# Bayes Estimation of the Size of a Finite Population: Capture/Recapture Sequential Sample Data

José Galvão Leite; Carlos Alberto de Bragança Pereira

*International Statistical Review / Revue Internationale de Statistique*, Vol. 58, No. 3. (Dec., 1990), pp. 201-213.

*International Statistical Review* (1990), **58**, 3, pp. 201–213. Printed in Great Britain


# Bayes Estimation of the Size of a Finite Population: Capture/Recapture Sequential Sample Data

**José Galvão Leite and Carlos Alberto de Bragança Pereira**

*Universidade de São Paulo, Instituto de Matemática e Estatistica, C. Postal 20570, CEP 01496, São Paulo, SP Brazil*

**Summary**

**The Bayes' estimator of the population size, based on data obtained by the general capture/recapture sequential sampling process, is introduced. Properties related to the information contained in the data are studied. Also, some large-sample properties are obtained by using standard martingale results. The strongest results are the almost sure convergence of the Bayes' estimator to the true population size and of the Bayes' risk to zero. The Bayes' properties presented are restricted to proper priors having finite second moments. It is shown that the maximum likelihood estimator also converges almost surely to the population size.**

*Key words:* Bayes' estimator; Bayes' risk; Capture/recapture sequential sampling process; Martingale and supermartingale; Maximum likelihood estimator; Sufficient statistic.

## 1 Introduction

The objective of the present study is to show that the Bayes' estimator of the population size $N$, in addition to being consistent, has interesting properties, which are not shared by alternative classical procedures. As in Leite, Oishi & Pereira (1987, 1988), here we deal with a finite and closed population of size $N$, from which, using the capture/recapture sampling procedure, $k(>1)$ samples of sizes $m_i \geqslant 1$ $(i = 1, 2, \ldots, k)$ are sequentially selected. The sampling design for the capture/recapture sequential process and its sampling probability distribution are presented in the following section. For complete details see Leite & Pereira (1987). Section 3 introduces the Bayes' estimator and the Bayes' risk. Bayes' estimation of $N$, based on the capture/recapture sequential sampling process, was also studied by Freeman (1972) and Zacks (1984). However, both studies considered only the simple one-by-one case, that is $m_1 = \ldots = m_k = 1$. Monotone properties of the Bayes' estimator are presented in § 4. Large-sample properties of the Bayes' risk and estimator are presented in §§ 5 and 6. We also show that related results can also be obtained for the maximum likelihood estimator discussed by Leite, Oishi & Pereira (1987, 1988).

With the sampling procedure used in this paper, a minimal sufficient statistic for $N$ is $T_k$, the number of distinct units selected in the whole sample. It is not difficult to see that this statistic converges almost surely to the true value of $N$. This strong and simple result is the basis of the large-sample properties discussed in §§ 5 and 6. Results of § 6 use the language of martingales and supermartingales.

It is important to notice that when the sample increases (more information is collected) the variance (predictive) of the Bayes' estimator increases as shown in § 6. For someone who is accustomed to looking for minimum variance estimators this may be very unintuitive. However, it must be understood that this variance is taken under the marginal (predictive) distribution of the data since, to compute the Bayes' estimator, the parameter $N$ is eliminated by integration under the conditional distribution of $N$ given the data, the posterior distribution. It is intuitively clear, on the other hand, that the Bayes' estimator would be perfect if the posterior mean, $E\{N \mid T_k\}$, is equal to $N$. In this case, the maximum variance is attained since

$$\text{Var}\{N\} = \text{Var}\{E\{N \mid T_k\}\} + E\{\text{Var}\{N \mid T_k\}\}.$$

Note that any reasonable person should try to decrease the posterior variance, and consequently its expectation $E\{\text{Var}\{N \mid T_k\}\}$, to zero. This corresponds to increasing the variance of the Bayes' estimator to Var $\{N\}$, its maximum possible value.

## 2 The Statistical Model

Consider a population of finite size, $N\ (\in \mathbb{N} = \{0, 1, \ldots\})$, which does not change in size or in form during the study time. From this population, $k(>1)$ samples are sequentially selected at random. Each sample is returned back to the population before the next one is selected. To obtain the relevant data for estimating $N$, the following steps are performed.

(i) The first random sample of size $m_1(\geqslant 1)$ is drawn, without replacement. After the sample units are marked they are returned to the population and the number $U_1 = m_1$ is recorded.

(ii) The $j$th $(j > 1)$ random sample of size $m_j$ is drawn, without replacement. The sample units that have been previously marked are immediately returned to the population. The remaining $U_j$ unmarked sample units are marked and returned to the population. The numbers $m_j$ and $U_j$ are recorded.

After the $k$ samples have been observed, the data random vector,

$$\mathbb{D}_k = (U_1, \ldots, U_k),$$

assumes the observed point

$$\mathcal{D}_k = (u_1, \ldots, u_k),$$

where $u_1 = m_1$ and $u_j \in \{0, 1, \ldots, m_j\}$ for $j = 2, \ldots, k$.

Note that the statistic

$$T_k = U_1 + \ldots + U_k$$

is the number of distinct units selected in the whole sampling process. Leite & Pereira (1987) show that this statistic is sufficient for $N$. Moreover, the *Likelihood kernel,* which is a minimal sufficient statistic (Zacks, 1981), is given by

$$\lambda(n, \mathcal{M}_k, t) = \mathcal{I}_t(n)\left\{(n-t)!\prod_{j=1}^{k}\binom{n}{m_j}\right\}^{-1} n!,$$

where $t$ is the observed value of $T_k$, $\mathcal{I}_t(n)$ is the indicator function of $\mathbb{N}_t = \{x \in \mathbb{N}, x \geqslant t\}$ evaluated at point $n$, and $\mathcal{M}_k$ is the vector $(m_1, \ldots, m_k)$. The families of probability distributions of $\mathbb{D}_k$ and $T_k$ (Leite & Pereira, 1987) are given by

$$\begin{aligned} P\{\mathbb{D}_k = \mathcal{D}_k \mid N = n\} &= \lambda(n, \mathcal{M}_k, t)\mathcal{I}_A(t)\left\{\prod_{j=2}^{k}\binom{t_{j-1}}{m_j - u_j}\right\} \div \left\{\prod_{j=1}^{k}(u_j)!\right\}, \\ P\{T_k = t \mid N = n\} &= \lambda(n, \mathcal{M}_k, t)\mathcal{I}_A(t)\sum_{i=0}^{t}\frac{(-1)^{t-i}}{i!\,(t-i)!}\prod_{j=1}^{k}\binom{i}{m_j}, \end{aligned} \qquad (2.1)$$

where $\mathscr{I}_A(t)$ is the indicator function of the set

$$A = \{x \in \mathbb{N}; \max\{m_1, \ldots, m_k\} \leqslant x \leqslant \min\{n, (m_1 + \ldots + m_k)\}\}$$

evaluated at point $t$ and for $j = 1, \ldots, k$, $t_j = u_1 + \ldots + u_j$. Note that $A$ depends on the value $n$ of $N$.

In the following sections, after introducing a prior probability function for $N$, $P\{N = n\} = \pi(n)$, we discuss some properties of the posterior probability function,

$$P\{N = n \mid \mathbb{D}_k = \mathscr{D}_k\} = P\{N = n \mid T_k = t\} = \pi(n \mid k, t),$$

of the posterior mean or Bayes' estimator,

$$E\{N \mid \mathbb{D}_k = \mathscr{D}_k\} = E\{N \mid T_k = t\} = \beta(k, t),$$

and of the posterior variance or Bayes' risk,

$$V\{N \mid \mathbb{D}_k = \mathscr{D}_k\} = V\{N \mid T_k = t\} = \rho(k, t).$$

Note that $\pi$, $\beta$ and $\rho$ also depend on $\mathscr{M}_k$. In the sequel, all the functions depending on $k$ also should depend on $\mathscr{M}_k$ and, following this rule, we write $\lambda(n, k, t)$ for $\lambda(n, \mathscr{M}_k, t)$. In addition, we let $S_j = m_1 + \ldots + m_j$ and $M_j = \max\{m_1, \ldots, m_j\}$, for all $j = 1, \ldots, k$.

## 3 Bayes' Estimation

Let $\pi$ be a prior probability function for $N$ and let

$$\mathbb{N}_t^\pi = \{x \in \mathbb{N}; x \geqslant t, \pi(x) > 0\}.$$

For all $t \in \mathbb{N}$ such that $M_k \leqslant t \leqslant S_k$ and $\mathbb{N}_t^\pi \neq \varnothing$, the posterior probability function of $N$ is given by

$$\pi(n \mid k, t) = \lambda(n, k, t)\kappa(k, t)\pi(n)\mathscr{I}_t^\pi(n), \tag{3.1}$$

where $\mathscr{I}_t^\pi(n)$ is the indicator function of $\mathbb{N}_t^\pi$ evaluated at point $n$ and

$$\kappa(k, t) = \left\{\sum_{n=t}^{\infty} \frac{n!\,\pi(n)}{(n-t)!\prod_{j=1}^{k}\binom{n}{m_j}}\right\}^{-1}. \tag{3.2}$$

Using the fact that (see Appendix 1)

$$\lambda(n, k, t) \leqslant \left\{1 - \frac{M_k - 1}{t}\right\}^{-t} \prod_{j=1}^{k} (m_j!), \tag{3.3}$$

one may easily prove that $\kappa(k, t)$ is positive and bounded. Note that $M_k \leqslant t \leqslant S_k$ is a natural restriction since: (a) $t < m_j$, for some $j$, would happen only if we had selections with replacement; and (b) $t > S_k$ would happen only if, before the selection process starts, there already existed marked population units.

It is difficult to define a workable conjugate class of distributions for this problem since, for some sample points, the sum of the likelihood over all possible values of $N$, $\{n \in \mathbb{N}; n \geqslant t\}$, diverges. For instance, considering the improper uniform measure on $\mathbb{N}$, for the one-by-one case where $m_1 = \ldots = m_k = 1$, $\pi(n \mid k, t)$ would not be defined for $t = k - 1$ and $t = k$ since,

$$\frac{1}{\kappa(k, t)} = \sum_{n=t}^{\infty} \lambda(n, k, t) = \sum_{n=t}^{\infty} \frac{n!}{n^k (n-t)!} \tag{3.4}$$

converges only if $t \leqslant k - 2$.

Considering only proper prior distributions is not restricting the practical applicability of the Bayes' method in the present problem. Usually the space (a lake for example) occupied by the population of interest is limited, permitting only the accommodation of a finite number of population units (fishes in the lake). Even when the maximum possible number of population units is taken to be very large, the Bayes' solution for the estimation problem is obtainable. Note that the choice of this supposed maximum number is a very important and delicate matter. For instance, if one observes a value of $T_k$, $t$, larger than the number chosen for this maximum, one must agree that the chosen prior opinion used was wrong. This is a example of a problem where *open-minded* prior must be used. Following a personal communication from David Blackwell in 1986, we consider as open-minded priors all probability functions that assign positive probabilities for all physically possible values of the parameter. This yields the restriction that the set $\mathbb{N}_t^\pi = \{x \in \mathbb{N}; x \geqslant t, \pi(x) > 0\}$ must be nonempty. This restriction creates a slight logical problem since it relates the prior distribution to the observation $t$. However in practice, by knowing the size of the location that accommodates the population, one may consider positive probabilities (although very small for some points) to all physically possible values of $N$. To avoid these problems, we will consider only proper prior distributions that assign positive probabilities to any non-negative integer.

Let $\pi(n)$ be a prior probability function with a finite second moment. For all $t \in \mathbb{N}$ such that $M_k \leqslant t \leqslant S_k$ and $\mathbb{N}_t^\pi \neq \varnothing$, the Bayes' Estimate (BE) of $N$ is given by

$$\beta(k, t) = \sum_{n=t}^{\infty} n\pi(n \mid k, t) = \kappa(k, t) \sum_{n=t}^{\infty} n\lambda(n, k, t)\pi(n). \tag{3.5}$$

Due to inequality (3.3) and to the fact that $\pi$ has a second moment, $\beta(k, t)$ is finite. For this one-by-one sampling, with $t \leqslant k$ and $\mathbb{N}_t^\pi \neq \varnothing$, we have

$$\beta(k, t) = \frac{\kappa(k, t)}{\kappa(k-1, t)}. \tag{3.6}$$

Before discussing the properties of the BE we present examples with Poisson prior distributions; that is $\pi(n) = (n!)^{-1}\theta^n \exp\{-\theta\}$, for $n \in \mathbb{N}$.

*Example* 1. For the one-by-one case with Poisson prior, the BE is given by

$$\beta(k, t) = \frac{E\{(N+t)^{-k+1}\}}{E\{(N+t)^{-k}\}},$$

where $E\{.\}$ is the expectation operator and $N$ is the random variable having the prior Poisson distribution with parameter $\theta > 0$. Table 1 presents the values of BE for $\theta = 20$ (prior mean or variance) and $k = 10$, 12 and 15. In order to evaluate the influence of the use of the prior information, we also present, in parentheses, the maximum likelihood estimates (MLE). The theory of the MLE under the general capture/recapture sampling process is presented by Leite, Oishi & Pereira (1987, 1988). In the present example, it is interesting to notice that the MLE and the BE yield close values for small, and more informative, values of $t$. The MLE diverges as $t$ increases. This fact shows that the influence of the prior information is stronger when the data is less informative; that is when $t$ is large.

*Example* 2. For the two-by-two case ($m_1 = \ldots = m_k = 2$) with Poisson prior, the BE is given by

$$\beta(k, t) = \left\{\sum_{n=t}^{\infty} \frac{n\theta^n}{[(n-1)n]^k (n-t)!}\right\} \Big/ \left\{\sum_{n=t}^{\infty} \frac{\theta^n}{[(n-1)n]^k (n-t)!}\right\} \quad (t = 2, 3, \ldots, 2k).$$

**Table 1**

*Bayes' estimates of N for Poisson prior with parameter $\theta = 20$ (maximum likelihood estimates)*

| $t$ | $k = 10$ | $k = 12$ | $k = 15$ |
|---|---|---|---|
| 1 | (1) 1·0450 | (1) 1·0061 | (1) 1·0006 |
| 2 | (2) 4·7857 | (2) 2·4358 | (2) 2·0628 |
| 3 | (3) 10·2049 | (3) 6·0156 | (3) 3·5704 |
| 4 | (4) 13·1565 | (4) 9·8999 | (4) 6·0322 |
| 5 | (5) 15·2625 | (5) 12·6860 | (5) 8·9280 |
| 6 | (8) 17·0382 | (7) 14·8565 | (6) 11·5406 |
| 7 | (12) 18·6396 | (9) 16·7130 | (8) 13·7788 |
| 8 | (19) 29·1317 | (12) 18·3865 | (10) 15·7434 |
| 9 | (42) 21·5481 | (18) 19·9408 | (12) 17·5215 |
| 10 | (∞) 22·9085 | (29) 21·4111 | (16) 19·1690 |
| 11 | — | (62) 22·8189 | (21) 20·7212 |
| 12 | — | (∞) 24·1781 | (30) 22·2011 |
| 13 | — | — | (48) 23·6244 |
| 14 | — | — | (100) 25·0023 |
| 15 | — | — | (∞) 26·3430 |

*Example* 3. If in the preceding examples we take $k = 3$, $m_1 = 1$, $m_2 = 2$ and $m_3 = 3$ then the BE is given by

$$\beta(3, t) = \left\{\sum_{n=t}^{\infty} \frac{n\theta^n}{(n-2)(n-1)^2 n^3 (n-t)!}\right\} \Big/ \left\{\sum_{n=t}^{\infty} \frac{\theta^n}{(n-2)(n-1)^2 n^3 (n-t)!}\right\} \quad (t = 3, 4, 5, 6).$$

## 4 Basic Properties of the Bayes' Estimator

The study presented here and in the remaining sections is restricted to proper prior distributions with finite second moments. Recall that: (a) the BE, $\beta(k, t)$, is a function of $\mathcal{M}_k = (m_1, \ldots, m_k)$; (b) $M_k \leq t \leq S_k$; and (c) $\mathbb{N}_t^\pi \neq \varnothing$. Also recall that the probability function $\pi$ is said to be degenerate if its support has only one point. The following results show that $\beta$ is a non-increasing function of $k$ and a non-decreasing function of $t$.

THEOREM 1. *For all $k \geq 2$ and $t \in \mathbb{N}$, let $M_k \leq t \leq S_k$ and $\mathbb{N}_t^\pi \neq \varnothing$. If $m_{k+1} \leq S_k$, then*

$$\beta(k, t) \geq \beta(k+1, t). \tag{4.1}$$

*Equality holds if the prior probability function, $\pi$, is degenerate.*

*Proof.* Considering the restrictions, define the following decreasing function of $n \in \mathbb{N}$ ($\mathcal{I}_t(n)$ is the indicator function of $\mathbb{N}_t$):

$$h(n) = \mathcal{I}_t(n) \Big/ \binom{n}{m_k}.$$

It follows (Lehmann, 1966) that

$$E\{h(N) \mid T_k = t\} E\{N \mid T_k = t\} \geq E\{Nh(N) \mid T_k = t\}.$$

It is simple to check that

$$E\{h(N) \mid T_k = t\} = \frac{\kappa(k, t)}{\kappa(k+1, t)}, \quad E\{Nh(N) \mid T_k = t\} = \frac{\kappa(k, t)}{\kappa(k+1, t)} E\{N \mid T_{k+1} = t\};$$

that is

$$\frac{\kappa(k, t)}{\kappa(k+1, t)}\beta(k, t) \geqslant \frac{\kappa(k, t)}{\kappa(k+1, t)}\beta(k+1, t),$$

which completes the proof. □

THEOREM 2. *For all $k \geqslant 2$ and $t \in \mathbb{N}$, let $M_k \leqslant t \leqslant S_k - 1$ and $N_t^\pi \neq \varnothing$. If $\pi(n \mid k, t)$ is not degenerate at point $t$, then*

$$\beta(k, t) \leqslant \beta(k, t+1). \tag{4.2}$$

*Equality holds if the prior probability function, $\pi$, is degenerate at any point but $t$.*

*Proof.* Considering the restrictions, we note that (cov is for covariance)

(i) $$\rho(k, t) = V\{N \mid T_k = t\} = \text{cov}\{N, N-t \mid T_k = t\},$$

(ii) $$E\{N(N-t) \mid T_k = t\} = \frac{\kappa(k, t)}{\kappa(k, t+1)}\beta(k, t+1),$$

(iii) $$E\{N - t \mid T_k = t\} = \frac{\kappa(k, t)}{\kappa(k, t+1)}.$$

These yield the following formula

$$\rho(k, t) = \frac{\kappa(k, t)}{\kappa(k, t+1)}\{\beta(k, t+1) - \beta(k, t)\}. \tag{4.3}$$

If $\pi(n \mid k, t)$ is not degenerate at point $t$, then $\rho$ is a well defined non-negative function and the proof is completed since both factors in the right side of (4.3) are non-negative. □

The following example will show that similar results do not hold for the function $\rho(k, t)$, the Bayes' risk.

*Example* 4. For the one-by-one case $(m_1 = \ldots = m_k = 1)$ with Poisson prior with parameter $\theta = 100$, we have

$$\rho(9, 5) = 99{\cdot}98439 < \rho(10, 5) = 100{\cdot}04162, \quad \rho(10, 6) = 99{\cdot}91608 > \rho(10, 7) = 99{\cdot}76708.$$

This example shows that $\rho$, unlike $\beta$, is neither monotone decreasing in $k$, for each fixed $t$, nor monotone increasing in $t$, for each fixed $k$. However, the results introduced in the sequel show that, for large samples, both $\rho$ and $\beta$ have desirable properties.

## 5 Large-Sample Properties

In this section we introduce two simple large-sample properties of the BE and discuss an interesting property of $T_k$, the sufficient statistic. As a consequence of this property, it is shown that the BE and the MLE converge almost surely to $N$ in the classical sense. For the two properties below, the value $t$ of $T_k$ is held fixed when $k$ increases.

Since we deal with a finite population, $\{m_j\}_{j\geqslant 1}$ is a bounded sequence of elements of $\mathbb{N}$ with $M = \max\{m_j; j \geqslant 1\}$. As before, $m_j$ is the size of the $j$th sample and, for all $t \in \mathbb{N}$ such that $t \geqslant M$ and $\mathbb{N}_t^\pi \neq \varnothing$, define $s = \min\{j \in \mathbb{N}, j \geqslant 2 \text{ and } S_j \geqslant t\}$. For all $k \geqslant s$, we have that $M_k \leqslant t \leqslant S_k$ and consequently both $\beta(k, t)$ and $\rho(k, t)$ are well defined. From Theorem 1 and the fact that $\beta(k, t) \geqslant 1$ for all $k \geqslant s$, the sequence $\{\beta(k, t)\}_{k\geqslant s}$ for a fixed $t$ has a finite limit when $k$ increases to infinite. The value of this limit is given by the following result that is proved in Appendix 2, since we have a long proof.

THEOREM 3. *For a fixed* $t \in \mathbb{N}$, *if* $t \geq M$, $\mathbb{N}_t^\pi \neq \varnothing$, *and* $\tau = \min \mathbb{N}_t^\pi$, *then*

$$\lim_{k \to \infty} \beta(k, t) = \tau.$$

*If* $\pi(t) > 0$, *then* $\tau = t$.

The convergence of the Bayes' risk sequence, $\{\rho(k, t)\}_{k \geq s}$, to zero is stated next.

THEOREM 4. *For a fixed* $t \in \mathbb{N}$, *if* $t \geq M$, *and* $\mathbb{N}_t^\pi \neq \varnothing$, *then*

$$\lim_{k \to \infty} \rho(k, t) = 0.$$

*Proof.* If $\pi(n \mid s, t)$ is degenerate then so is $\pi(n \mid k, t)$ for all $k \geq s$ and the result holds. If $\pi(n \mid s, t)$ is not degenerate then, for all $k \geq s$, $\pi(n \mid k, t)$ is not degenerate at the point $t$ and from (4.3) we have that

$$\rho(k, t) = \{\beta(k, t) - t\}\{\beta(k, t+1) - \beta(k, t)\}.$$

Since $M_k \leq t \leq S_k$, $\beta(k, t+1)$ is well defined and, using Theorem 3, the present result will follow. □

The following simple, strong result is the most important large-sample property under the classical statistics perspective. Together with the above results it shows that the BE is also a good estimator under the classical perspective. This result is formally presented in Corollary 1 of § 6.

THEOREM 5. *Considering only the process defined by* $\{P\{T_k = t \mid N = n\}\}_{k \geq 1}$, *the minimal sufficient statistic*, $T_k$, *converges almost surely to* $n$, *for any fixed value*, $n \in \mathbb{N}$, *of* $N$.

*Proof.* To prove this result we consider an analogy with the random selection of balls in an urn. First consider the one-by-one case; that is consider an urn with $n$ balls from which we select sequentially and randomly, with replacement, $k$ balls. If $t$ is the number of distinct balls selected in the first $k$ selections, then $(t/n)^m$ is the probability that only these $t$ distinct balls are going to be selected in the next $m$ draws. It is clear that when $m$ increases this probability decreases. With similar arguments we can prove that $T_k$ converges to $n$ almost surely as $k$ increases to infinity. Now suppose that more than one ball is drawn without replacement in each of the $k$ selection steps. It is clear that in this case the velocity of the convergence increases. Then, the proof for the one-by-one case solves in fact the general case. □

Putting together Theorems 3, 4 and 5 we can conclude that, in the case of an open-minded prior, the Bayes' estimator converges almost surely, under the process $\{P\{T_k = t \mid N = n\}\}_{k \geq 1}$, to the value $n$ of the population size, $N$. This pointwise convergence can also be proved for the maximum likelihood estimator introduced by Leite, Oishi & Pereira (1987, 1988). Recall that the MLE is given by $\hat{N}$ which is (i) equal to $T_k$ if $T_k = M_k$, (ii) equal to $\infty$ if $T_k = S_k$, and (iii) equal to $T_k + R_k - 1$ if $M_k < T_k < S_k$, where

$$R_k = \min \left\{ n \in \mathbb{N} : \prod_{j=1}^{k} (T_k + n - m_j) < n(T_k + n)^{k-1} \right\}.$$

To obtain the convergence of the MLE one only needs to prove that the MLE assumes only one value for large $k$ and that $R_k$ converges to one. The proofs of these facts are simple but long.

## 6 Large Samples, Martingales, and Supermartingales

In the present section we study large-sample properties under the Bayes' model. The conditional probability space of $\mathbb{D}_k$ given $N$, the statistical model, and the probability space of $N$, the prior model, are carefully stated in order to produce the precise definition of the joint probability space of $(N, \mathbb{D}_k)$.

As before, we consider a bounded sequence of integers, $\{m_j\}_{j\geq 1}$, with

$$M = \sup\{m_j; j \geq 1\}, \quad M_j = \max\{m_1, \ldots, m_j\}, \quad S_j = m_1 + \ldots + m_j \quad (j \geq 1).$$

Define also the following sets:

$\mathbb{N}_M = \{x \in \mathbb{N}; x \geq M\}$,

$\mathbb{A}_j = \{0, 1, \ldots, m_j\}$ for all $j \geq 1$, and

$\Omega = \{\omega = (\omega_1, \omega_2, \ldots); \omega_j \in \mathbb{A}_j, j = 1, 2, \ldots\}$.

Let $\mathscr{F}$ be the $\sigma$-algebra generated by the sets

$$\{\omega \in \Omega; \omega_1 \in \mathbb{B}_1 \subset \mathbb{A}_1, \omega_2 \in \mathbb{B}_2 \subset \mathbb{A}_2, \ldots, \omega_j \in \mathbb{B}_j \subset \mathbb{A}_j\} \quad (j = 1, 2, \ldots).$$

The measurable space $(\Omega, \mathscr{F})$ is the space of experimental observations based in $\mathbb{D}_k$.

Recalling that, for all $j \geq 1$, $t_j = u_1 + \ldots + u_j$, where $u_1 = m_1$ and $u_j \in \mathbb{A}_j$, we consider the family $\{P_n; n \in \mathbb{N}_M\}$ of probability measures on $(\Omega, \mathscr{F})$, defined for all positive integers as

$$P_n\{\omega \in \Omega; \omega_1 \in \mathbb{A}_1, \omega_2 \in A_2, \ldots, \omega_k \in \mathbb{A}_k\} = P\{\mathbb{D}_k = \mathscr{D}_k \mid N = n\}, \qquad (6.1)$$

where $P\{\mathbb{D}_k = \mathscr{D}_k \mid N = n\}$ is defined by (2.1). Analogously, we can write

$$P_n\{\omega \in \Omega; \omega_1 + \ldots + \omega_k = t; M_j \leq t \leq \min\{S_j, n\}, j \leq k\} = P\{T_k = t \mid N = n\}. \quad (6.2)$$

The triplet $(\Omega, \mathscr{F}, \{P_n; n \in \mathbb{N}_M\})$ is the *Statistical Space* or *Statistical Model*. Consider the $\sigma$-algebra $\mathscr{E}$ of subsets of $\mathbb{N}_M$ and a probability distribution $\pi$ on $(\mathbb{N}_M, \mathscr{E})$. The probability space $(\mathbb{N}_M, \mathscr{E}, \pi)$ is the *Prior Model*. To complete the construction of the Bayes' framework, we define the following entities:

(a) the cartesian product $\Omega^* = \mathbb{N}_M \times \Omega$;

(b) the smallest $\sigma$-algebra, $\mathscr{F}^*$, containing the set of cartesian products of elements of $\mathscr{E}$ times the elements of $\mathscr{F}$, that is

$$\mathscr{F}^* = \sigma(\{E \times F, E \in \mathscr{E}, F \in \mathscr{F}\});$$

(c) a probability measure $\Pi$ on $(\Omega^*, \mathscr{F}^*)$ defined by

$$\Pi(E \times F) = \sum_{n \in E} P_n(F)\pi(n)$$

for all $E \in \mathscr{E}$ and $F \in \mathscr{F}$.

The triplet $(\Omega^*, \mathscr{F}^*, \Pi)$ is the Bayesian model. For every $k > 1$ and all points $(n, \omega) \in \Omega^*$, the quantities of interest, $N$, $\mathbb{D}_k$ and $T_k$ can be viewed as random entities defined on $(\Omega^*, \mathscr{F}^*, \Pi)$ as follows:

$$N(n, \omega) = n, \quad \mathbb{D}_k(n, \omega) = (\omega_1, \omega_2, \ldots, \omega_k), \quad T_k(n, \omega) = \sum_{i=1}^{k} \omega_i.$$

Using the Bayesian structure defined above, we can state the following results about the random sequence $\{T_k\}_{k\geq 1}$. Below, we write $X_k \to Y\,[p]$ (or $X_k \to Y$ a.s. $[p]$) to indicate that, when $k$ increases to $\infty$, $X$ converges in probability (or almost surely) to $Y$

under the probability model $p$. In fact any statement followed by a.s. $[p]$ means that the statement is true almost surely (equivalently, with probability 1) under the probability model $p$.

LEMMA 1. *We have that:*

(i) *for every fixed* $k \geq 1$, $T_k \leq N$ *a.s.* $[\Pi]$; *that is*

$$\Pi\left(\left\{(n, \omega) \in \Omega^*; \sum_{i=1}^{k} \omega_i \leq n\right\}\right) = 1;$$

(ii) $T_k \to n$ $[P_n]$; *that is, for any* $n \in \mathbb{N}$,

$$\lim_{k\to\infty} P_n\left(\left\{\omega \in \Omega; \sum_{i=1}^{k} \omega_i = n\right\}\right) = 1;$$

(iii) $T_k \to N[\Pi]$; *that is*

$$\lim_{k\to\infty} \Pi\left(\left\{(n, \omega) \in \Omega^*; \sum_{i=1}^{k} \omega_i = n\right\}\right) = 1.$$

Item (i) is consequence of the definition of $P_n$ since we had the restriction

$$P_n\left(\left\{\omega \in \Omega; \sum_{i=1}^{k} \omega_i > n\right\}\right) = 0$$

for every $k \geq 1$.

The proof of item (ii) is left to Appendix 3, and to prove item (iii) we recall the Bounded Convergence Theorem to write

$$\lim_{k\to\infty} \Pi\{T_k = N\} = \sum_{n\in\mathbb{N}} \pi(n) \lim_{k\to\infty} P_n\left(\left\{\omega \in \Omega; \sum_{i=1}^{k} \omega_i = n\right\}\right) = \sum_{n\in\mathbb{N}} \pi(n) = 1.$$

The next result is a formal version of Theorem 5 which is a consequence of Lemma 1.

COROLLARY 1. *We have that:*

(iv) $T_k \to n$ *a.s.* $[P_n]$; *that is, for any* $n \in \mathbb{N}$,

$$P_n\left(\left\{\omega \in \Omega; \lim_{k\to\infty} \sum_{i=1}^{k} \omega_i = n\right\}\right) = 1;$$

(v) $T_k \to N$ *a.s.* $[\Pi]$; *that is*

$$\Pi\left(\left\{(n, \omega) \in \Omega^*; \lim_{k\to\infty} \sum_{i=1}^{k} \omega_i = n\right\}\right) = 1.$$

*Proof.* The proof is simple. The sequence $\{T_k\}_{k\geq 1}$ is nondecreasing and by definition $T_k \leq n$ a.s. $[P_n]$ and, from Lemma 1, $T_k \leq N$ a.s. $[\Pi]$. Consequently, there exists a random variable $L$ such that $T_k \to L$ a.s. $[P_n]$ and $T_k \to L$ a.s. $[\Pi]$. Using again Lemma 1, $T_k \to n$ $[P_n]$ and $T_k \to N$ $[\Pi]$ imply that $L = n$ a.s. $[P_n]$ and $L = N$ a.s. $[\Pi]$. □

Consider the increasing sequence, $\{\mathscr{F}_k\}_{k\geq 1}$, of sub-$\sigma$-algebras of $\mathscr{F}^*$ induced by the experimental observable sequence, $\{\mathbb{D}_k\}_{k\geq 1}$. That is $\mathscr{F}_k = \{\mathbb{D}_k^{-1}(\mathbb{A}): \mathbb{A} \subset (\mathbb{A}_1 \times \ldots \times \mathbb{A}_k)\}$. The Bayes' estimator, $\beta_k$, is defined as the conditional expectation of $N$ given $\mathscr{F}_k$; that is, for all $k \geq 1$,

$$\beta_k = E\{N \mid \mathscr{F}_k\}. \tag{6.3}$$

Recall that we are considering only prior distributions with finite second moments. The

Bayes' risk is then defined here as the conditional expectation of $(N-\beta_k)^2$ given $\mathcal{F}_k$ and we write, for all $k>1$,

$$\rho_k = E\{(N-\beta_k)^2 \mid \mathcal{F}_k\}. \tag{6.4}$$

Note that, since $T_k$ is a sufficient statistic, the Bayes' estimate, $\beta(k, t)$, and the posterior variances, $\rho(k, t)$, introduced before are in fact the observed values of $\beta_k$ and $\rho_k$, respectively. We list below some standard properties. Here and in remaining part of the paper all the results are related only to the Bayes' model $\Pi$.

LEMMA 2. *We have that:*

(vi) $\{\beta_k\}_{k\geq 1}$ *is a martingale relative to* $\{\mathcal{F}_k\}_{k\geq 1}$;

(vii) $\{\rho_k\}_{k\geq 1}$ *is a supermartingale relative to* $\{\mathcal{F}_k\}_{k\geq 1}$;

(viii) $\{\beta_k\}_{k\geq 1}$ *converges almost surely* $[\Pi]$ *to a random variable defined on* $(\Omega^*, \mathcal{F}^*, \Pi)$;

(ix) $E\{\rho_k\} \geq E\{\rho_{k+1}\}$; *that is, is nondecreasing in expectation.*

*Proof.* The proof here is also straightforward since (vi) and (vii) are direct consequences of standard properties of conditional expectations, (viii) is Theorem 4.3 of Doob (1953, p. 331), and (ix) is a direct consequence of (vii). □

Next we introduce the main results of this paper. We recall the fact that we are considering proper priors with finite second moment.

THEOREM 6. *We have that:*

(x) *the Bayes' estimator converges almost surely* $[\Pi]$ *to the random variable (population size)* $N$; *that is* $\beta_k \to N$ *a.s.* $[\Pi]$;

(xi) *the Bayes' risk converges almost surely* $[\Pi]$ *to zero; that is* $\rho_k \to 0$ *a.s.* $[\Pi]$.

The proof of Theorem 6 is left to Appendix 4 because, although short, it is very technical. This theorem is important since it shows a strong result for the Bayes' estimator and also shows that a good stopping rule shall depend on the Bayes' risk.

We end this paper with a result about the variance (predictive) of the Bayes' estimator. Note that the Bayes' estimator is a function of the data and its moments are based on the marginal distribution of the data, called predictive distribution.

COROLLARY 2. *The variance of the Bayes' estimator increases to the prior variance as the number of samples increases; that is* $\operatorname{Var}\{\beta_k\} \uparrow \operatorname{Var}\{N\}$ *as* $k \to \infty$.

*Proof.* To prove this result we recall that $\{\operatorname{Var}\{\beta_k\}\}_{k\geq 1}$ is nondecreasing and $\operatorname{Var}\{\beta_k\} \leq \operatorname{Var}\{N\}$. Then

$$\lim \operatorname{Var}\{\beta_k\} \leq \operatorname{Var}\{N\}.$$

On the other hand, using Fatou's Lemma, we have that

$$\liminf_k E\{\beta_k^2\} \geq E\left\{\liminf_k \beta_k^2\right\} = E\{N^2\}.$$

Hence, we have that $\lim \operatorname{Var}\{\beta_k\} \geq \operatorname{Var}\{N\}$ and the result follows. □

A final remark is that Lemma 1 and Corollary 1 are results related to $T_k$, the sufficient statistic. From them we can conclude that $T_k$ has strong properties under both classical and Bayesian views. These properties may be used to state desired properties of the maximum likelihood estimator introduced for the first time in Leite, Oishi & Pereira (1987). Under the Bayes' view the MLE is the posterior mode under the improper uniform prior. However the Bayesian material presented in this paper would be appropriate if we consider a large but finite support to the uniform prior and making it a proper probability distribution. Hence, both $[P_n]$ and $[\Pi]$ play important roles.

This paper is focused only on the investigation of Bayes' estimation properties. It is not our objective to examine stopping rules. However, it is clear that, besides cost, a good stopping rule must depend on the difference between the number of units selected up to a certain stage $j$, $S_j$, and the number of distinct units among those, $T_j$. A large difference (correspondingly, a small risk) could be substantial evidence that almost all members of the population have been selected. Should one continue sampling in such a situation?


## Acknowledgement

We are indebted to Professor Shelly Zacks who introduced us to the problem and gave many valuable comments. We are also grateful to the referees for useful suggestions that improved the format of this final version. Rio Scientific Center from IBM Brazil, partially supported this work.


## Appendix 1: Proof of (3.3)

To show the inequality we let $M_k = M$ and $S_k = S$, and rewrite the likelihood kernel as follows:

$$\begin{aligned}\lambda(n, k, t) &= \left\{\prod_{j=1}^{k} m_j!\right\}\left\{\prod_{j=0}^{t-1} (n-j)\right\}\left\{\prod_{j=1}^{k}\prod_{i=0}^{m_j-1} (n-i)\right\}^{-1} \\ &\leqslant \left\{\prod_{j=1}^{k} m_j!\right\}\left\{\prod_{j=0}^{t-1} (n-j)\right\}\left\{\prod_{j=1}^{k} (n-M+1)^{m_j}\right\}^{-1} \\ &= \left\{\prod_{j=1}^{k} m_j!\right\}\left\{\prod_{j=0}^{t-1} (n-j)\right\}(n-M+1)^{-S} \\ &= \left\{\prod_{j=1}^{k} m_j!\right\}\left\{\prod_{j=0}^{t-1} \left(1-\frac{j}{n}\right)\right\}\left\{\left(1-\frac{M-1}{n}\right)^{t}(n-M+1)^{S-t}\right\}^{-1} \\ &\leqslant \left\{\prod_{j=1}^{k} m_j!\right\}\left\{1-\frac{M-1}{n}\right\}^{-t} \leqslant \left\{\prod_{j=1}^{k} m_j!\right\}\left\{1-\frac{M-1}{t}\right\}^{-t}.\end{aligned}$$

## Appendix 2: Proof of Theorem 3

For all $k \geqslant s$, we can write

$$\beta(k, t) = \left\{a'\frac{(\tau-t)!}{\pi(\tau)\tau!} + \tau\right\} \Big/ \left\{a\frac{(\tau-t)!}{\pi(\tau)\tau!} + 1\right\},$$

where

$$a = \sum_{n=\tau+1}^{\infty} \frac{(n!)\pi(n)}{(n-t)!} \prod_{j=1}^{k} \left\{\binom{\tau}{m_j} \div \binom{n}{m_j}\right\},$$

$$a' = \sum_{n=\tau+1}^{\infty} \frac{n(n!)\pi(n)}{(n-t)!} \prod_{j=1}^{k} \left\{\binom{\tau}{m_j} \div \binom{n}{m_j}\right\}.$$

Then it is enough to show that

$$\lim_{k\to\infty} a = \lim_{k\to\infty} a' = 0.$$

For all $k \geqslant t+1$,

$$\prod_{j=1}^{k} \left\{ \binom{\tau}{m_j} \div \binom{n}{m_j} \right\} \leqslant (\tau/n)^k.$$

Consequently,

$$a \leqslant \tau^t \sum_{n=\tau+1}^{\infty} (\tau/n)^{k-t} \frac{(n!)\pi(n)}{(n)^t(n-t)!} \leqslant \tau^t \sum_{n=\tau+1}^{\infty} \pi(n)(\tau/n)^{k-t}$$
$$\leqslant \tau^t[\tau/(\tau+1)]^{k-t} \sum_{n=t+1}^{\infty} \pi(n) < (\tau+1)^t[\tau/(\tau+1)]^k.$$

This last term converges to zero as $k$ increases to infinity. Similarly, we would prove that $a'$ converges to zero as $k$ increases to infinity and the proof is completed.

**Appendix 3: Proof of Lemma 1**

Only item (ii) remains to be proved. Note that, for each fixed $n$ such that $\pi(n) > 0$, there exists a positive integer $k_1$ (depending on $n$ and on the sequence $\{m_j\}_{j\geqslant 1}$), such that $n \leqslant S_k$ for every $k \geqslant k_1$. Then, for all $k \geqslant k_1$

$$P_n\left(\left\{\omega \in \Omega; \sum_{i=1}^{k} \omega_i = n\right\}\right) = n! \sum_{i=0}^{n} \left\{(-1)^{n-i}[i!\,(n-i)!]^{-1} \prod_{j=1}^{k} \left[\binom{i}{m_j} \div \binom{n}{m_j}\right]\right\}$$
$$= 1 + n! \sum_{i=0}^{n-1} \left\{(-1)^{n-i}[i!\,(n-i)!]^{-1} \prod_{j=1}^{k} \left[\binom{i}{m_j} \div \binom{n}{m_j}\right]\right\}.$$

The second term of the right-hand side of this expression converges to zero as $k \to \infty$ since, for $0 \leqslant i < n$,

$$0 \leqslant \prod_{j=1}^{k} \left\{\binom{i}{m_j} \div \binom{n}{m_j}\right\} \leqslant (i/n)^k \to 0,$$

as $k \to \infty$. Consequently,

$$\lim_{k\to\infty} P_n\left(\left\{\omega \in \Omega; \sum_{i=1}^{k} \omega_i = n\right\}\right) = 1.$$

**Appendix 4: Proof of Theorem 6**

Using Theorem 4.3 of Doob (1953, p. 331), we have that

$$\lim_{k\to\infty} \beta_k = E\{N \mid \mathscr{F}_\infty\} \text{ a.s. } [\Pi],$$

where $\mathscr{F}_\infty$ is the smallest $\sigma$-algebra containing

$$\bigcup_{k>1} \mathscr{F}_k.$$

Since $T_k$ is $\mathscr{F}_\infty$-measurable (because it is $\mathscr{F}_k$-measurable and $\mathscr{F}_k \subset \mathscr{F}_\infty$) and $T_k \to N$ a.s. $[\Pi]$, we have that $\lim_k \sup T_k$ is $\mathscr{F}_\infty$-measurable and

$$\limsup_k T_k = N \text{ a.s. } [\Pi].$$

Then

$$E\{N \mid \mathscr{F}_\infty\} = N \text{ a.s. } [\Pi],$$

concluding the proof of item $(x)$.

To prove item (xi) we use Theorem 9.4.4 of Chung (1974, p. 334) to conclude that $\{\rho_k\}_{k\geqslant 1}$ converges a.s. [Π] to a random variable since

$$E\{\rho_k\} \leqslant \text{Var}\{N\} < \infty.$$

From item (x) we know that $\beta_k^2 \rightarrow N^2$ a.s. [Π].
To conclude the proof we need to show that

$$E\{N^2 \mid \mathscr{F}_k\} \rightarrow N^2 \text{ a.s. } [\Pi].$$

Since $E\{N^2\} < \infty$, from Theorem 4.3 of Doob (1953, p. 331) we have that

$$\lim_{k\rightarrow\infty} E\{N^2 \mid \mathscr{F}_k\} = E\{N^2 \mid \mathscr{F}_\infty\} \text{ a.s. } [\Pi].$$

On the other hand, since $T_k^2$ is $\mathscr{F}_\infty$-measurable and converges a.s. [Π] to $N^2$; we conclude that $\lim_k \sup T_k^2$ is also $\mathscr{F}_\infty$-measurable and is equal to $N^2$ a.s. [Π]. Hence,

$$E\{N^2 \mid \mathscr{F}_\infty\} = N^2 \text{ a.s. } [\Pi],$$

concluding the proof. □

## References


Chung, K.L. (1974). *A Course in Probability Theory*, 2nd ed. New York: Academic Press.
Doob, J.L. (1953). *Stochastic Process*. New York: Wiley.
Freeman, P.R. (1972). Sequential estimation of the size of a population. *Biometrika* **59**, 9–17.
Lehmann, E.L. (1966). Some concepts of dependence. *Ann. Math. Statist.* **37**, 1137–1153.
Leite, J.G. (1986). Exact estimates of the size of a finite and closed population (in Portugese). Doctoral dissertation, University of São Paulo.
Leite, J.G., Oishi, J. & Pereira, C.A. de B. (1987). Exact ML estimate of the size of a finite and closed population: capture/recapture sequential sample data. *Prob. Eng. Info. Sci.* **1**, 225–236.
Leite, J.G., Oishi, J. & Pereira, C.A. de B. (1988). A note on the exact maximum likelihood estimation of the size of a finite and closed population. *Biometrika* **75**, 178–180.
Leite, J.G. & Pereira, C.A. de B. (1987). An urn model for the capture/recapture sequential sampling process. *Sequential Anal.* **6**, 179–186.
Zacks, S. (1981). *Parametric Statistical Inference: Basic Theory and Modern Approaches*. Oxford: Pergamon.
Zacks, S. (1984). Bayes sequential estimation of the size of a finite population. University of São Paulo, Brazil. RT-MAE-8404.


## Résumé


On introduit, pour l'effectif d'une population, un estimateur de Bayes calculable sur des données obtenues par le processus d'échantillonnage progressif capture-recapture. On étudie des propriétés relatives à l'information contenue ces données. Quelques propriétés des grands échantillons sont aussi obtenues par l'emploi de résultats standards pour les martingales. Les résultats les plus forts sont la convergence presque sûre de l'estimateur de Bayes vers l'effectif réel de la population et la convergence du risque de Bayes vers zero. Les propriétés de Bayes présentées sont valables pour des probabilités a priori qui sont des vraies probabilités avec des moments d'ordre second finis. On démontre que l'estimateur du maximum de vraisemblance converge aussi presque sûrement vers l'effectif réel de la population.


[*Received July* 1989, *accepted April* 1990]