RT-MAE-9105

ON THE ESTIMATION OF THE SIZE OF A FINITE AND CLOSED POPULATION

by

Heleno Bolfarine, José G. Leite

and

Josemar Rodrigues

Palavras Chaves: Likelihood function; Profile likelihood; Posterior mode; Laplace approximation.

Classificação AMS: 60E05
(AMS Classification)

ON THE ESTIMATION OF THE SIZE OF A FINITE AND CLOSED POPULATION

Heleno Bolfarine, José G. Leite
Josemar Rodrigues
Universidade de São Paulo, Departamento de Estatistica-IME
Caixa Postal No. 20570- Agencia Iguatemi
São Paulo, S.P.- BRASIL

Abstract

In this paper, we consider the problem of estimating the size N of a finite and closed population, using data obtained from capture-recapture experiments. By defining an appropriate model, we investigate the maximum of the likelihood, of the profile likelihood and of an orthogonal adjusted profile likelihood (Cox and Reid, 1987) function. We show that they all may present infinity as the maximum likelihood estimator of N. This seems to be a caracteristic of the likelihood approach in this problem. Further, we present a Bayesian approach with minimum prior information as a way of countering this difficulty. Exact analytical expressions for the posterior modes are also obtained.

1. Introduction

Consider a finite and closed animal population of size N, which is unknown. By using data obtained from a capture-recapture sequential experiment, we investigate likelihood and Bayesian approaches for estimating the unknown population size N. Inferences about N based on data obtained by special cases of this sampling process were considered by many authors. See for example the reference list by Seber (1982). More recently, Leite et al. (1988) derived an exact analytical expression for the maximum likelihood estimator, conditional on the observed sample sizes. The animal population is considered to be closed, that is, during the time of the study, we assume that it does not change in number nor in form. From this population, random samples are sequentially selected according to the the capture-recapture sampling scheme. Before each of the samples are returned to the population, the animals not seen before (unmarked) are marked, returned to the population and their number is recorded. As in Castledine (1981), we assume that the capture probabilities are the same for all animals but, it may change during the realization of the study.

In Section 2 we introduce the model and derive the likelihood function corresponding to the observed data. We show that the maximum likelihood estimator (MLE) of N depend strongly on the capture probabilities and it may even be infinite. In Section 3 we consider the profile likelihood of N witch follows from the likelihood function derived in Section 2. By studying its behaviour, we provide some situations where its maximum is infinite. Section 4 introduces an orthogonal (local) transformation in such a way that the parameter of interest N is locally orthogonal to the other (nuisance) parameters in the model. With this transformation, estimation of N is less afected by the presence of the nuisance parameters. Following the lines of Cox and Reid (1987), we propose an adjusted orthogonal profile likelihood function for the parameter N and show that the

MLE of N that follows from this likelihood may also be infinite. Thus, both, the profile likelihood and the orthogonal adjusted profile likelihood fuctions present the inconvenience of having maximum likelihood estimators with no finite moments. In Section 5 we consider a Bayesian approach with minimum prior information on the unknown parameters. We show that the mode of the posterior distribution is always finite and derive exact explicit analytical expressions for it. In Section 6, the mode of a profile posterior distribution obtained by using the Laplace method for integrals (see Tierney and Kadane, 1986) is also investigated.

2. The Model and the Likelihood Function

As mentioned above, we write N for the unknown population size, s (≥ 2) for the number of samples taken, X_i for the number of unmarked animals in the i-th sample, Y_i for the number of marked animals in the i-th sample ($Y_1 = 0$) and M_i for the number of marked animals just before the i-th sample, $i = 1, \ldots, s$. We note that $M_1 = 0$ and that $M_{i+1} = M_i + X_i = \sum_{j=1}^{i} X_j$, $i = 1, \ldots, s$. Let $n_i = X_i + Y_i$ be the number of animals captured in the i-th sample, $i = 1, \ldots, s$. Furthermore, we assume that the animals are captured independently, with probability p_i in the i-th sample, $i = 1, \ldots, s$. As emphasized before, it is assumed that the population remains closed throughout the realization of the experiment.

Under the above assumptions, it follows that

$$X_i | p_i \sim B(N - M_i, p_i; x_i)$$

and

$$Y_i | p_i \sim B(M_i, p_i; y_i),$$

$i = 1, \ldots, s$, where $B(n, p; x)$ denotes the probability of x successes in n trials of a binomial experiment with success probability p. To complete the notation, let $\mathbf{p} = (p_1, \ldots, p_s)'$ and $\mathcal{D} = \{(x_i, y_i); i = 1, \ldots, s, y_1 = 0\}$, the observed data. From the independence of X_i and Y_i, it follows that the likelihood function corresponding to the data $\mathcal{D}$ is given by

$$(1) \qquad L(N, \mathbf{p}) \propto \frac{N!}{(N - M_{s+1})!} \prod_{i=1}^{s} p_i^{n_i} (1 - p_i)^{N - n_i}, \quad N \geq M_{s+1}$$

with the restriction that

$$max\{n_i; 1 \leq i \leq s\} \leq M_{s+1} \leq \sum_{i=1}^{s} n_i.$$

It is not difficult to see that the maximum of the likelihood function (1) occurs at the solution of the following equations:

$$(2) \qquad \hat{p}_i = \frac{n_i}{\hat{N}},$$

and

(3) $$\hat{N} = \frac{M_{s+1}}{1 - \prod_{i=1}^{s}(1 - \hat{p}_i)},$$

$i = 1, \ldots, s$. Note from equations (2) and (3) that the maximum likelihood estimator of N depends strongly on the maximum likelihood estimators of p_i, $i = 1, \ldots, s$ and that the maximum likelihood estimator of N is the solution to the equation

(4) $$\prod_{i=1}^{s}(1 - \frac{n_i}{N}) = 1 - \frac{M_{s+1}}{N}.$$

In the next lemma, we present a solution to the equation (4) in a special situation.

Lemma 2.1. *If $M_{s+1} = \sum_{i=1}^{s} n_i$, then $\hat{N} = \infty$.*

Proof. For $s \geq 2$ and $0 < x_i < 1$, $i = 1, \ldots, s$, we have

$$\prod_{i=1}^{s}(1 - x_i) > 1 - \sum_{i=1}^{s} x_i.$$

Since $0 < n_i/N < 1$, $i = 1, \ldots, s$, and $M_{s+1} = \sum_{i=1}^{s} n_i$, it follows that

$$\prod_{i=1}^{s}(1 - \frac{n_i}{N}) > 1 - \sum_{i=1}^{s} \frac{n_i}{N} = 1 - \frac{M_{s+1}}{N}.$$

Thus, $\hat{N} = \infty$ is the only possible solution to (4).

Note that Lemma 2.1 provides situation where the maximum likelihood estimator (solution to equation (4)) is infinite.

3. The Profile Likelihood Function

It follows from (1) and (2) that the profile likelihood of N, corresponding to the observed data $\mathcal{D}$ is

(5) $$L_p(N|\mathcal{D}) \propto \frac{N!}{(N - M_{s+1})!} \prod_{i=1}^{s} (\frac{n_i}{N})^{n_i}(1 - \frac{n_i}{N})^{N - n_i}; \; N \geq M_{s+1}.$$

We denote by $h(N)$ the kernel of the likelihood (5), that is, let

(6) $$h(N) = \frac{N!}{(N - M_{s+1})! N^{\sum_{i=1}^{s} n_i}} \prod_{i=1}^{s}(1 - \frac{n_i}{N})^{N - n_i}; \; N \geq M_{s+1}.$$

The next two Lemmas illustrate the behaviour of $L_p(N|\mathcal{D})$ in two special cases. In the sequel, $\hat{N}$ denotes the maximum of $L_p(N|\mathcal{D})$ given in (5).

Lemma 3.1. *If $M_{s+1} < \sum_{i=1}^{s} n_i$ then $\hat{N} < \infty$.*

Proof. From (6), we may write

$$h(N) = \frac{N(N-1)\ldots(N-(M_{s+1}-1))}{N^{M_{s+1}}N^{\sum_{i=1}^{s} n_i - M_{s+1}}}\prod_{i=1}^{s}(1-\frac{n_i}{N})^{N-n_i}$$

$$= (1-\frac{1}{N})(1-\frac{2}{N})\ldots(1-\frac{(M_{s+1}-1)}{N})\frac{1}{N^{\sum_{i=1}^{s} n_i - M_{s+1}}}\prod_{i=1}^{s}(1-\frac{n_i}{N})^{N-n_i}$$

(7) $$\to 1.0.\prod_{i=1}^{s} e^{-n_i} = 0,$$

as $N \to \infty$. Thus, (7) implies that $\hat{N} < \infty$.

Lemma 3.2. *If $M_{s+1} = \sum_{i=1}^{s} n_i$ and*

(8) $$(\sum_{i=1}^{s} n_i)^2 - \sum_{i=1}^{s} n_i - 2\sum_{i=1}^{s} n_i^2 \geq 0,$$

then $\hat{N} = \infty$.

Proof. Since $M_{s+1} = \sum_{i=1}^{s} n_i$, it follows from (6) that

$$h(N) = \frac{N(N-1)\ldots(N-(M_{s+1}-1))}{N^{M_{s+1}}}\prod_{i=1}^{s}(1-\frac{n_i}{N})^{N-n_i}$$

(9) $$= (1-\frac{1}{N})\ldots(1-\frac{(M_{s+1}-1)}{N})\prod_{i=1}^{s}(1-\frac{n_i}{N})^{N-n_i}$$

$$< e^{-\frac{(1+2+\ldots+(M_{s+1}-1))}{N}}\prod_{i=1}^{s}(e^{-n_i/N})^{N-n_i}$$

$$= e^{-\frac{M_{s+1}(M_{s+1}-1)}{2N} - M_{s+1} + \sum_{i=1}^{s} n_i^2/N}$$

$$= e^{-((\sum_{i=1}^{s} n_i)^2 - \sum_{i=1}^{s} n_i - 2\sum_{i=1}^{s} n_i^2)/2N - \sum_{i=1}^{s} n_i}.$$

Thus, if

$$(\sum_{i=1}^{s} n_i)^2 - \sum_{i=1}^{s} n_i - 2\sum_{i=1}^{s} n_i^2 \geq 0,$$

then

(10) $$h(N) < e^{-\sum_{i=1}^{s} n_i}; \quad N \geq M_{s+1}.$$

On the other hand, it follows from (9) that

$$h(N) \to 1.\prod_{i=1}^{s} e^{-n_i} = e^{-\sum_{i=1}^{s} n_i}, \tag{11}$$

as $N \to \infty$. The result follows then from (10) and (11).

Hence, Lemma 3.2 provides conditions under which the maximum of the profile likelihood (5), $\hat{N}$, is infinite. In the next examples, we consider some situations where the conditions in Lemma 3.2 are satisfied.

Example 3.1. Suppose that $s \geq 3$, $n_1 = \ldots = n_s = 1$ and $M_{s+1} = \sum_{i=1}^{s} n_i = s$. Since

$$s^2 - s - 2s = s^2 - 3s = s(s-3) \geq 0,$$

condition (8) is satisfied and, from Lemma 3.2, it follows that $\hat{N} = \infty$.

Example 3.2. Suppose that $s = 3$, $n_1 = 40$, $n_2 = 60$, $n_3 = 30$ and that $M_4 = \sum_{i=1}^{3} n_i = 130$. Then, condition (8) is satisfied, since

$$(\sum_{i=1}^{3} n_i)^2 - \sum_{i=1}^{3} n_i - 2\sum_{i=1}^{3} n_i^2 = 4570.$$

Hence, from Lemma 3.2, it follows that $\hat{N} = \infty$.

4. An Adjusted Orthogonal Profile Likelihood

As seen in the previous sections, both, the likelihood and the profile likelihood may have $\hat{N} = \infty$ as their maximum. In this section, we introduce a local orthogonal transformation of (N, p) and, based on this transformation, we consider an adjusted orthogonal likelihood function. For this adjusted likelihood, we show that its maximum may also be $\hat{N} = \infty$. Inspired on Cox and Reid (1987), the next definition considers conditions under which N and a parameter vector $\mu = (\mu_1, \ldots, \mu_n)$ are locally orthogonal.

Definition 4.1. *The parameters N and μ are (locally) orthogonal if*

$$\frac{\partial^2 lnL(N,\mu)}{\partial N \partial \mu_i}|_{(N=\hat{N};\mu=\hat{\mu})} = 0, \tag{12}$$

where $(\hat{N}, \hat{\mu})$ is the maximum likelihood estimator of (N, μ).

Note that the condition required by Cox and Reid (1987) is related to the Fisher information matrix of N and μ. On the other hand, (12) is related to the observed information of N and μ, that is, the second derivatives with respect to N and μ_i evaluated at the maximum likelihood estimators. Now, let

$$\mu_i = Np_i, \tag{13}$$

$i = 1, \ldots, s$.

Lemma 4.1. *N is locally orthogonal to μ.*

Proof. It follows from (1) and (13) that the likelihood function of N and μ is given by

$$L(N, \mu|\mathcal{D}) \propto \frac{N!}{(N - M_{s+1})!} \prod_{i=1}^{s} \left(\frac{\mu_i}{N}\right)^{n_i} \left(1 - \frac{\mu_i}{N}\right)^{N-n_i}; \tag{14}$$

$N \geq M_{s+1}$. It is easy to check that

$$\hat{\mu}_i = n_i,$$

$i = 1, \ldots, s$. Thus,

$$\frac{\partial^2 \ln L(N, \mu|\mathcal{D})}{\partial N \partial \mu_i}\Big|_{(\mu=\hat{\mu}, N=\hat{N})} = \frac{n_i - n_i}{(\hat{N} - n_i)^2} = 0,$$

$i = 1, \ldots, s$, from where the result follows.

Now, following the lines of Cox and Reid (1987), we propose the following adjusted orthogonal likelihood for N

$$L_a(N|\mathcal{D}) \propto L_p(N|\mathcal{D}) \left\{ det\left[-\frac{\partial^2 \ln L(N, \mu|\mathcal{D})}{\partial \mu_i \partial \mu_j}\Big|_{\mu=\hat{\mu}} \right] \right\}^{-1/2}. \tag{15}$$

Notice that the likelihood (15) may also be justified in terms of the Laplace approximation to the marginal posterior distribution of N, as considered for example in Sweeting (1987).

After some algebraic manipulations, it can be shown that

$$\left\{ det\left[-\frac{\partial^2 \ln L(N, \mu|\mathcal{D})}{\partial \mu_i \partial \mu_j}\Big|_{\mu=\hat{\mu}} \right] \right\}^{-1/2} = N^{s/2} \prod_{i=1}^{s} \left(\frac{n_i}{N}\right)^{1/2} \left(1 - \frac{n_i}{N}\right)^{1/2}.$$

Thus, we may write

$$L_a(N|\mathcal{D}) \propto \frac{N! N^{s/2}}{(N - M_{s+1})!} \prod_{i=1}^{s} \left(\frac{n_i}{N}\right)^{n_i+1/2} \left(1 - \frac{n_i}{N}\right)^{N-n_i+1/2}, \tag{16}$$

$N \geq M_{s+1}$.

Lemma 4.2. *Let $\hat{N}$ be the value of N that maximizes $L_a(N|\mathcal{D})$ given in (16).*

(i) If $M_{s+1} < \sum_{i=1}^{s} n_i$, then $\hat{N} < \infty$;

(ii) If $M_{s+1} = \sum_{i=1}^{s} n_i$ and $(\sum_{i=1}^{s} n_i)^2 - 2\sum_{i=1}^{s} n_i^2 \geq 0$, then $\hat{N} = \infty$.

Proof. Notice that we may write

$$L_a(N|\mathcal{D}) \propto \frac{N!}{(N - M_{s+1})! N^{\sum_{i=1}^{s} n_i}} \prod_{i=1}^{s} \left(1 - \frac{n_i}{N}\right)^{N-n_i+1/2}.$$

If $M_{s+1} < \sum_{i=1}^{s} n_i$ then

$$\frac{N!}{(N-M_{s+1})!N^{\sum_{i=1}^{s} n_i}} \prod_{i=1}^{s}(1-\frac{n_i}{N})^{N-n_i+1/2}$$

$$= (1-\frac{1}{N})(1-\frac{2}{N})\dots(1-\frac{(M_{s+1}-1)}{N})\frac{1}{N^{\sum_{i=1}^{s} n_i - M_{s+1}}} \prod_{i=1}^{s}(1-\frac{n_i}{N})^{N-n_i+1/2}$$

$$\to 1.0.e^{-\sum_{i=1}^{s} n_i} = 0,$$

as $N \to \infty$, which proves (i).

On the other hand, if $M_{s+1} = \sum_{i=1}^{s} n_i$ and $(\sum_{i=1}^{s} n_i)^2 - 2\sum_{i=1}^{s} n_i^2 \geq 0$ then

$$\frac{N!}{(N-M_{s+1})!N^{\sum_{i=1}^{s} n_i}} \prod_{i=1}^{s}(1-\frac{n_i}{N})^{N-n_i+1/2}$$

$$= (1-\frac{1}{N})\dots(1-(\frac{M_{s+1}-1}{N}))\prod_{i=1}^{s}(1-\frac{n_i}{N})^{N-n_i+1/2}$$

$$< e^{-\frac{(1+\dots+(M_{s+1}-1))}{N}} \prod_{i=1}^{s}(e^{-\frac{n_i}{N}})^{N-n_i+1/2}$$

$$= e^{-\frac{M_{s+1}(M_{s+1}-1)}{2N}} \prod_{i=1}^{s} e^{-n_i+\frac{n_i^2}{N}-\frac{n_i}{2N}}$$

$$= e^{\frac{-(\sum_{i=1}^{s} n_i)^2+2\sum_{i=1}^{s} n_i^2}{2N}} e^{-\sum_{i=1}^{s} n_i}$$

$$\leq e^{-\sum_{i=1}^{s} n_i}.$$

Since

$$\frac{N!}{(N-M_{s+1})!N^{\sum_{i=1}^{s} n_i}} \prod_{i=1}^{s}(1-\frac{n_i}{N})^{N-n_i+1/2} \to e^{-\sum_{i=1}^{s} n_i},$$

(ii) follows.

5. Bayesian Estimation of N

As shown in the previous sections, the maximum likelihood estimator of N which follows from the likelihood, profile likelihood and adjusted orthogonal likelihood may be infinite. In this section, we present a Bayes estimator of N with minimal (noninformative) prior information which is always finite. Further, we exibt an explicit analytical expression for this estimator.

As in Section 4, we transform the parameter $(N, \mathbf{p})$ to the parameter (N, μ), where $\mu_i = Np_i$, as suggested by Lemma 4.1. Thus, in this new parametrization, it follows that the likelihood function of N and μ is given by

$$L(N,\mu|\mathcal{D}) \propto \frac{N!}{(N-M_{s+1})!} \prod_{i=1}^{s}(\frac{\mu_i}{N})^{n_i}(1-\frac{\mu_i}{N})^{N-n_i}. \tag{17}$$

Moreover, since $0 < \mu_i < N$, for all i, we consider the prior

$$\pi(N,\mu) = \pi(\mu|N)\pi(N) = \prod_{i=1}^{s} \pi(\mu_i|N)\pi(N) = \prod_{i=1}^{s} \frac{1}{N}\frac{1}{(N+1)^2}$$

$$(18) \qquad = \frac{1}{N^s(N+1)^2},$$

where we consider the proper noninformative prior

$$\pi(N) = \frac{1}{(N+1)^2}.$$

It follows then from (17) and (18) that the joint posterior of N and μ is given by

$$(19) \qquad \pi(N,\mu|\mathcal{D}) \propto \frac{N!}{(N-M_{s+1})!} \prod_{i=1}^{s} (\frac{\mu_i}{N})^{n_i}(1-\frac{\mu_i}{N})^{N-n_i} \frac{1}{N^s(N+1)^2}; \; N \geq M_{s+1}.$$

Integrating out μ from the joint distribution (19), it follows that the marginal posterior distribution of N is given by

$$\pi(N|\mathcal{D}) \propto \frac{N!}{(N-M_{s+1})!N^s(N+1)^2} \prod_{i=1}^{n} \int_0^N (\frac{\mu_i}{N})^{n_i}(1-\frac{\mu_i}{N})^{N-n_i} d\mu_i$$

$$(20) \qquad = \frac{N!}{(N-M_{s+1})!(N+1)^2 \prod_{i=1}^{s} \binom{N+1}{n_i+1}}; \; N \geq M_{s+1}.$$

Lets denote by $\hat{N}$ the mode of the posterior density (20).

Lemma 5.1. *If $M_{s+1} = max\{n_1, \ldots, n_s\}$ then, $\hat{N} = M_{s+1}$ is the unique mode of $\pi(N|\mathcal{D})$.*

Proof. Notice that we may write

$$\pi(N|\mathcal{D}) \propto \frac{\binom{N+1}{M_{s+1}+1}}{(N+1)^3 \prod_{i=1}^{s} \binom{N+1}{n_i+1}}; \quad N \geq M_{s+1}.$$

Without loss of generality we may assume that $M_{s+1} = n_s$. Thus,

$$\pi(N|\mathcal{D}) \propto \frac{1}{(N+1)^3 \prod_{i=1}^{s-1} \binom{N+1}{n_i+1}}; \quad N \geq n_s,$$

from where the result follows.

We consider now the case where $M_{s+1} > max\{n_1, \dots, n_s\}$. Let

$$K(N) = \frac{N!}{(N - M_{s+1})!(N+1)^2 \prod_{i=1}^{s} \binom{N+1}{n_i+1}}; \quad N \geq M_{s+1}$$

and define the function

$$g(x) = (1 - M_{s+1}x)^{-1}(1+x)^{-(s+2)} \prod_{i=1}^{s}(1 - n_i x); \quad x \in [0, \frac{1}{M_{s+1}}).$$

It is clear that the posterior mode $\hat{N}$ of (20) maximizes $K(N)$. Moreover, after some algebraic manipulations, it may be noticed that

$$g(\frac{1}{N+1}) = \frac{K(N+1)}{K(N)}; \quad N \geq M_{s+1}.$$

The next result discuss the behavior of the function g in the interval $(0, \frac{1}{M_{s+1}})$.

Lemma 5.2. *If $M_{s+1} > max\{n_1, \dots, n_s\}$, then the equation $g(x) = 1$ has only one nonnull root x_o in the interval $(0, 1/M_{s+1})$. Further, if $x \in (0, x_o)$ then $g(x) < 1$. On the other hand, if $x \in (x_o, 1/M_{s+1})$ then $g(x) > 1$.*

Proof. Consider the functions

$$g_1(x) = \prod_{i=1}^{s}(1 - n_i x) \text{ and } g_2(x) = (1 - M_{s+1}x)(1+x)^{s+2},$$

for all real x. It can be verified that the first and second derivatives of $g_1(.)$ are such that

$$g_1'(x) = \sum_{i=1}^{s}(-n_i) \prod_{j \neq i}^{s}(1 - n_j x) < 0$$

and

$$g_1''(x) = \sum_{i=1}^{s} \sum_{j \neq i}^{s} n_i n_j \prod_{k \neq i,j}^{s} (1 - n_k x) > 0,$$

for all x in the interval $(0, 1/M_{s+1})$. Hence, g_1 is a continuous, convex and decreasing function in the interval $[0, 1/M_{s+1}]$. Furthermore, the function g_2 is continuous and its first and second derivatives are given by

$$g_2'(x) = [-M_{s+1}(s+3)x + s + 2 - M_{s+1}](1+x)^{s+1}$$

and

$$g_2''(x) = [-M_{s+1}(s+2)(s+3)x + (s+1)(s+2 - M_{s+1}) - M_{s+1}(s+3)](1+x)^{s},$$

respectively, for all real x. Hence, if $M_{s+1} \geq s+2$, then $g_2'(x) < 0$ and $g_2''(x) < 0$ for all $x \in (0, 1/M_{s+1})$, that is, g_2 is a concave and decreasing function in the interval $[0, 1/M_{s+1}]$. At the origin $g_1(0) = g_2(0) = 1$ and at the point $1/M_{s+1}$, $g_1(1/M_{s+1}) > 0$ and $g_2(1/M_{s+1}) = 0$. Further,

$$g_2'(0) - g_1'(0) = s + 2 - M_{s+1} + \sum_{i=1}^{s} n_i > 0.$$

Thus, there exists a positive real number δ such that $g_2'(x) - g_1'(x) > 0$, for all $x \in [0, \delta)$. From the Mean Value Theorem, it follows that $g_2(x) > g_1(x)$, for all $x \in (0, \delta)$. Consequently, there is a unique point $x_o \in (0, 1/M_{s+1})$ such that

$$\begin{aligned} &g_1(x_o) = g_2(x_o), \\ &g_1(x) < g_2(x), \quad \text{for all } x \in (0, x_o) \text{ and} \\ &g_1(x) > g_2(x), \quad \text{for all } x \in (x_o, 1/M_{s+1}). \end{aligned}$$

The result then follows from the fact that g is the restriction of g_1/g_2 in the interval $[0, 1/M_{s+1})$.

We consider now the case where $M_{s+1} < s + 2$. In this situation,

$g_2'(x) > 0$ for all $x \in (0, \frac{s+2-M_{s+1}}{M_{s+1}(s+3)})$

and

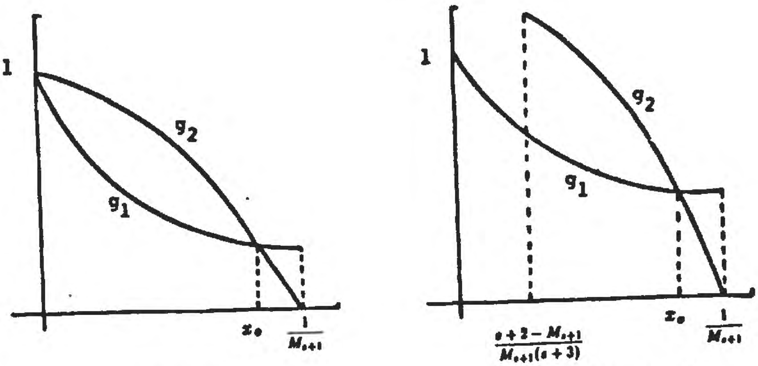
$g_2'(x) < 0$ and $g_2''(x) < 0$, for all $x \in (\frac{s+2-M_{s+1}}{M_{s+1}(s+3)}, \frac{1}{M_{s+1}})$,

which implies that g_2 is increasing in the interval $[0, \frac{s+2-M_{s+1}}{M_{s+1}(s+3)}]$ and decreasing and concave in the interval $[\frac{s+2-M_{s+1}}{M_{s+1}(s+3)}, \frac{1}{M_{s+1}}]$. Thus, there exists a unique point $x_o \in (\frac{s+2-M_{s+1}}{M_{s+1}(s+3)}, \frac{1}{M_{s+1}})$, such that

$$\begin{aligned} &g_1(x_o) = g_2(x_o), \\ &g_1(x) < g_2(x), \quad \text{for all } x \in (0, x_o), \text{ and} \\ &g_1(x) > g_2(x), \quad \text{for all } x \in (x_o, \frac{1}{M_{s+1}}). \end{aligned}$$

Since $g = \frac{g_1}{g_2}$ in the interval $[0, \frac{1}{M_{s+1}})$, the proof is complete.

Figure 1 below illustrate the behaviour of the functions g_1 and g_2 in the interval $[0, \frac{1}{M_{s+1}}]$ in a general situation.



$(M_{s+1} \geq s+2)$ $\qquad$ $(M_{s+1} < s+2)$

Figure 1. Functions g_1 and g_2 in a general situation

We note that the root of the equation $g(x) = 1$, $x \in [0, \frac{1}{M_{s+1}})$, is not of the form $x_o = \frac{1}{n}$, for all $n \in \mathcal{N}^* = \{1, 2, \ldots, \}$. Indeed, the equation $g(\frac{1}{n}) = 1$, for some $n \in \mathcal{N}^*$, $n > M_{s+1}$ is equivalent to

$$\text{(21)} \qquad \sum_{i=0}^{s+2} \binom{s+2}{i} n^i - n^2 \prod_{i=1}^{s} (n - n_i) - M_{s+1} \sum_{i=1}^{s+2} \binom{s+2}{i} n^{i-1} = \frac{M_{s+1}}{n}.$$

The left hand side of equation (21) is an integer. On the other hand, since the right hand side is not an integer, it follows that $x_o \neq \frac{1}{n}$, for all $n > M_{s+1}$. Let $m = max\{n_1, \ldots, n_s\}$ and

$$n_{s+1} = min\{n \in \mathcal{N}^*; \prod_{j=1}^{s} (M_{s+1} + n - n_j) < \frac{n}{(M_{s+1} + n)^3} (M_{s+1} + n + 1)^{s+2}\}.$$

The main result of this section is stated next.

Theorem 5.1. *For all $s \geq 2$ there is a unique mode of $\pi(N|\mathcal{D})$, $\hat{N}$, namely*

$$\hat{N} = \begin{cases} m; & M_{s+1} = m, \\ M_{s+1} + n_{s+1} - 1; & M_{s+1} > m. \end{cases}$$

Proof. From Lemma 5.1, $\hat{N} = M_{s+1}$ if $M_{s+1} = m$. If $m < M_{s+1}$, it follows from Lemma 5.2 that exists $n_o \in \mathcal{N}^*$, $n_o > 1$, such that for $n \in \mathcal{N}^*$,

$$g\left(\frac{1}{M_{s+1} + n}\right) \begin{cases} < 1; & n \in \mathcal{N}^*, \ n \geq n_o, \\ > 1; & n \in \mathcal{N}^*, \ n \leq n_o - 1. \end{cases}$$

Thus,

$$n_e = min\{n \in \mathcal{N}^*; g(\frac{1}{M_{s+1}+n}) < 1\}$$

$$= min\{n \in \mathcal{N}^*; \prod_{j=1}^{s}(M_{s+1}+n-n_j) < \frac{n}{(M_{s+1}+n)^3}(M_{s+1}+n+1)^{s+2}\},$$

that is, $n_e = n_{s+1}$ and from

$$g(\frac{1}{M_{s+1}+n}) = \frac{K(M_{s+1}+n)}{K(M_{s+1}+n-1)}, \quad n \in \mathcal{N}^*,$$

it follows that

$$K(M_{s+1}+n_{s+1}-1) > K(M_{s+1}+n_{s+1}-2) > \ldots > K(M_{s+1}),$$

and

$$K(M_{s+1}+n_{s+1}-1) > K(M_{s+1}+n_{s+1}) > K(M_{s+1}+n_{s+1}+1) > \ldots.$$

Thus, $\hat{N} = M_{s+1}+n_{s+1}-1$ is the unique mode of $\pi(N|\mathcal{D})$.

As a direct consequence of Theorem 5.1, we have

Corollary 5.1. *If $M_{s+1} > m$, then $\hat{N} = M_{s+1}$ if and only if*

$$\prod_{j=1}^{s}(M_{s+1}+1-n_j) < (\frac{1}{M_{s+1}+1})^3(M_{s+1}+2)^{s+2}.$$

6. An Approximate Posterior Distribution

An alternative way of eliminating the nuisance parameter p is by maximization, by using the Laplace method for integrals (Tierney and Kadane, 1986). In the sequel we illustrate the implications of this approach to the estimation of N. Numerical studies which are reported below show that the approximation is very accurate for approximating posterior densities of the population size N. The prior density we consider for (N, p) is given by

$$\pi(N,\mathrm{p}) = \pi(\mathrm{p}|N)\pi(N) = \frac{1}{(N+1)^2}. \tag{22}$$

Notice that we are using a uniform prior for p, that is,

$$\pi(\mathrm{p}) = 1, \quad 0 < p_i < 1,$$

$i = 1, \ldots, s$. Let $\hat{\mathrm{p}}(N)$ be the maximum of the joint posterior density $\pi(N, \mathrm{p}|\mathcal{D})$ for fixed N. It can be shown that $\hat{\mathrm{p}}(N) = (n_1/N, \ldots, n_s/N)'$. Furthermore, according to (4.1) in Tierney an Kadane (1986), it follows that the Laplace approximation for the posterior density of N is given by

$$\pi_L(N|\mathcal{D}) \propto \pi_p(N|\mathcal{D})|j(N,(\frac{n_1}{N},\ldots,\frac{n_s}{N}))|^{-1/2}, \tag{23}$$

where

$$|j(N,(\frac{n_1}{N},\ldots,\frac{n_s}{N}))| = det[-\frac{\partial^2 log(\pi(N,\mathbf{p}|\mathcal{D}))}{\partial \mathbf{p}^2}|_{\mathbf{p}=\hat{\mathbf{p}}(N)}],$$

and

$$\pi_p(N|\mathcal{D}) = \pi(N,\hat{\mathbf{p}}(N)|\mathcal{D}) \propto \frac{N!}{(N-M_{s+1})!(N+1)^2}\prod_{i=1}^{s}(\frac{n_i}{N})^{n_i}(1-\frac{n_i}{N})^{N-n_i}, \tag{24}$$

$N \geq M_{s+1}$, which is obtained from the joint posterior distribution $\pi(N,\mathbf{p}|\mathcal{D})$ by replacing $\mathbf{p}$ for its maximum likelihood estimator (with fixed N), $\hat{\mathbf{p}} = (\frac{n_1}{N},\ldots,\frac{n_s}{N})'$. After some algebraic manipulations, it can be verified that

$$|j(N,\hat{\mathbf{p}}(N))| = \frac{N^s}{\prod_{i=1}^{s}\frac{n_i}{N}(1-\frac{n_i}{N})}. \tag{25}$$

From (23)-(25), it follows then that the Laplace approximation for the marginal posterior distribution of N, is given by

$$\pi_L(N|\mathcal{D}) \propto \frac{N!}{(N-M_{s+1})!(N+1)^2N^{\frac{s}{2}}}\prod_{i=1}^{s}(\frac{n_i}{N})^{n_i+1/2}(1-\frac{n_i}{N})^{N-n_i+1/2}, \quad N \geq M_{s+1}. \tag{26}$$

Lemma 6.1. *The posterior mode $\hat{N}$ of $\pi_L(N|\mathcal{D})$ is always finite.*

Proof. Notice that we may write the posterior density (26) as

$$\pi_L(N|\mathcal{D}) \propto (1-\frac{1}{N})\ldots(1-\frac{(M_{s+1}-1)}{N})\frac{1}{(N+1)^2N^{\sum_{i=1}^{s}n_i-M_{s+1}+s}} \tag{27}$$

$$\cdot\prod_{i=1}^{s}(1-\frac{n_i}{N})^{N-n_i+1/2}.$$

From (27), it follows that if $M_{s+1} < \sum_{i=1}^{s}n_i$, then as $N \to \infty$,

$$\pi_L(N|\mathcal{D}) \to 1.0.e^{-\sum_{i=1}^{s}n_i} = 0.$$

On the other hand, if $M_{s+1} = \sum_{i=1}^{s}n_i$, it follows from (27) that as $N \to \infty$

$$\pi_L(N|\mathcal{D}) \to 1.0.e^{-\sum_{i=1}^{s}n_i} = 0.$$

Thus, in both cases, $\hat{N} < \infty$, which concludes the proof.

In order to get an explicit expression for the posterior mode of the density (26), we derive first an approximate expression for it by using the Stirling approximation.

Lemma 6.2. *For moderate and large N,*

$$\pi(N|\mathcal{D}) \dot{\propto} \frac{N!}{(N-M_{s+1})!(N+1)^2 N^s \prod_{i=1}^{s} \binom{N}{n_i}}, \tag{28}$$

where $\dot{\propto}$ means approximately proportional to.

Proof. From (26) it follows that

$$\pi_L(N|\mathcal{D}) \propto \frac{N!}{(N-M_{s+1})!(N+1)^2 N^{s/2}} \prod_{i=1}^{s} \frac{\sqrt{2\pi} n_i^{n_i+1/2} e^{-n_i} \sqrt{2\pi}(N-n_i)^{N-n_i+1/2} e^{-N+n_i}}{2\pi e^{-N} N^{N+1/2} N^{1/2}}$$

$$\dot{\propto} \frac{N!}{(N-M_{s+1})!(N+1)^2 N^s} \prod_{i=1}^{s} \frac{1}{\binom{N}{n_i}}, \quad N \geq M_{s+1},$$

by using the Stirling approximation, which concludes the proof.

Theorem 6.1. *If $M_{s+1} = max\{n_1, \ldots, n_s\}$, then $\hat{N} = M_{s+1}$ is the unique mode of the posterior density (28).*

Proof. Without loss of generality, we may assume that $M_{s+1} = n_s$. Since, from (28),

$$\pi_L(N|\mathcal{D}) \dot{\propto} \frac{\binom{N}{M_{s+1}}}{(N+1)^2 N^s \prod_{i=1}^{s} \binom{N}{n_i}}$$

$$= \frac{1}{(N+1)^2 N^s \prod_{i=1}^{s-1} \binom{N}{n_i}}, \quad N \geq n_s,$$

which implies that $\hat{N} = n_s = M_{s+1}$, as was to be proved.

We discuss now the case where $M_{s+1} > max\{n_1, \ldots, n_s\}$. Let

$$K(N) = \frac{N!}{(N-M_{s+1})!(N+1)^2 N^s \prod_{i=1}^{s} \binom{N}{n_i}}, \quad N \geq M_{s+1}$$

and define the function

$$g(x) = (1 - M_{s+1}x)^{-1}(1+x)^{-2}(1-x)^s \prod_{i=1}^{s}(1-n_i x),$$

$x \in [0, \frac{1}{M_{s+1}})$. It is clear that the posterior mode $\hat{N}$ of (28) maximizes $K(N)$. Further, it may be checked that

$$g(\frac{1}{N+1}) = \frac{K(N+1)}{K(N)}.$$

The proof of the next lemma parallels that of the Lemma 5.2 and is therefore ommited.

Lemma 6.3. *If $M_{s+1} > max\{n_1, \ldots, n_s\}$, then the equation $g(x) = 1$ has only one nonnull root x_o in the interval $(0, \frac{1}{M_{s+1}})$. For $x \in (0, x_o)$, $g(x) < 1$ and for $x \in (x_o, 1/M_{s+1})$, $g(x) > 1$.*

Let $m = max\{n_1, \ldots, n_s\}$ and

$$n_{s+1} = min\{n \in \mathcal{N}^*; (M_{s+1}+n-1)^s \prod_{i=1}^{s}(M_{s+1}+n-n_i) < n(M_{s+1}+n+1)^2(M_{s+1}+n)^{2s-3}\}.$$

Theorem 6.2. *For all $s \geq 2$ if $M_{s+1} = m$, then the unique mode of (28) is $\hat{N} = m$. Otherwise, if $M_{s+1} > m$, we have two possibilities:*

(i) if

$$(M_{s+1}+n_{s+1}-2)^s \prod_{i=1}^{s}(M_{s+1}+n_{s+1}-1-n_i) > (n_{s+1}-1)(M_{s+1}+n_{s+1})^2(M_{s+1}+n_{s+1}-1)^{2s-3},$$

then, the unique mode of (28) is $\hat{N} = M_{s+1} + n_{s+1} - 1$;

(ii) if

$$(M_{s+1}+n_{s+1}-2)^s \prod_{i=1}^{s}(M_{s+1}+n_{s+1}-1-n_i) = (n_{s+1}-1)(M_{s+1}+n_{s+1})^2(M_{s+1}+n_{s+1}-1)^{2s-3},$$

then, (28) has two modes, $\hat{N} = M_{s+1} + n_{s+1} - 1$ and $\hat{N} = M_{s+1} + n_{s+1} - 2$.

Proof. If $M_{s+1} = m$, it follows from Theorem 6.1 that $\hat{N} = m$. If $M_{s+1} > m$, it follows from Lemma 6.3 that there exists $n_o \in \mathcal{N}^*$, such that

$$g(\frac{1}{M_{s+1}+n})\begin{cases} < 1, & n \in \mathcal{N}^*,\ n \geq n_o \\ \geq 1, & n = n_o - 1\ (n_o - 1 \in \mathcal{N}^*) \\ > 1, & n \in \mathcal{N}^*,\ n < n_o - 1. \end{cases}$$

Thus,

$$n_o = min\{n \in \mathcal{N}^*; g(\frac{1}{M_{s+1}+n}) < 1\} = min\{n \in \mathcal{N}^*;$$
$$(M_{s+1}+n-1)^s \prod_{i=1}^{s}(M_{s+1}+n-n_i) < n(M_{s+1}+n+1)^2(M_{s+1}+n)^{2s-3}\},$$

that is, $n_s = n_{s+1}$ and assuming that condition (i) is satisfied we have that

$$g(\frac{1}{M_{s+1}+n})\begin{cases} < 1; & n \in \mathcal{N}^*, n \geq n_{s+1} \\ > 1; & n \in \mathcal{N}^*, n \leq n_{s+1} - 1. \end{cases}$$

Moreover, since

$$g(\frac{1}{M_{s+1}+n}) = \frac{K(M_{s+1}+n)}{K(M_{s+1}+n-1)}, \quad n \in \mathcal{N}^*,$$

it follows that

$$K(M_{s+1}+n_{s+1}-1) > K(M_{s+1}+n_{s+1}-2) > \ldots > K(M_{s+1})$$

and

$$K(M_{s+1}+n_{s+1}-1) > K(M_{s+1}+n_{s+1}) > K(M_{s+1}+n_{s+1}+1) > \ldots,$$

which implies that the unique mode of (28) is

$$\hat{N} = M_{s+1} + n_{s+1} - 1.$$

On the other hand, assuming that condition (ii) is satisfied, we have

$$g(\frac{1}{M_{s+1}+n})\begin{cases} < 1; & n \in \mathcal{N}^*, n \geq n_{s+1}, \\ = 1, & n = n_{s+1} - 1 \ (n_{s+1} - 1 \in \mathcal{N}^*), \\ > 1, & n \in \mathcal{N}^*, n < n_{s+1} - 1. \end{cases}$$

Furthermore, from the fact that

$$g(\frac{1}{M_{s+1}+n}) = \frac{K(M_{s+1}+n)}{K(M_{s+1}+n-1)}, \quad n \in \mathcal{N}^*,$$

it follows that

$$K(M_{s+1}+n_{s+1}-1) = K(M_{s+1}+n_{s+1}-2) > K(M_{s+1}+n_{s+1}-3) > \ldots > K(M_{s+1})$$

and

$$K(M_{s+1}+n_{s+1}-1) > K(M_{s+1}+n_{s+1}) > K(M_{s+1}+n_{s+1}+1) > \ldots,$$

which implies that the two modes of (28) are given by

$$\hat{N} = M_{s+1} + n_{s+1} - 1 \quad \text{and} \quad \hat{N} = M_{s+1} + n_{s+1} - 2,$$

so that the proof is now complete.

Computing programs (very short) can easily be written to compute the Bayes estimators that follows from Theorems 5.1 and 6.2. As the numerical results that follow show, the Laplace approximation given in Theorem 6.2 is very accurate.

Numerical Illustration 1. In this application, we consider the sunfish data that appears in Castledine (1981). For this population, $s = 14$ capture-recaptures were performed. The number of animals captured in each of the samples (n_i) were: 10,27,17,7,1,5,6,15,9,18,16,5, 7,19. For this experiment, $M_{s+1}=122$. Using Theorem 5.1, it follows that the Bayes estimator of N is $\hat{N}=206$. The Laplace approximation given in Theorem 6.2 yields exactly the same estimator.

Numerical Illustration 2. In this illustration, we consider several different situations. The results are recorded in Table 1 below.

Table 1. Examples of Posterior modes from Theorems 5.1 and 6.2

$(n_1,\ldots,n_s)$	M_{s+1}	$\hat{N}$ from Theorem 5.1	$\hat{N}$ from Theorem 6.2
(40,60,30)	130;128	1196;722	1195;722
(40,60)	100;90	698;196	698;196
(1,5,8)	14;12	23;15	23;15
(40,60,80)	180;179	2241;1863	2240;1863
(3,3,4,4,5)	19;17	35;26	35;26
(15,20,25,30,50)	140;130	1181;468	1182;468

References

Castledine, B. J. (1981). A Bayesian analysis of multiple capture-recapture sampling for a closed population. Biometrika, 68, 197-210.

Cox, D.R. and Reid, N. (1987). Orthogonal parameters and approximated conditional inference (with discussion). Journal of the Royall Statistical Society, B, 49, 1-39.

Leite, J.G., Oishi, J. and Pereira, C.A.B. (1988). A note on the exact maximum likelihood estimation of the size of a finite and closed population. Biometrika, 75, 178-180.

Seber G.A.F. (1982). *The Estimation of Animal Abundance*, 2nd ed. London: Griffin.

Sweeting, T. (1987). Comments on "Orthogonal parameters and approximated conditional inference", Journal of the Royall Statistical Society, B, 49,1-39.

Tierney L. and Kadane J. (1986). Accurate approximations for the posterior moments and marginal densities. Journal of the American Statistical Association, 81, 82-86.

ULTIMOS RELATORIOS TECNICOS PUBLICADOS

<u>1990</u>

9001 - JR., HODGES, J.L. RAMSEY, P.H. & WECHSLER, S.: Improved Significance Probabilities of the Wilcoxon Test, 30p.

9002 - PAULA, G.A.: Bias Correction for Exponential Family Nonlinear Models, 10p.

9003 - SCHONMANN, R.H. & TANAKA, N.I.: One Dimensional Caricature of Phase Transition, 16p.

9004 - ZACKS, S. & BOLFARINE, H.: Maximum Likelihood Prediction in Finite Populations, 15p.

9005 - ZACKS, S. & BOLFARINE, H.: Equivariant Prediction of the Population Variance Under Location-Scale Superpopulation Models. 8p.

9006 - BOLFARINE, H.: Finite Population Prediction Under Error in Variables Superpopulation Models, 14p.

9007 - BOLFARINE, H.: Ratio and Regression Estimators Under Error-In-Variables Superpopulation, 16p.

9008 - BOTTER, D.A. & MOTTA, J.M.: Experiments With Three-Treatment Three-Period Crossover Design: Analysis Through the General Linear Models, 15p.

9009 - SCHONMANN, R.H.: Finite Size Scaling Behavior of a Biased Majority Rule Cellular Automaton, 11p.

9010 - BUENO, V.C.: A Martingale Version of Multistate Monotone Systems, 5p.

9011 - CORDEIRO, G.M. & FERRARI, S.L.P.: A Modified Score Tests Statistic Having Chi-Squared Distribution to Order, 12p.

9012 - NEVES, E.J. & SCHONMANN, R.H.: Behavior of Droplets for a Class a Glauber Dynamics at very Low Temperature, 33pg.

9013 - FERRARI, P.A.: Schock Fluctuations in Asymmetric Simple Exclusion, 26p.

9014 - MAGALHÃES, M.N. & TANAKA, N.I.: Reverse Process in M/M/1 Queues, 7p.

9015 - CORDEIRO, G.M. & PAULA, G.A.: Estimation Large Sample Parametric Tests and Diagnostic for Non-Exponential Family Nonlinear Models, 31p.

9016 - CORDEIRO, G.M., FERRARI, S.L.P. & PAULA, G.A.: Improved Score Tests for Generalized Linear Models, 25p.

9017 - ESTON, V.R., GALVES, A. JACOBI, C.M., & LANGENVIN, R.: Dominance Switch Between Two Interacting Species and Large Deviations, 8p.

9018 - IRONY, T.Z., PEREIRA, C.A.B. & BARLOW, R.E.: Bayesian Models for Quality Assurance, 22p.

9019 - BOLFARINE, H. & ZACKS, S., Bayes and Minimax Prediction in Finite Populations, 16p.

9020 - BOLFARINE, H.: RODRIGUES, J. & ZACKS, S.: Some Asymptotic Results in Finite Populations, 13p.

9021 - BOLFARINE, H., CORDANI, L.K.: Estimation of a Structural Linear Regression Model With a Known Reliability Ratio, 9pg.

9022 - PAULINO, C.D.M., PEREIRA, C.A.B.: Bayesian Analysis of Categorical Data Informatively Censored. São Paulo, 18pg

9023 - FERRARI, P.A., MARTINEZ, S. & PICCO, P.: Some Properties of Quasi Stationary Distributions in the Birth and Death Chains, 11pg.

9024 - BOLFARINE, H., SANDOVAL, M.C.: Prediction of the finite population distribution function under Gaussian superpopultion models, 13 pg.

9025 - BUENO, V.C. & NORROS, I.; Component importance through compensator transforms, 7p

1991

9101 - RAVISHANKAR, K.; Convergence of a random motion to a point, 11p.

9102 - FERRARI, P.A. & KAVISHANKAR, R.; Shocks in asymmetric exclusion automata, 17p.

9103 - ANDRE, C.D.S. & PERES, C.A.; An Algorithm for the Msae Estimation of the Multistage Dose-Responde model, 14p.

9104 - FERRARI, P.A.; Microscopic shocks in one dimensional driven systems, 21p.

The complete list of Relatórios do Departamento de Estatística, IME-USP, will be sent upon request.

- Departamento de Estatística
  IME-USP - Caixa Postal 20.570
  01498 - São Paulo, Brasil