A RESULT ABOUT THE MAXIMUM LIKELIHOOD ESTIMATE OF A POPULATION SIZE

Helder Paulo ZACHARIAS[1]
José Galvão LEITE[1]
Carlos Alberto Ribeiro DINIZ[1]

▪ ABSTRACT: In some cases it is possible that the maximum likelihood estimate of a population size be infinite. In this article a result showing such situation is presented.

▪ KEYWORDS: Estimation of population size; maximum likelihood estimate.

1 Introduction

The results of a capture-recapture design are usually used for estimating the size of a population. A capture-recapture design could initially consist of a randomly selected sample of an animal population, that is tagged and released (or part of them) to the population. In each of k ($k \geq 2$) of those type of sampling, a random number of animals is selected and those animals that were not captured before are now tagged and then are all (or part of them) released to the population.

Several authors have studied the problem of estimating the size of a population. Among them, Otis et al., 1978; Seber, 1982; Leite et al., 1988; Pollock, 1991. In this work it is considered a situation where the maximum likelihood estimate of a population size N can be infinite.

[1]Departamento de Estatística da Universidade Federal de São Carlos, Caixa Postal 676, CEP: 13565-905, São Carlos, SP, E-mail: dcad@power.ufscar.br

2 Statistical Model and the Likelihood Function

Let N be the population size; k the number of samples $(k \geq 2)$; p_i the probability that an animal can be captured in the ith sample, independently of the others, $i = 1, 2, \ldots, k$; $\mathbf{p} = (p_1, p_2, \ldots, p_k)$, the k-dimensional vector of those probabilities; n_i the number of captured animals in the ith sample, $i = 1, 2, \ldots, k$; m_i the number of tagged animals captured in the ith sample, $i = 1, 2, \ldots, k$; $(m_1 = 0)$; $M_i = \sum_{j=1}^{i-1} (n_j - m_j)$, $i = 2, 3, \ldots, k$, the number of tagged animal presented in the population exactly before of the selection of the ith sample, with $M_1 = 0$; $r = \sum_{i=1}^{k} (n_i - m_i)$ the total number of different animals captured during the process of the k sampling and $D = \{n_1, m_1; n_2, m_2; \ldots; n_k, m_k\}$ statistics or data referent to the experiment.

If the selected samples are independent then

$$P(n_1, m_1; n_2, m_2; \ldots; n_k, m_k | N, \mathbf{p}) = \prod_{i=2}^{k} \binom{M_i}{m_i} \tag{1}$$

$$\times \prod_{i=1}^{k} \binom{N - M_i}{n_i - mi} p_i^{n_i} (1 - p_i)^{N - n_i}. \tag{2}$$

Now, as

$$\prod_{i=1}^{k} \binom{N - M_i}{n_i - mi} = \left(\prod_{i=1}^{k} \frac{1}{(n_i - m_i)!} \right) \frac{N!}{(N - r)!},$$

the expression (2) becomes

$$P(n_1, m_1; n_2, m_2; \ldots; n_k, m_k | N, \mathbf{p}) =$$

$$\frac{N!}{(N - r)!} \prod_{i=2}^{k} \binom{M_i}{m_i} \prod_{i=1}^{k} \frac{1}{(n_i - m_i)!} p_i^{n_i} (1 - p_i)^{N - n_i}.$$

Thus, the likelihood function, for $N \geq r, 0 < p_i < 1$, can be written as

$$\begin{aligned} L(N, \mathbf{p} | D) &= P(n_1, m_1; n_2, m_2; \ldots; n_k, m_k | N, \mathbf{p}) \\ &\propto \binom{N}{r} \prod_{i=1}^{k} p_i^{n_i} (1 - p_i)^{N - n_i}. \end{aligned} \tag{3}$$

The likelihood function resumes all relevant information involved with the data about the model parameters. In the next section it is presented the maximum likelihood estimates (MLE) of N and $\mathbf{p}$.

3 Maximum Likelihood Estimates

The maximum likelihood estimates of N and $\mathbf{p}$, denoted by $\hat{N}$ and $\hat{\mathbf{p}}$, respectively, are given by the following theorem.

Theorem 1. *The maximum likelihood estimate of the vector parameters* $(N, \mathbf{p})$, *is given by* $(\widehat{N}, \widehat{\mathbf{p}})$, *where* $\widehat{N}$ *is approximately equal to the solution of the equation* $1 - \frac{r}{N} = \prod_{i=1}^{k} \left(1 - \frac{n_i}{N}\right)$, $N > r$, *and* $\hat{\mathbf{p}} = (\hat{p}_1, \hat{p}_2, \ldots, \hat{p}_k)$ *with* $\hat{p}_i = \frac{n_i}{\hat{N}}$, $i = 1, 2, \ldots, k$;

Proof. Since $(\widehat{N}, \widehat{\mathbf{p}})$ is the point of maximum of $\ln K(N, \mathbf{p})$, where $K(N, \mathbf{p})$ is the kernel of $L(N, \mathbf{p}|D)$, that is, $K(N, \mathbf{p})$ is the factor of $L(N, \mathbf{p}|D)$ that depends on N and $\mathbf{p}$, then from the relation (3) we have

$$\ln K(N, \mathbf{p}) = \ln \binom{N}{r} + \sum_{i=1}^{k} \left[n_i \ln p_i + (N - n_i) \ln (1 - p_i)\right],$$

where $N \geq r$ and $0 < p_i < 1$, $i = 1, 2, \ldots, k$. The point $(\widehat{N}, \widehat{\mathbf{p}})$ satisfy the $k + 1$ equations given by

$$\begin{aligned} \left.\frac{\partial \ln K(N, \mathbf{p})}{\partial p_i}\right|_{(N,\mathbf{p})=(\widehat{N},\widehat{\mathbf{p}})} &= 0, \; i = 1, 2, \ldots, k; \\ \ln K(\widehat{N}, \widehat{\mathbf{p}}) - \ln K(\widehat{N} - 1, \widehat{\mathbf{p}}) &= 0. \end{aligned}$$

The last equation above, which is equivalent to

$$K(\widehat{N}, \widehat{\mathbf{p}}) = K(\widehat{N} - 1, \widehat{\mathbf{p}}),$$

corresponds formally to

$$\left.\frac{\partial \ln K(N, \mathbf{p})}{\partial N}\right|_{(N,\mathbf{p})=(\widehat{N},\widehat{\mathbf{p}})} = 0,$$

since N is a integer variable assuming values greater or equal to r (Darroch, 1958).

Thus,

$$\frac{\partial \ln K(N,\mathbf{p})}{\partial p_i} = \frac{n_i}{p_i} - \frac{N-n_i}{1-p_i} = 0 \Rightarrow \frac{n_i}{p_i} = \frac{N-n_i}{1-p_i} \Rightarrow \hat{p}_i = \frac{n_i}{\hat{N}}, 1 \le i \le k,$$

where $\hat{N}$ is approximately equal to the solution of the equation

$$K(N,\mathbf{p}) = K(N-1,\mathbf{p}), N \ge r+1,$$

that is,

$$\begin{aligned}
\binom{N}{r}\prod_{i=1}^{k} p_i^{n_i}(1-p_i)^{N-n_i} &= \binom{N-1}{r}\prod_{i=1}^{k} p_i^{n_i}(1-p_i)^{N-n_i-1}, N \ge r+1 \\
&\Rightarrow\ 1-\frac{r}{N} = \prod_{i=1}^{k}(1-p_i), N \ge r+1 \\
&\Rightarrow\ 1-\frac{r}{N} = \prod_{i=1}^{k}\left(1-\frac{n_i}{N}\right), N \ge r+1,
\end{aligned}$$

which finalizes the prove.

Observe that $\hat{N} = \infty$ when $r = \sum_{i=1}^{k} n_i$. In order to demonstrate that $\hat{N} = \infty$ is the unique solution of $1 - \frac{\sum_{i=1}^{k} n_i}{N} = \prod_{i=1}^{k}\left(1-\frac{n_i}{N}\right), N \ge \sum_{i=1}^{k} n_i + 1$, it is enough to show that for any real number $x_1, x_2, \ldots, x_k$ such that $0 < x_i < 1$, $1 \le i \le k$,

$$1 - \sum_{i=1}^{k} x_i < \prod_{i=1}^{k}(1-x_i), \tag{4}$$

for all $k \ge 2$.

The relation (4) is proved by finite induction about $k \ge 2$.

For $k = 2$ and for any $x_1, x_2 \in (0,1)$ we have $(1-x_1)(1-x_2) = 1-(x_1+x_2)+x_1x_2 > 1-(x_1+x_2)$, which proves relation (4). Suppose now that (4) is true for any natural k, $k \ge 2$. Therefore, for any $x_1, x_2, \ldots, x_k, x_{k+1}$ such that $0 < x_i < 1$, $i = 1, 2, \ldots, k+1$, we have

$$\begin{aligned}
\prod_{i=1}^{k+1}(1-x_i) &= \left(\prod_{i=1}^{k}(1-x_i)\right)(1-x_{k+1}) > \left(1-\sum_{i=1}^{k} x_i\right)(1-x_{k+1}) \\
&= 1-\sum_{i=1}^{k+1} x_i + \left(\sum_{i=1}^{k} x_i\right) x_{k+1} > 1-\sum_{i=1}^{k+1} x_i,
\end{aligned}$$

which shows that the relation (4) is valid for the natural number $k+1$, proving (4).

The Theorem 1 is illustrated by presenting, in Table 1, the approximate MLEs of N for some values of k, $n_1, n_2, \ldots, n_k$ and r. Observe that the values of the MLEs increase considerably as the values of r approximate the sum of the $n_i^{'}s$.

Table 1- Approximate maximum likelihood estimates of N

k	n_j	r	MLE
	$n_1 = 40$	80	120
2	$n_2 = 60$	90	240
		98	1200
		99	2400
	$n_1 = 1$	8	8
3	$n_2 = 5$	12	26
	$n_3 = 8$	13	52
	$n_1 = 76$	390	410
	$n_2 = 176$	440	483
5	$n_3 = 220$	580	798
	$n_4 = 245$	700	1402
	$n_5 = 182$	850	6256
		890	34824
	$n_1 = 18$	90	91
	$n_2 = 29$	130	140
	$n_3 = 51$	170	214
7	$n_4 = 55$	210	348
	$n_5 = 49$	260	893
	$n_6 = 36$	290	4108
	$n_7 = 61$	295	9330
		298	37527

4 Conclusions

In this work it is showed that the values of the MLEs increase as the values of r approximate of the sum of the $n_i^{'}s$. Also it is showed that the MLE of the population size, N, is infinite when $r = \sum n_i$. For those cases where there is the possibility of infinite maximum likelihood estimate of the population size is convenient to use a Bayesian approach in order to estimate of N. Results in this direction can be seen in Zacharias *et al.* (2000a), (2000b), (2001), Leite *et al.* (2000), George & Robert (1992).

Acknowledgments

The authors wish to thank the referees for helpful comments and suggestions. The research of José Galvão Leite is partially supported by CNPq/Brazil, grant 301339/91-0.

ZACHARIAS, H. P.; LEITE, J. G.; DINIZ, C.A.R. Um resultado sobre o estimador de máxima verossimilhança do tamanho de uma população.*Rev. Mat. Estat.* (São Paulo), v. 20, p. 155-161, 2002.

- *RESUMO: Em alguns casos é possível que o estimador de máxima verossimilhança do tamanho de uma população seja infinito. Neste artigo é apresentado um resultado mostrando tal situação.*

- *PALAVRAS-CHAVE: Estimação do tamanho de uma população, estimador de máxima verossimilhança.*

References

DARROCH, J. N. The multiple-recapture census, I: estimation of a closed population. *Biometrika,* v.45, p.343-59, 1958.

GEORGE, E. I.; ROBERT, C. P. Capture-recapture estimation via Gibbs sampling. *Biometrika,* v.79, n.4, p.677-83, 1992.

LEITE, J. G.; OISHI, J.; PEREIRA, C. A. B. A note on the exact maximum likelihood estimation of the size of a finite and closed population. *Biometrika*, v.75, p.178-80, 1988.

LEITE, J. G.; RODRIGUES, J.; MILAN, L. A. A Bayesian analysis for estimating the number of species in a population using nonhomogeneos Poisson process. *Statistics and Probability Letters,* v.48, p.153-161, 2000.

OTIS, D. L.; BURNHAM, K. P.; WHITE, G. C.; ANDERSON, D. R. Statistical inference for capture data on closed animal population. *Wildl. Monogr.* v.62, p.1-135, 1978.

POLLOCK, K. H. Modeling capture, recapture, and removal statistics for estimation of demographic parameters for fish and wildlife populations: Past, present, and future. *J. Am. Statist. Assoc.*, v.86, p.225-238, 1991.

SEBER, G. A. F. *The estimation of animal abundance and related parameters*. London: Charles Griffin, 1982. 398p.

ZACHARIAS, H. P.; LEITE, J. G.; DINIZ, C. A. R. Estimating population size for capture-recapture model via Gibbs sampling. Submitted, 2001.

ZACHARIAS, H. P.; LEITE, J. G.; DINIZ, C. A. R. A bayesian approach on the estimation of the size of a population.. Technical Report, n.58, ISSN 0104-0499, DEs-UFSCar, 2000a. 24p.

ZACHARIAS, H. P.; LEITE, J. G.; DINIZ, C. A. R. The use of noninformative priors on the Bayesian estimation of the size of a population. Technical Report, n.57, ISSN 0104-0499, DEs-UFSCar, 2000b. 24p.

Recebido em 12.10.2002.