $N = 15,000$, larger differences between those summary statistics are observed.

For each of the four cases Fig. 1 describes a comparison between the estimated values of $s.d._{\text{Asy.}}(\hat{N})$ and $s.d._{\text{Boot.}}(\hat{N})$. We can observe that for large values of N and large values of capture probabilities (upper left pannel) there is a very good agreement between the asymptotic and bootstrap standard deviations for $\hat{N}$. Such agreement is still very good for large N and small values of capture probabilities (upper right pannel). However, when the population size is small, either for small (lower right pannel) or large (lower left pannel) capture probabilities, the bootstrap standard deviations tend to be larger than the asymptotic ones. Therefore, for small N the estimated asymptotic standard deviations of $\hat{N}$ tend to underestimate the true value.

For simulated populations with $N < 15,000$ individuals (not shown here) we observed even more extreme differences between the bootstrap and the asymptotic standard deviations, with last ones tending to underestimate the true uncertainty about $\hat{N}$. Therefore, expression (5) is not suitable for estimating uncertainty about $\hat{N}$ for small populations. These findings are though consistent with the asymptotic nature of the proposed methodology.

Acknowledgements The author is grateful to Antonio Eduardo Gomes and the anonymous referees for their insightful and helpful comments and suggestions which considerably improved the article.

References

Alho, J. M. (1990). Logistic regression in Capture–recapture models. *Biometrics*, *46*, 623–635.
Buckland, S. T. (1980). A modified analysis of the Jolly–Seber Capture–recapture model. *Biometrics*, *36*, 419–35.
Buckland, S. T., Garthwaite, P. H. (1991). Quantifying precision of mark–recapture estimates using the bootstrap and related methods. *Biometrics*, *47*, 255–68.
da Silva, C. Q., Zeh, J., Madigan, D., Laake, J., Rugh, D., Baraff, L., Koski, W., Miller, G. (2000). Capture–recapture estimation of bowhead whale population size using photo-identification data. *Journal of Cetacean Research and Management*, *2*, 45–61.
Efron, B. (1981). Nonparametric standard errors and confidence intervals. (With discussion). *Canadian Journal of Statistics*, *9*, 139–172.
Huggins, R. M. (1989). On the statistical analysis of capture experiments. *Biometrika*, *76*, 133–140.
Huggins, R. M. (1991). Some practical aspects of a conditional likelihood approach to capture experiments. *Biometrics*, *47*, 725–732.
Rivest, L. P., Potvin, F., Crépeau, H., Daigle, G. (1995). Statistical methods for aerial surveys using double-count technique to correct visibility bias. *Biometrics*, *51*, 461–470.
Sanathanan, L. (1972). Estimating the size of a multinomial population. *The Annals of Mathematical Statistics*, *43*, 142–152.
Schweder, T. (2003). Abundance estimation from multiple photo surveys: confidence distributions and reduced likelihoods for bowhead whales off Alaska. *Biometrics*, *59*, 974–983.
Sekar, C. C., Deming, W. E. (1949). On a method of estimating birth and death rates and the extent of registration. *Journal of the American Statistical Association*, *44*, 101–115.
Seber, G. A. F. (1982). *The estimation of animal abundance and related parameters*. 2nd edn. London: Charles Griffin and Company Ltd.