

Bootstrap Methods for Capture-Recapture Sampling

Metodi Bootstrap nel Campionamento per Cattura-Ricattura

Giancarlo Manzi, Fulvia Mecatti

Department of Statistics, University of Milan-Bicocca,
Via Bicocca degli Arcimboldi, 8, 20126, Milan, Italy
giancarlo.manzi@statistica.unimib.it, fulvia.mecatti@unimib.it

Riassunto: Lo scopo del presente lavoro è quello di estendere al caso del campionamento per cattura-ricattura il metodo bootstrap per la stima della varianza di stimatori costruiti su campioni da popolazioni finite. Nel campionamento da popolazioni di animali, non è raro il caso in cui alcuni animali, già catturati una volta, mostrino una accresciuta familiarità nei confronti del contatto umano, mentre altri tendano a nascondersi. In questi casi, le probabilità di inclusione possono risultare modificate. In questo lavoro si presentano due applicazioni dell'algoritmo bootstrap per il campionamento πPS proposto da Mecatti (2000) adattate al caso del campionamento per cattura-ricattura. La prima riguarda la stima della varianza dell'usuale stimatore di Petersen della numerosità della popolazione. La seconda utilizza la stessa stima come numerosità delle popolazioni empiriche bootstrap su cui si basa l'algoritmo di Mecatti. Il lavoro si conclude con due simulazioni su dati reali.

Keywords. Finite Population Sampling, Petersen Estimator, πPS Sampling, Variance Estimation.

1. Introduction

Animal counting for scientific purposes has always represented a challenging task for ecologists and biologists. Many restrictions have to be taken into consideration when dealing with populations of animals: geographical restrictions, animal identification restrictions, animal sighting or capture restrictions and so on. With regard to the sampling design, problems related with the inclusion probabilities structure have often led to biased results. The main animal behaviour that conducts to such unsatisfactory results is perhaps the *trap-shy/trap-happy* behaviour: after the first capture, the following captures are affected by the tendency of some animals to be familiar with the human contact. Other animals inclination is instead towards a trap-shy behaviour. Consequently, even if the inclusion probabilities could be assumed equal at the first capture, at the second selection the inclusion probability structure could result deeply modified. For example, badgers are considered to have a highly variable behaviour. In Minta and Mangel (1989), where an original Monte Carlo method is proposed, an example based on real data where badgers were snow tracked in a closed area is presented. Fifteen of the badgers were radio tagged and known to occupying the area. In a period of two months, badgers were snow tracked towards terminal holes in which they were identified if previously radio tagged. A total number of 107 badgers could be followed. Radio telemetry revealed that 68 of the tracks were generated by marked badgers and 39 by unmarked badgers. Out of the original 15 marked badgers, one was sighted 11 times, another one 9 times and, for example, two of them were never sighted.

It is clear that in the second sample some of the badgers are more likely to be selected than others. With this in mind, in Section 2, two applications of a bootstrap algorithm adapted for capture-recapture sampling are proposed: first the sighting probabilities are considered as inclusion probabilities; secondly the naïve use of the Mecatti's algorithm is considered by using the Petersen estimate of the population size in constructing the bootstrap empirical population. In Section 3, the two applications are evaluated by using real data from Minta and Mangel (1989) and some initial results are given.

2. Extensions of Mecatti's method to capture-recapture or capture-resight sampling

In Minta and Mangel (1989) is argued that sighting of animals is a sort of *visually recapturing* and it can be seen as an alternative approach to estimation. Practical differences between capture-recapture and capture-resight sampling are also discussed. Note that for estimation purposes they could be considered methodologically equivalent, *i.e.* the capture event is assumed equivalent to the sighting event. In our proposals, in drawing the first sample, the Lincoln-Petersen assumptions are considered. The first proposal aims at establishing an empirical probability distribution of sightings among animals in the second sample, considering both marked and unmarked animals, so that the bootstrap method by Mecatti (2000) for πPS sampling appears adequate. This empirical probability distribution could be thought as the sighting πPS structure of the animal population. In a sense, the more an animal is sighted the more it is expected to be sighted in the recapture sample: hence its inclusion probability results modified. Let M be the number of marked (or sighted) animals, namely the first sample (or the *capture-sample*) size, n be the total number of re-sightings (or recaptures), namely the second sample (or the *recapture-sample*) size, and m be the total number of marked animals in the second sample. Let $m_i, i = 1, \dots, M$, denotes the label of the i -th marked animal, nm_j be the label of the j -th unmarked animal, with $j = 1, \dots, \hat{U}$ where $\hat{U} = \hat{N} - M$ denotes the estimator of the number of unmarked animals in the population and \hat{N} denotes the customary Petersen estimator of the population size N . Moreover, let ms_i be the total number of sightings for animal i , *i.e.* its unique sighting in the first sample plus the sightings in the second sampling. The estimated number of sightings \hat{nms}_i for animal i is defined as $\hat{nms}_i = \frac{n - m}{\hat{N} - M}$. Note that this means that an equal probability distribution P of sightings among unmarked animals in the recapture sample is assumed. Thus, the estimated sampling probability distribution \hat{P} follows from the estimated sighting probability distribution of animals as given by

$$\hat{P}(m_i) = \frac{ms_i}{M + n} \quad \text{and} \quad \hat{P}(nm_j) = \frac{\hat{nms}_j}{M + n} \quad (1).$$

With the purpose of estimating the variance $V\left(\hat{N}\right)$ of the Petersen estimator, the

following bootstrap algorithm is proposed (*Algorithm 1*).

Step 1: Chose the number B of bootstrap iterations, sufficiently large. Obtain probabilities (1) for labels m_i and nm_j . *Step 2.* For b in (1: B) (b is the current bootstrap iteration):

Step 2.a: Draw a with replacement sample of size n from labels $\{m_1, \dots, m_M, nm_1, \dots, nm_{\hat{U}}\}$

with inclusion probabilities (1). *Step 2.b:* Calculate the number \hat{m}^b of marked labels in the bootstrap sample obtained in Step 2.a. *Step 2.c:* Calculate the bootstrap Petersen estimate of population size $\hat{N}^b = M \cdot \frac{n}{\hat{m}^b}$. *Step 2.d:* Add the quantity \hat{N}^b as a new item

in the vector S^* . *Step 3:* Calculate the variance of the values in S^* that is our bootstrap variance estimation.

In the second proposal (*Algorithm 2*) the sample value of the estimator \hat{N} is used as a device for applying the naïve Mecatti bootstrap algorithm (2000).

3. Two simulations based on a real case study

The two proposals above have been analysed empirically by performing two simulation studies. First we considered the badgers counting reported in Minta and Mangel (1989). The Petersen estimate results $\hat{N} = 24$, so that $\hat{U} = 9$, $\hat{nms}_i = 39/9 = 4.33$ and $\hat{P}(nm_i) = 0.035$. The variance of the Petersen estimator, calculated as the inverse of the Fisher information (Bailey, 1951) is 2.99. According to Algorithm 1, a sequence of 10 runs was performed with $B = 10,000$. The related variances and mean are reported in Table 1 (left). Monte Carlo variances calculated using 10,000 Minta-Mangel bootstrap estimates (Minta and Mangel 1989) are also presented. A second simulation was conducted on Algorithm 2 by considering the badgers skull length. For this simulation we have supposed the distribution of the $\hat{N} = 24$ badgers skull length measurements obtained as drawn from a Normal distribution with mean 12.404 and variance 0.107, according to previous similar studies (see, among others, Hidaka *et al.* 1998). A sequence of 10 runs was performed with $B = 10,000$ based on an original sample of size $M = 15$. For each simulation run, the variance of the sample mean was calculated with Algorithm 2 (without replacement), with the classical formula as explained, for example, in Cicchitelli *et al.* 1992, p. 225, and by using a naïve bootstrap with both the original sample and all the bootstrap samples drawn with replacement. Results are listed Table 1 (right).

Table 1. Results from two simulation studies for Algorithms 1 and 2

Simulation No.	Simulations for Algorithm 1		Simulations for Algorithm 2		
	Algorithm 1 Bootstrap Variance Estimate	Monte Carlo Minta-Mangel Variance Estimate	Algorithm 2 Bootstrap Variance Estimate for Sample Mean	Classical Variance Estimate for Sample Mean	Naive Bootstrap Variance Estimate for Sample Mean
1	2.22	0.063	0.0031	0.7838	1.6032
2	2.28	0.063	0.0035	1.0067	1.8394
3	2.34	0.058	0.0023	1.1567	1.9395
4	2.34	0.058	0.0014	1.3745	2.4701
5	2.26	0.058	0.0032	1.2595	3.9494
6	2.25	0.063	0.0027	1.3529	3.5699
7	2.21	0.063	0.0032	1.5338	3.2987
8	2.24	0.063	0.0036	1.2900	2.2293
9	2.24	0.058	0.0033	1.2406	4.0943
10	2.19	0.063	0.0031	0.7838	1.6032
Mean	2.26	0.061	0.0029	1.2221	2.7771

The estimates generated by the two proposed algorithms appear quite stable among all bootstrap simulation runs performed. The average of the Algorithm 1 bootstrap variance estimates (2.26) is very close to the Fisher information-based variance estimate (2.99). The Monte Carlo Minta-Mangel variance estimate could be considered suspect because it reaches only two different values among ten simulations (0.058 and 0.063), although it is very close to zero. Results from Algorithm 2 are less far from the classical variance estimate than the naïve bootstrap estimates, confirming the idea that when the independence structure is compromised, a modified bootstrap needs to be invoked in order to solve the unfitness problem related to classical bootstrap procedure. Future research will focus on confidence intervals coverage and on more complicated estimators.

References

- Bailey N.T.J. (1951) On estimating the size of mobile population from recapture data, *Biometrika*, 38, 293-306
- Cicchitelli G, Herzel A. and Montanari G.E. (1992) *Il Campionamento Statistico*, Il Mulino, Bologna
- Hidaka S, Matsumoto M, Hiji H, Ohsako S, Nishinakagawa H. (1998) Morphology and morphometry of skulls of raccoon dogs, *Nyctereutes procyonoides* and badgers, *Meles meles*, *J. Vet. Med. Sci.*, 60(2), 161-167
- Mecatti F. (2000) Bootstrapping Unequal Probability Samples, *Statistica Applicata*, Vol.12, n.1, 67-77
- Minta S. and Mangel M. (1989) A Simple Population Estimate Based on Simulation for Capture-Recapture and Capture-Resight Data, *Ecology*, 70(6), 1738-1751