Journal of Statistical Planning and Inference 25 (1990) 363-380
North-Holland

BAYES SEQUENTIAL ESTIMATION OF THE SIZE OF A FINITE POPULATION

S. ZACKS

State University of New York at Binghamton, NY 13901, U.S.A.

C.A. de B. PEREIRA and J.G. LEITE

I.M.E., University of São Paulo, Brazil

Received October 1987
Recommended by K. Hinkelmann

Abstract: Bayes estimators for estimating the size of a finite closed population are studied for the capture-recapture sequential sampling. In particular we investigate the properties of the posterior mean and variance. Computing formula is developed to overcome some numerical difficulty. Sequential stopping rules are discussed. The sampling distribution of stopping variables and of associated estimators is determined recursively, and the operating characteristics of some procedures are studied.

Key words: Population size; Bayes sequential estimation; stopping rules; distribution of stopping times.

1. Introduction

The problem of estimating the size of a finite population is of interest in many areas of ecology, wild life management and other areas. The reader is referred to the book of Seber (1985) for a discussion of the various methods developed during the years. In Section 8 we provide a list of selected references. Our objective is to study Bayesian sequential estimation, including the question of stopping rules. There are only a few papers in the literature dealing with the Bayesian estimation and design of the sampling. We should mention however the studies of Freeman (1972) and Castledine (1981).

The present study deals with the estimation of the size of a finite closed population, by sequential capture-recapture sampling. We consider a general model, in which the size of the samples drawn (captured) in each stage is not necessarily predetermined, but could also be random. It is assumed that the samples can be considered as simple random without replacement. After the first sample is drawn all its elements are tagged and returned to the population. A second random sample is later drawn again. The number of untagged elements in the new sample is recorded, all elements are tagged and returned. In this manner we continue until the new information obtained on the population size is small compared to the cost of sampling and the information already gathered.

In Section 2 we provide a general formulation of the sampling model, the random variables and the likelihood function. In contrast to many papers, which assume a one-by-one capture–recapture, we treat the more general case, of samples of size M_k, $k=1,2,\ldots$ In Section 3 we consider the Bayesian framework and the extent to which the prior distribution is compatible with the observed likelihood function of the population size, N.

In Section 4 we study the properties of Bayes estimators, and their associated Bayes risk, for the squared-error loss. Several theorems are proven concerning monotonicity properties of the Bayes estimators and their limiting behavior. In particular we prove that these estimators converge almost surely, as the number of samples k grows, to the true population size N. As illustrated in Section 5, there are difficult numerical problems associated with the computation of the Bayes estimators and their posterior risk. These numerical difficulties were mentioned also by Freeman (1972). We show how these numerical difficulties can be overcome analytically. Applying the theory of formal power-series (Henrici, 1974), we derive an alternative computing formula for the case of Poisson prior distributions, which yields accurated results, even for very large samples.

The question of determining sequential stopping rules is very important. Samuel (1968) studied this problem in the context of maximum likelihood estimation (classical approach). Freeman (1972) performed dynamic programming to determine the stopping boundaries for Bayes truncated stopping rules. In Section 6 the problem of determining the Bayes sequential stopping rule is discussed. We consider the class of stopping rules, which require stopping as soon as the posterior risk drops below a specified level.

Section 7 is devoted to the development of the sampling distributions of the stopping variables, under a given population size N. Evaluating these sampling distributions we can compare the expected value of the Bayes estimator at stopping, the mean-absolute-deviation of the Bayes estimator from N at stopping, the expected total number of observations required and other characteristics of interest. Such comparisons can also evaluate different stopping rules. It is shown that if we use Poisson priors with small prior mean, λ, we can obtain excellent estimates, by stopping as soon as the posterior risk is smaller than 0.1. The effect of the sample size, M, on these results is only mild.

2. The sampling process and the statistical model

Consider a population of size N. It is assume that, during the duration of the study, the population is closed, i.e., there is no migration, births or deaths.

Samples of size M (possibly random) are drawn from the population at random and without replacement. After a random sample is drawn, all the population elements in the sample wich have not been previously observed are tagged. The sample is then returned to the population and another random sample is drawn. This is the process of 'capture–recapture'. Rules according to which sampling is terminated will be discussed later.

Let M_k $(k=1,2,\ldots)$ be th size of the k-the sample. Let U_k $(k=1,2,\ldots)$ denote the number of untagged elements in the k-th sample, and let $T_k=\sum_{j=1}^{k} U_j$. Obviously, $U_1=M_1$, $0\le U_k\le M_k$ for all $k\ge 2$; and the population size N is greater or equal to $M^*=\max_{1\le i}\{M_i\}$. Notice that the statistic T_k designates the total number of distinct elements of the population, which have been observed in the first k samples. The population size N is always greater or equal to T_k.

Finally, we assume that different samples are conditionally independent given $M_1, M_2, \ldots$, and that M_k are independent of N. Let $\boldsymbol{D}_k=(U_1,\ldots,U_k)$. Consider the points $\boldsymbol{d}_k=(u_1,\ldots,u_k)$ and $\boldsymbol{M}_k=(M_1,\ldots,M_k)$, where $u_j\in\{0,1,\ldots,M_j\}$, $j=1,\ldots,k$. The likelihood function of N, given $\{\boldsymbol{D}_k=\boldsymbol{d}_k\}$ and $\boldsymbol{M}_k$, is

$$L(N;\,\boldsymbol{d}_k,\boldsymbol{M}_k)=\prod_{j=1}^{k}\frac{\binom{N-t_{j-1}}{u_j}\binom{t_{j-1}}{M_j-u_j}}{\binom{N}{M_j}}$$

$$=\frac{\prod_{j=2}^{k}\binom{t_{j-1}}{M_j-u_j}}{\prod_{j=1}^{k}u_j!}\,I_{A_k}(t_k)K(N;\,k,\boldsymbol{M}_k,t_k), \tag{2.1}$$

where, $t_j=\sum_{i=0}^{j}u_i$, $j=0,1,2,\ldots,k$, $u_0\equiv 0$,

$$K(N;\,k,\boldsymbol{M}_k,t_k)=\begin{cases}\dfrac{N!}{(N-t_k)!\,\prod_{j=1}^{k}\binom{N}{M_j}}, & N\ge t_k,\\ 0, & N<t_k,\end{cases} \tag{2.2}$$

and $I_{A_k}(x_k)$ is the indicator function of the set

$$A_k=\left\{x;\ x=0,1,\ldots,\max_{1\le j\le k}\{M_j\}\le x\le\min\left(N,\sum_{j=1}^{k}M_j\right)\right\}.$$

$K(N;\,k,\boldsymbol{M}_k,t_k)$ is the likelihood kernel, and as can be immediately verified, the minimal sufficient statistic for N, after k samples, is T_k. As shown by Leite and Pereira (1987), the p.d.f. of T_k is

$$P\{T_k=t\mid N=n,\boldsymbol{M}_k\}=K(n;\,k,\boldsymbol{M}_k,t)I_A(t)\sum_{i=0}^{t}(-1)^{t-i}\frac{\prod_{j=1}^{k}\binom{i}{M_j}}{i!\,(t-i)!}. \tag{2.3}$$

3. The Bayesian framework

Let $\pi(n)$ denote a prior probability function defined on the set $\mathbb{N}^* = \{0, 1, 2, \ldots\}$ of nonnegative integers. The posterior probability function, after k samples, given $(\boldsymbol{M}_k, \boldsymbol{U}_k)$ depends only on the minimal sufficient statistics $(\boldsymbol{M}_k, T_k)$, and can be written in terms of the likelihood-kernel (2.2), for $T_k = t$, as

$$\pi(n \mid k, \boldsymbol{M}_k, t) = \pi(n) C(k, \boldsymbol{M}_k, t) K(n;\, k, \boldsymbol{M}_k, t) I_t(n), \tag{3.1}$$

where $I_t(n)$ is the indicator function of the set $\mathbb{N}_t = \{t, t+1, \ldots\}$ and

$$C(k, \boldsymbol{M}_k, t) = \left[\sum_{n=t}^{\infty} \pi(n) K(n;\, k, \boldsymbol{M}_k, t)\right]^{-1}. \tag{3.2}$$

Notice that $1/C(k, \boldsymbol{M}_k, t)$ assumes value zero if $\pi(n) = 0$ for all $n \geq t$. Thus, the posterior probability function is well defined only if the prior probability function, $\pi(n)$, is consistent with the data; i.e., only if $\pi(n)$ assumes positive values for some values of n in $\mathbb{N}_t$. Thus, if we define the set $\mathbb{N}_t^{\pi} = \{n;\ n \geq t \text{ and } \pi(n) > 0\}$ then we should require that $\mathbb{N}_t^{\pi}$ will not be empty, for every observable value t of T_k. This condition is trivially satisfied by every prior probability function which is supported by $\mathbb{N}^*$. This creates some logical difficulty, since N is a priori finite. However, if an upperbound for N is not available, one would be on a safer ground by applying a p.d.f. whose support is $\mathbb{N}^*$. Furthermore, we restrict attention in the present paper only to functions $\pi(n)$ yielding a finite value of $C(k, \boldsymbol{M}_k, t)$ for all $0 \leq t \leq k$, all $\boldsymbol{M}_k$; and all $k = 1, 2, \ldots$. The improper probability measure on $\mathbb{N}^*$, $\pi(n) = c$, with $M_j = 1$ for all $j = 1, \ldots, k$, yields $C^{-1}(k, \boldsymbol{M}_k, t) = \sum_{n=t}^{\infty} n!/((n-t)!\, n^k)$, which is finite only if $t \leq k-2$. Thus, such an improper prior should not be considered.

We show now that, if $t \geq M_k^* = \max_{1 \leq j \leq k} \{M_j\}$, then $K(n;\, k, \boldsymbol{M}_k, t)$ is bounded above by a function independent of n. Notice that, since $T_1 = M_1$, the above requirement is satisfied whenever all M_j assume the same value M. Furthermore, $T_k \leq \sum_{j=1}^{k} M_j$. Thus,

$$\begin{aligned}
K(n;\, k, \boldsymbol{M}_k, t) &= \frac{(\prod_{j=1}^{k} M_j!) \prod_{j=1}^{t} (n-t+j)}{\prod_{j=1}^{k} \prod_{i=1}^{M_j} (n-i+1)} \\
&\leq \frac{(\prod_{j=1}^{k} M_j!) \prod_{i=1}^{t} (n-t+j)}{(n - M_k^* + 1)^{\sum_{j=1}^{k} M_j}} \\
&\leq \frac{(\prod_{j=1}^{k} M_j!)}{(1 - (M_k^* - 1)/n)^t (n - M_k^* + 1)^{\sum_{j=1}^{k} M_j - t}} \\
&\leq \left(\sum_{j=1}^{k} M_j!\right) \Big/ \left(1 - \frac{M_k^* - 1}{t}\right)^t.
\end{aligned} \tag{3.3}$$

Thus, if $\sum_{n=0}^{\infty} \pi(n) < \infty$, and $M_k^* \le t \le \sum_{j=1}^{k} M_j$, then

$$C^{-1}(k, M_k, t) = \sum_{n=t}^{\infty} \pi(n) K(n;\, k, M_k, t) < \infty,$$

and the posterior distribution is well defined.

4. Bayes estimators for squared-error loss

The Bayes estimator for a squared-error loss is the mean of the posterior distribution. The corresponding posterior Bayes risk is the variance of the posterior distribution. Accordingly, if the prior p.d.f. is a proper prior distribution, having a finite second moment, and if $\mathbb{N}_t^{\pi} \ne \emptyset$ and $M_k^* \le t \le \sum_{j=1}^{k} M_j$, the Bayes estimator of N, given $T_k = t$, exists and is given by the expected value of the posterior distribution $\pi(n \mid k, M_k, t)$, i.e.,

$$B(k,t) = \sum_{n=t}^{\infty} n\pi(n \mid k, M_k, t). \tag{4.1}$$

A prior probability function $\pi(n)$ is called degenerate at n^* if $\pi(n) = I\{n = n^*\}$. Formula (4.1) does not hold if π is degenerate at $n^* < t$. However, if $\pi(n) = I\{n = n^*\}$ for some $n^* \in \mathbb{N}_t^*$ then $B(k,t) = n^*$ for all (k,t).

Let $R(k,t)$ denote the posterior variance of N, given $T_k = t$. In the present section we present some general results concerning the functions $B(k,t)$ and $R(k,t)$. Computational formulae and examples, for the case of Poisson prior distributions, will be given in the next section.

Theorem 1. *For all $k \ge 2$, assume that $M_{k+1}^* \le t \le S_k$ and $\mathbb{N}_t^{\pi} \ne \emptyset$, where $M_{k+1}^* = \max_{1 \le j \le k+1} M_j$, $S_k = \sum_{j=1}^{k} M_j$. If $M_{k+1} \le S_k$ then*

$$B(k,t) \ge B(k+1,t). \tag{4.2}$$

Equality holds in (4.2) *if $\pi(n)$ is degenerate.*

Proof. Under the above conditions, define over $\mathbb{N}^*$ the function

$$h(n) = I_t(n) \Big/ \binom{n}{M_{k+1}}.$$

$h(n)$ is a decreasing function of n over $\mathbb{N}_t^{\pi}$. Thus

$$E_{\pi}\{h(N) \mid T_k = t\}\, E_{\pi}\{N \mid T_k = t\} \ge E_{\pi}\{h(N)N \mid T_k = t\}. \tag{4.3}$$

It is straightforward to check that

$$E_{\pi}\{h(N) \mid T_k = t\} = \frac{C(k, M_k, t)}{C(k+1, M_{k+1}, t)}, \tag{4.4}$$

and

$$E_\pi\{h(N)N \mid T_k = t\} = \frac{C(k, M_k, t)}{C(k+1, M_{k+1}, t)} B(k+1, t). \tag{4.5}$$

From (4.3)-(4.5) we obtain

$$\frac{C(k, M_k, t)}{C(k+1, M_{k+1}, t)} B(k, t) \geq \frac{C(k, M_k, t)}{C(k+1, M_{k+1}, t)} B(k+1, t). \tag{4.6}$$

Under the conditions on t and π, $0 < C(k, M_k, t)/C(k+1, M_{k+1}, t) < \infty$ and hence $B(k,t) \geq B(k+1, t)$. If $\pi(n)$ is degenerate at a value $n^* \in \mathbb{N}_t^\pi$ then, obviously, $B(k,t) = n^*$ for all (k,t). □

The next theorem establishes that, under general conditions, $B(k,t)$ is non-decreasing in t, for a fixed k.

Theorem 2. *For all $k \geq 2$, let $M_k^* \leq t \leq S_k - 1$ and let $\mathbb{N}_{t+1}^\pi \neq \emptyset$. Then*

$$B(k,t) \leq B(k, t+1). \tag{4.7}$$

Proof. Notice that $V_\pi\{N \mid T_k = t\} = \text{cov}_\pi\,(N, N-t \mid T_k = t)$. Thus,

$$R(k,t) = E_\pi\{N(N-t) \mid T_k = t\} - B(k,t) E_\pi\{N - t \mid T_k = t\}. \tag{4.8}$$

But,

$$E_\pi\{N(N-t) \mid T_k = t\} = \frac{C(k, M_k, t)}{C(k, M_k, t+1)} B(k, t+1), \tag{4.9}$$

and

$$E_\pi\{N - t \mid T_k = t\} = \frac{C(k, M_k, t)}{C(k, M_k, t+1)}. \tag{4.10}$$

From (4.8)-(4.10) we obtain that,

$$R(k,t) = \frac{C(k, M_k, t)}{C(k, M_k, t+1)} [B(k, t+1) - B(k, t)]. \tag{4.11}$$

Finally, since $R(k,t) \geq 0$ for all k, t satisfying the conditions of the theorem, and since $0 < C(k, M_k, t)/C(k, M_k, t+1) < \infty$, inequality (4.7) holds. Equality at (4.7) obviously holds if $\pi(n) = I(n = n^*)$ where $n^* \geq t + 1$. □

What happens when $k \to \infty$? Let $\{M_j\}_{j=1}^\infty$ be a sequence of elements of $\mathbb{N}^*$ and let $M^* = \max_{1 \leq j} \{M_j\}$. For each positive integer t such that $\mathbb{N}_t^\pi \neq \emptyset$ and $t \geq M^*$, define

$$k_t = \min\{j;\ j \geq \mathbb{N}^*, j \geq 2 \text{ and } S_j \geq t\}. \tag{4.12}$$

For all $k \geq k_t$, $M_k^* \leq t \leq S_k$ and the conditions of Theorem 1 hold. Thus, we obtain:

Theorem 3. *For a fixed $t \in \mathbb{N}^*$ such that, $t \geq M^*$ and $\mathbb{N}_t^\pi \neq \emptyset$, if $\pi(n)$ has a finite mean, then*

$$\lim_{k \to \infty} B(k,t) = \tau_t, \tag{4.13}$$

where $\tau_t = \min\{\mathbb{N}_t^\pi\}$.

Proof. For all $k \geq k_t$, the Bayes estimator $B(k,t)$ can be written as

$$B(k,t) = \frac{\tau_t + A^*(k,t)}{1 + A(k,t)}, \tag{4.14}$$

where

$$A(k,t) = \frac{(\tau_t - t)!}{\pi(\tau_t)\tau_t!} \sum_{n=\tau_t+1}^{\infty} \frac{n!\,\pi(n)}{(n-t)!} \prod_{j=1}^{k} \frac{\binom{\tau_t}{m_j}}{\binom{n}{m_j}}, \tag{4.15}$$

and

$$A^*(k,t) = \frac{(\tau_t - t)!}{\pi(\tau_t)\tau_t!} \sum_{n=\tau_t+1}^{\infty} \frac{n(n!)\,\pi(n)}{(n-t)!} \prod_{j=1}^{k} \frac{\binom{\tau_t}{m_j}}{\binom{n}{m_j}}. \tag{4.16}$$

But, for all $n \geq \tau_k$,

$$\prod_{j=1}^{k} \frac{\binom{\tau_j}{m_j}}{\binom{n}{m_j}} \leq \left(\frac{\tau_t}{n}\right)^k. \tag{4.17}$$

Hence,

$$\begin{aligned}
\pi(\tau_t)A(k,t) &\leq \tau_t^t \sum_{n=\tau_t+1}^{\infty} \left(\frac{\tau_t}{n}\right)^{k-t} \frac{n!\,\pi(n)}{n^t(n-t)!} \\
&\leq \tau_t^t \sum_{n=\tau_t+1}^{\infty} \left(\frac{\tau_t}{n}\right)^{k-t} \pi(n) \\
&\leq \tau_t^t \left(\frac{\tau_t}{1+\tau_t}\right)^{k-t} \sum_{n=\tau_t+1}^{\infty} \pi(n) \\
&\leq \tau_t^t \left(\frac{\tau_t}{1+\tau_t}\right)^{k-t} \to 0 \quad \text{as } k \to \infty.
\end{aligned} \tag{4.18}$$

Similarly, since $E_\pi\{N\} < \infty$,

$$\pi(\tau_t)A^*(k,t) \leq \tau_t^t E_\pi\{N\} \left(\frac{\tau_t}{1+\tau_t}\right)^{k-t} \to 0 \quad \text{as } k \to \infty. \tag{4.19}$$

Thus,

$$\lim_{k\to\infty} A(k,t) = \lim_{k\to\infty} A^*(k,t) = 0. \tag{4.20}$$

(4.14) and (4.20) yield (4.13). □

We remark that if π has positive prior probabilities for all $n \in \mathbb{N}^*$ then $\lim_{k\to\infty} B(k,t) = t$ for each fixed t.

We are interested in the almost-sure limit of $B(k, T_k)$. To simplify notation, we assume that $M_j = M$ for *all* $j = 1, 2, \ldots$; i.e., all samples are of equal size. T_1 assumes the value M w.p. 1 and T_k, $k \geq 2$ satisfy $M \leq T_k \leq kM$. Let $\Omega = \{\omega = (\omega_1, \omega_2, \ldots);\ \omega_j \in \{M, M+1, \ldots, jM\},\ j = 1, 2, \ldots\}$; let $\mathscr{B}_k$ denote the Borel σ-field generated by the random variables $(T_1, \ldots, T_k)$ and let $\mathscr{F}_\infty$ denote the smallest Borel σ-field containing $\bigcup_{j=1} \mathscr{B}_j$. For each $k \geq 1$, we consider on $\mathscr{B}_k$ the probability measure $\mathrm{P}_N^{(k)}(\cdot)$ induced by (2.1). In particular, $\mathrm{P}_N^{(k)}\{T_k = t\}$ is given by (2.3). We define on $\mathscr{F}_\infty$ the probability measure $\mathrm{P}_N(\cdot)$ such that, for each $B \in \mathscr{B}_k$, $\mathrm{P}_N\{B\} \equiv \mathrm{P}_N^{(k)}\{B\}$. The triplet $(\Omega, \mathscr{F}_\infty, \{\mathrm{P}_N\colon N \in \mathbb{N}^*\})$ is the statistical space.

For a given $N \in \mathbb{N}^*$, the sequence T_k is a.s. bounded by N. Moreover $T_k \leq T_{k+1}$. Thus, $\lim_{k\to\infty} T_k = \tilde{N}$ a.s. Moreover, from (2.3),

$$\mathrm{P}_N^{(k)}\{T_k = N\} = 1 + N! \sum_{i=0}^{N-1} (-1)^{N-i}(i!(N-i)!)^{-1} \left[\frac{\binom{i}{M}}{\binom{N}{M}}\right]^k. \tag{4.21}$$

As we have shown before,

$$0 \leq \left[\frac{\binom{i}{M}}{\binom{N}{M}}\right]^k \leq \left(\frac{i}{N}\right)^k \to 0 \quad \text{as } k \to \infty.$$

Hence,

$$\lim_{k\to\infty} \mathrm{P}_N^{(k)}\{T_k = N\} = 1.$$

Thus, $T_k \to N$ a.s. $[\mathrm{P}_N]$. Combining this with the result of Theorem 3 we obtain:

Corollary 1.

$$\lim_{k\to\infty} B(k, T_k) = N \quad \text{a.s. } [\mathrm{P}_N]. \tag{4.22}$$

From (4.10) and (4.11) we can write, for all $k \geq k_t$,

$$R(k,t) = [B(k,t) - t][B(k,t+1) - B(k,t)]. \tag{4.23}$$

Hence,

$$R(k, T_k) = [B(k, T_k) - T_k][B(k, T_k + 1) - B(k, T_k)]. \tag{4.24}$$

Since $T_k \to N$ a.s. $[\mathrm{P}_N]$ as $k \to \infty$ we obtain:

Corollary 2.

$$\lim_{k\to\infty} R(k,T_k)=0 \quad \text{a.s. } [\mathrm{P}_N]. \tag{4.25}$$

In the Bayesian framework we extend the above asymptotic results by defining the prior probability space $(\mathbb{N}_M^*, \mathscr{G}, \pi)$, where $\mathbb{N}_M^*=\{n\colon n\in\mathbb{N}^* \text{ and } n\ge M\}$ and $\mathscr{G}$ is the σ-field generated by all the subsets of $\mathbb{N}_M^*$. We consider the Cartesian products $\Omega^*=\Omega\times\mathbb{N}_M^*$; $\mathscr{F}^*=\mathscr{F}_\infty\times\mathscr{G}$ and the product measure $\mathscr{P}_\pi$ such that, for every $F\times G\in\mathscr{F}^*$,

$$P_\pi(F\times G)=\sum_{n\in G} P_n(F)\pi(n).$$

We define $B_k=E_\pi\{N\mid \mathscr{B}_k^*\}$ and $R_k=E_\pi\{(N-B_k)^2\mid\mathscr{B}_k\}$. Then we obtain:
(i) $\{B_k,\mathscr{B}_k\}_{k=1}^\infty$ is a martingale;
(ii) $\{R_k,\mathscr{B}_k\}_{k=1}^\infty$ is a super martingale;
(iii) $B_k\to N$ a.s. $[\pi]$;
(iv) $R_k\to 0$ a.s. $[\pi]$.
The technical details are not of much interest and will omitted. The asymptotic results obtained above will guide us in the investigation of the properties of sequential stopping rules, which will be discussed in the following sections.

5. Computing the Bayes estimator and the Bayes risk for Poisson priors

Let $\pi_\lambda(n)$ be the Poisson p.d.f. with mean λ; $0<\lambda<\infty$. In the present section we develop computing formula for $B(k,t)$, under $\pi_\lambda(n)$. $R(k,t)$ is computed according to (4.24). Since $\pi_\lambda(n)>0$ for all $n\in\mathbb{N}^*$, $\mathbb{N}_t^{\pi_\lambda}\neq\emptyset$ for all $t>M$. We focus attention on the case of $M_j=M$ for all $j\ge 1$. Under these assumptions $B(k,t)$ exists for all (k,T_k).

Substituting $\pi_\lambda(n)=\mathrm{e}^{-\lambda}\lambda^n/n!$ for the prior distribution we obtain

$$B(k,t)=t+\lambda\,\frac{\sum_{n=0}^\infty \lambda^n a_n(k,t+1,M)}{\sum_{n=0}^\infty \lambda^n a_n(k,t,M)}, \tag{5.1}$$

where

$$a_n(k,t,M)=\frac{1}{n!\,(\prod_{i=1}^M (n+t+1-i))^k},\quad n\ge 0.$$

Formula (5.1) is inappropriate for numerial computations, since the values of $a_n(k,t+1,M)$ decrease to zero very rapidly. For example, if $n=10$, $M=10$, $t=50$ and $k=20$ we obtain

$$a_{10}(20,50,10)=1\bigg/\left[3\,628\,800\prod_{i=1}^{10}(61-i)^{20}\right]=\exp(-818.1125).$$

This number is too small for actual computer evaluation. It is obvious that each term in the numerator and the denominator of the ratio of power-series in (5.1) should be inflated by sufficiently large function of k and t. In the present section we derive a formula which yields excellent numerical results. All the numerical results presented later were computed on a ZENITH PC, with TURBO PASCAL, according to this formula. For this purpose we apply a method of inverting the power-series, $\sum_{n=0}^{\infty} \lambda^n a_n$, with $a_0 \neq \emptyset$. This method is discussed in Henrici (1974, pp. 17).

First we invert $\sum_{n=0}^{\infty} \lambda^n a_n(k,t,M)$. Let

$$\left(\sum_{n=0}^{\infty} \lambda^n a_n(k,t,M)\right)^{-1} = \sum_{n=0}^{\infty} \lambda^n b_n(k,t,M), \tag{5.2}$$

where

$$b_0(k,t,M) = \frac{1}{a_0(k,t,M)} = \prod_{i=1}^{M} (t+1-i)^k, \tag{5.3}$$

and for $n \geq 1$,

$$b_n(k,t,M) = (-1)^n \left(\prod_{i=1}^{M} (t+1-i)^k\right) D_n(k,t,M), \tag{5.4}$$

where $D_n(k,t,M)$ is the Wronski determinant

$$D_n(k,t,M) = \begin{bmatrix} d_1(k,t,M) & \cdots & \cdots & d_n(k,t,M) \\ 1 & d_1(k,t,M) & \cdots & d_{n-1}(k,t,M) \\ & 1 & & \vdots \\ 0 & & \ddots & \vdots \\ & & 1 & d_1(k,t,M) \end{bmatrix}, \tag{5.5}$$

with

$$d_j(k,t,M) = \frac{1}{j!} \prod_{i=1}^{M} \left(\frac{t+1-i}{t+j+1-i}\right)^k, \quad j = 1, \ldots, n. \tag{5.6}$$

Notice that $D_1(k,t,M) = d_1(k,t,M)$. The determinant $D_n(k,t,M)$ can be computed by the following recursive formula:

$$D_0(k,t,M) = 1, \tag{5.7a}$$

$$D_n(k,t,M) = \sum_{j=0}^{n-1} (-1)^j d_{j+1}(k,t,M) D_{n-j-1}(k,t,M), \quad n \geq 1. \tag{5.7b}$$

Multiplying $\sum_{n=0}^{\infty} \lambda^n a_n(k,t+1,M)$ by $\sum_{n=0}^{\infty} \lambda^n b_n(k,t,M)$ we obtain the expression

$$B(k,t) = t + \lambda \sum_{n=0}^{\infty} \lambda^n B_n(k,t,M), \tag{5.8}$$

where

$$B_n(k,t,M) = \sum_{l=0}^{n} \frac{(-1)^l}{(n-l)!} \prod_{i=1}^{M} \left(\frac{t+1-i}{t+n-l+2-i}\right)^k D_l(k,t,M). \tag{5.9}$$

Substitution (5.9) in $\sum_{n=0}^{\infty} \lambda^n B_n(k,t,M)$ and changing the order of summation we obtain

$$\sum_{n=0}^{\infty} \lambda^n B_n(k,t,M)$$
$$= \sum_{l=0}^{\infty} (-1)^l \lambda^l D_l(k,t,M) \sum_{n=0}^{\infty} \lambda^n \frac{1}{n!} \prod_{j=1}^{M} \left(\frac{t+1-j}{n+t+2-j} \right)^k. \tag{5.10}$$

Moreover, according to (5.7),

$$\sum_{l=0}^{\infty} (-1)^l \lambda^l D_l(k,t,M) = 1 + \sum_{n=1}^{\infty} (-1)^n \lambda^n \sum_{j=1}^{n} (-1)^{j-1} d_j(k,t,M) D_{n-j}(k,t,M)$$
$$= 1 - \sum_{j=0}^{\infty} (-1)^j \lambda^j D_j(k,t,M) \sum_{l=1}^{\infty} \lambda^l d_l(k,t,M). \tag{5.11}$$

Thus,

$$\sum_{l=0}^{\infty} (-1)^l \lambda^l D_l(k,t,M) = \frac{1}{1+\sum_{j=1}^{\infty} \lambda^j d_j(k,t,M)}. \tag{5.12}$$

Substituting (5.12) in (5.10) we obtain the formula

$$B(k,t) = t + \lambda \frac{\sum_{n=0}^{\infty} \lambda^n \frac{1}{n!} \prod_{j=1}^{M} \left(\frac{t+1-j}{n+t+2-j} \right)^k}{\sum_{n=0}^{\infty} \lambda^n \frac{1}{n!} \prod_{j=1}^{M} \left(\frac{t+1-j}{n+t+1-j} \right)^k}. \tag{5.13}$$

Accurate computation of $B(k,t)$, and $R(k,t)$, according to (5.13) can be readily done. We sum the series in (5.13) up to $n = N(\lambda) = \text{Int}(\lambda + 5\sqrt{\lambda})$, where $\text{Int}(x)$ denotes the integer part of x.

In Table 1 we present values of $B(k,t)$ and $R(k,t)$, for Poisson priors with means $\lambda = 20$, 50, and with $M = 1$. The values of T_k were realized by simulation from a population of size $N = 50$.

Table 1
Values of $B(k,t)$ and $R(k,t)$ for Poisson priors

		$\lambda = 20$		$\lambda = 50$	
k	T_k	$B(k,T_k)$	$R(k,T_k)$	$B(k,T_k)$	$R(k,T_k)$
3	2	19.008	19.979	49.001	49.998
13	10	20.660	15.516	48.071	48.603
40	25	30.480	7.073	45.984	34.280
70	33	35.926	3.458	42.961	15.692
100	42	44.128	2.383	48.495	8.893
150	48	48.969	1.025	50.670	3.157
300	50	50.053	0.053	50.134	0.136

Table 2
Bayes estimates and posterior risk for $\lambda = 20$, $M = 420$

k	T_k	$B(k, T_k)$	$R(k, T_k)$
1	420	420.050	0.052
2	810	814.706	4.765
3	1125	1129.969	5.008
4	1233	1236.811	3.835
5	1542	1546.104	4.125
6	1648	1651.442	3.457
7	1749	1751.940	2.951

The numerical results presented in Table 1 show that, when M and k are small, $B(k,t)$ and $R(k,t)$ are sensitive to the assumed values of λ. But, as k grows T_k becomes the dominant factor, and the estimates are close even if the λ values differ considerably. For large values of k we see that

$$B(k, T_k) \approx T_k + R(k, T_k). \tag{5.14}$$

Formula (5.13) yields good numerical results even for large values M and T_k. In Table 2 we present the results, computed according to this formula when $M_k = 420$, and the T_k values are large. This might be the case, for example, when we sample fish from a pond, and each time we throw a net a large number of fish are caught.

Approximation (5.14) is illustrated also in Table 2, despite the small values of k. This is due to the large values of T_k.

6. Bayes sequential stopping rules

Let $g_{k+1}(u \mid k, M_k, t)$ denote the predictive p.d.f. of U_{k+1}, given $T_k = t$, i.e.,

$$g_{k+1}(u \mid k, M_k, t) = \sum_{n=t}^{\infty} \pi(n \mid k, M_k, t) \frac{\binom{n-t}{u}\binom{t}{M_{k+1}-u}}{\binom{n}{M_{k+1}}}. \tag{6.1}$$

Let $R(k,t)$ denote the posterior risk after k samples, given $T_k = t$. Let $\varrho(k,t)$ denote the risk associated with the optimal stopping policy. Let $c(M_k)$ denote the cost [$] of observing the k-th sample. Let γ [$] denote a cost factor, which represents the penalty for one unit of posterior risk. Then, $\varrho(k,t)$ satisfies the functional equation

$$\varrho(k,t) = \min\{\gamma R(k,t), c(M_{k+1}) + \sum_{u=0}^{M_{k+1}} g_{k+1}(u \mid k, M_k, t)\varrho(k+1, t+u)\}. \tag{6.2}$$

For given values of $\{c(M_k)\}_{k=1}^{\infty}$ and γ, it is optimal to stop after k samples, if k is the smallest non-negative integer for which $\gamma R(k,t)=\varrho(k,t)$ (see DeGroot (1970)).

An explicit solution for $\varrho(k,t)$ is not available. One can obtain numerical approximations by a sequence of functions $\{\varrho^{(K)}(k,t), k=1,\dots,K\}$, which are evaluated by the method of Dynamic programming. $\varrho^{(K)}(k,t)$ represent the risk associated with the optimal policy, which is truncated at K, i.e., no more than K additional samples can be drawn, after observing the frist k samples. By backward induction we define

$$\varrho^{(0)}(k+K,t) = \gamma R(k+K,t), \tag{6.3a}$$

$$\varrho^{(j)}(k+K-j,t) = \min\{\gamma R(k+K-j,t), c(M_{k+K-j+1}) + \sum_{u=0}^{M_{k+K-j+1}} g_{k+K-j+1}(u \mid k+K-j, M_{k+K-j}, t) \cdot \varrho^{(j-1)}(k+K-j+1, t+u)\} \tag{6.3b}$$

for $j=1,\dots,K$. The function $\varrho^{(0)}(k+K,t)$ has to be tabulated for all $M^*_{k+K}\le t\le S_{k+K}$ and $\varrho^{(j)}(k+K-j,t)$ has to be tabulated for all $M^*_{k+K-j}\le t\le S_{k+K-j}$, where $M^*_v=\max_{1\le k\le v}\{M_k\}$ and $S_v=\sum_{k=1}^{v} M_v$.

Freeman (1972) performed extensive numerical dynamic programming, in order to determine the optimal stopping boundaries for truncated sequential sampling, when $M_k=1$ and $\pi(n)\propto n^{-v}$, $v\ge 2$. The stopping boundaries obtained by Freeman

Table 3
The boundary $b^*_m(k)$ for the Poisson prior with $\lambda=50$, $m=5$, 10, 15, 20, $\varepsilon=0.1$

	m			
k	5	10	15	20
3	0	11	16	22
4	0	12	18	25
5	6	13	20	27
6	7	15	23	30
7	8	16	25	33
8	8	18	27	36
9	9	19	29	39
10	10	21	32	43
15	14	29	44	58
20	18	37	55	74
25	22	45	67	90
30	26	53	79	106
35	30	61	91	122
40	34	69	104	138
45	38	77	116	154
50	42	85	128	170

can be approximated by linear boundaries of the form $\tilde{T}_k = b_0 + b_1 k$. The stopping time is

$$\tilde{k} = \min\{k \in \mathbb{N}:\ T_k \le b_0 + b_1 k\}. \tag{6.4}$$

Since we are interested in stopping rules which allow for large numbers of samples, k, and since $R(k, T_k) \to 0$ a.s. $[\pi]$, we will consider the class of (suboptimal) stopping variables:

$$K_\varepsilon = \min\{k \ge 2:\ R(k, T_k) < \varepsilon\}, \tag{6.5}$$

for values of ε close to zero. In Section 7 we study the operating characteristics of this Bayes stopping rule. In Table 3 we provide, for each k the first value of $t \ge m$ for which $R(k,t) > 0.1$. These values have been computed for the case of Poisson prior with $\lambda = 50$. Notice that these values, to be designated by $b_m^*(k)$, are the boundary of the continuation region. As soon as T_k falls below $b_m(k)$ we stop sampling.

Plotting the values of $b_m(k)$ versus k we see that these stopping boundaries are approximately linear. In Figure 1 we present these boundaries for Poisson priors with $\lambda = 10$, $\lambda = 5$, $\varepsilon = 0.05$, 0.20 and various values of M. It is interesting to notice that the boundaries for k_ε are similar to those obtained by Freeman (1972) by the method of dynamic programming.

7. Sampling distribution of stopping times and the use of sequential estimators

Suppose that all samples are of size $M_k = m$. The formula developed below can be easily generalized for any sequence of M values. Let $\{b_m(k)\}_{k=1}^{\infty}$ be a sequence of integer values, such that $b_m(k) < km$ for all $k = 1, 2, \ldots$ and $b_m(k) \uparrow \infty$ monotonically. Define the corresponding stopping times

$$K(b_m) = \min\{k \in \mathbb{N}:\ T_k \le b_m(k)\}. \tag{7.1}$$

We have shown that $T_k \to N$ a.s. $[\mathrm{P}_N]$. Since N is a finite integer, $K(b_m) < \infty$ a.s. We develop now the probability distribution of $K(b_m)$, for a fixed value, N, of the population size.

Let

$$h_k^{(N,m)}(t) = \mathrm{P}_N\{T_k = t, K(b_m) \ge k\}. \tag{7.2}$$

Obviously, $h_1^{(N,m)}(t) = I\{t = m\}$. Since $T_k = T_{k-1} + U_k$, we can write, for $k \ge 2$,

$$h_k^{(N,m)}(t) = \sum_{u = b_m(k-1)+1}^{m(k-1)\wedge t} h_{k-1}^{(N,m)}(u)\,\mathrm{P}_N\{U_k = t-u \mid T_{k-1} = u\}, \tag{7.3}$$

where $m(k-1) \wedge t = \min\{t, m(k-1)\}$. Moreover,

$$\mathrm{P}_N\{U_k = t-u \mid T_{k-1} = u\} = \frac{\binom{N-u}{t-u}\binom{u}{m-t+u}}{\binom{N}{m}}, \tag{7.4}$$

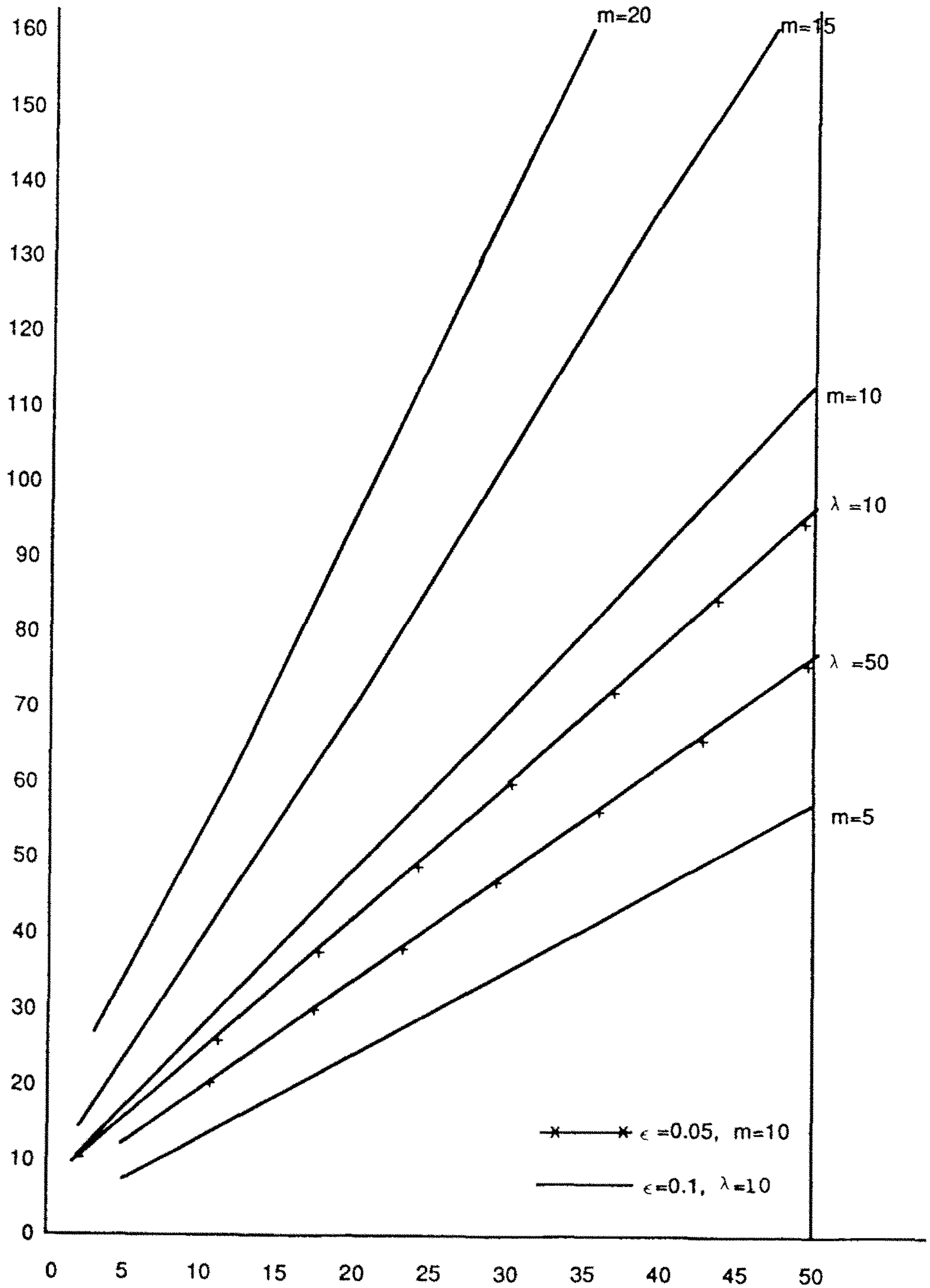

Fig. 1. Stopping boundaries for K_ε.

is the hyper-geometric p.d.f., $h(t-u;\ N, N-u, m)$, where

$$h(j;\ N, A, m) = \frac{\binom{A}{j}\binom{N-A}{m-j}}{\binom{N}{m}}, \quad j = 0, \ldots, m.$$

Since $h_1^{(N,m)}(t) = I\{t = m\}$ we obtain

$$h_2^{(N,m)}(t) = \begin{cases} 0, & \text{if } t < m \text{ or } t > 2m, \\ h(t-m;\ N, N-m, m) & \text{if } m \le t \le 2m. \end{cases} \tag{7.5}$$

Furthermore, for every $k \ge 3$,

$$h_k^{(N,m)}(t) = \sum_{u=b_m(k-1)+1}^{m(k-1)\wedge t} h_{k-1}^{(N,m)}(u)\, h(t-u;\ N, N-u, m). \tag{7.6}$$

The probability distribution of $K(b_m)$ can be determined by computing recursively $h_k^{(N,m)}(t)$ and evaluating

$$\mathrm{P}_{N,m}\{K(b_m) = k\} = \sum_{j=b_m(k-1)+1}^{b_m(k)} h_k^{(N,m)}(j). \tag{7.7}$$

If stopping occurs at $K(b_m) = \tilde{k}$, $T_{\tilde{k}}$ can assume a value in the set $\{b_m(\tilde{k}-1)+1, \ldots, b_m(\tilde{k})\}$. For these values of $T_{\tilde{k}}$ the Bayes estimator assumes the value $B(\tilde{k}, T_{\tilde{k}})$. The probability distribution function of $B(\tilde{k}, t)$, for $t \in \{b_m(\tilde{k}-1)+1, \ldots, b_m(\tilde{k})\}$ is $h_{\tilde{k}}^{(N,m)}(t)$. In Table 4 we present the p.d.f. of $B(\tilde{k}, T_{\tilde{k}})$ and of $K(b_m)$ for the case of a population with size $N = 50$, $m = 15$, $\varepsilon = 0.10$. The boundary $b_m(k)$ is that of Figure 1, which corresponds to the Poisson case with $\lambda = 10$; i.e., $b_m(k) = \mathrm{Int}(7.5 + 3.25k) - 1$.

Table 4
Probability distribution of $B(\tilde{k}, T_{\tilde{k}})$ and of $K(b_m)$ for $N = 50$, $m = 15$, $\varepsilon = 0.1$

k	T_k	$B(\tilde{k}, T_{\tilde{k}})$	$P\{B(\tilde{k}, T_{\tilde{k}})\}$	$P\{K(b_m) = k\}$
12	43	43.068	0.000002	
	44	44.078	0.000015	
	45	45.089	0.000248	
	46	46.100	0.002899	0.003164
13	47	47.077	0.007771	
	48	48.087	0.069500	
	49	49.098	0.313078	0.390349
14	50	50.077	0.606488	0.606488
sum	–	–	1.000000	1.000000

Table 5
Operating characteristics of K_ε, with $\varepsilon = 0.1$ and $N = 100$

λ	m	$b_m(k)$[a]	MEAN[b]	MAD[c]	ASN[d]
10	5	$1.0 + k$	99.041	1.0163	449.7
	10	$3.0 + 2k$	99.093	0.9749	442.7
	15	$6.5 + 3k$	99.152	0.9237	433.7
	20	$8.7 + 4k$	99.166	0.9093	419.9
50	5	$1.0 + 0.8k$	99.880	0.2528	618.0
	10	$4.0 + 1.6k$	99.912	0.2490	599.8
	15	$7.0 + 2.4k$	99.912	0.2473	584.7
	20	$10 \ + 3.2k$	99.892	0.2452	576.5

[a] Take the integer part.
[b] $\text{MEAN} = E^{(N,m)}\{B(\tilde{k}, T_{\tilde{k}})\}$.
[c] $\text{MAD} = E^{(N,m)}\{|B(\tilde{k}, T_{\tilde{k}}) - N|\}$.
[d] $\text{ASN} = mE^{(N,m)}\{K_\varepsilon\}$.

We see in Table 4 that over 99% of the stopping time are realized with $T_{\tilde{k}} \geq 49$, after 13 or 14 samples. The expected value of the Bayes estimator at stopping is $E^{(N,m)}\{B(\tilde{k}, T_{\tilde{k}}\} = 49.596$. The expected number of total number of observations is $E^{(N,m)}\{m, \tilde{k}\} = 204.05$. The mean absolute deviation of $B(\tilde{k}, T_{\tilde{k}})$ from the true value of N is given by $E^{(N,m)}\{|B(\tilde{k}, T_{\tilde{k}}) - N|\} = 0.4974$. By evaluating the probability distribution of $K(b_m)$ and of $B(\tilde{k}, T_{\tilde{k}})$, we can compare different sampling procedures, to determine what should be the sample size, m, if it is under our control and what value of λ to choose for our prior. We can also compare in this manner different stopping rules.

In Table 5 we present the values of $E^{(N,m)}\{B(\tilde{k}, T_{\tilde{k}}\}$, $E^{(N,m)}\{|B(\tilde{k}, T_{\tilde{k}}) - N|\}$ and $E^{(N,m)}\{K(b_m)m\}$, for various values of m and $\lambda = 10$, 50. As seen in this table, it is desirable to draw large samples, if possible. But this is not an important factor. For each λ, $mE^{(N,m)}\{K(b_m)\}$ varies slowly. The interesting result is that the stopping rule should be determined by calculating $R(k, t)$ with small prior mean λ.

References

Berg, S. (1987). Estimation of the size of a finite population in sequential sampling: Truncated combinatorial numbers, *Sequential Anal.* **6**, 165–177.

Chapman, D.G. (1954). The estimation of biological population. *Ann. Math. Statist.* **25**, 1–15.

Castledine, B.J. (1981). A Bayesian analysis of multiple recapture sampling for a closed population. *Biometrika* **67**, 197–210.

Darling, D.A. and H. Robbins (1967). Finding the size of a finite population. *Ann. Math. Statist.* **38**, 1392–1398.

Darroch, J.N. (1958). The multiple-recapture census, I. Estimation of a closed population. *Biometrika* **45**, 343–359.

Darroch, J.N. (1959). The multiple-recapture census, II. Estimation when there is immigration or death. *Biometrika* **46**, 336- 351.

DeGroot, M.H. (1970). *Optimal Statistical Decisions.* McGraw-Hill, New York.

Efron, B. and R. Thisted (1976). Estimating the number of unseen species: How many words did Shakespeare know. *Biometrika* **63**, 435–447.

Freeman, P.R. (1972). Sequential estimation of the size of the population. *Biometrika* **59**, 9–17.

Good, I.J. (1953). The population frequency of species and the estimation of population parameters. *Biometrika* **40**, 237–264.

Good, I.J. and G.H. Toulimin (1956). The number of new species and the increase in population coverage, when a sample is increased. *Biometrika* **43**, 45–63.

Goodman, L.A. (1953). Sequential sampling tagging for population size problems. *Ann. Math. Statist.* **24**, 56–69.

Henrici, P. (1974). *Applied and Computational Complex Analysis*, Vol. 1. John Wiley, New York.

Jolly, G.M. (1965). Explicit estimates from capture-recapture data with both death and immigration-stochastic model. *Biometrika* **52**, 225–247.

Jolly, G.M. (1982). Mark-recapture models with parameters constant in time. *Biometrics* **38**, 301–321.

Leite, J.G. and C.A. de B. Pereira (1987). An urn model for the capture-recapture sequential sampling process. *Sequential Anal.* **6**, 179–186.

Leite, J.C., J. Oishi and C.A. de B. Pereira (1988). Exact ML estimate of a Finite Population size: capture-recapture sequential sample data. *Probab. Inging. Inform. Sci.* **1**, 225–236.

Pollock, K.H. and M. Otto (1983). Robust estimation of population size in closed animal population from capture-recapture experiments. *Biometrics* **39**, 1035–1049.

Pollock, K.H., J.E. Hines and J.D. Nichols (1985). Goodness-of-fit tests for open capture-recapture models. *Biometrics* **41**, 399–410.

Samuel, E. (1968). Sequential Maximum Likelihood estimation of the size of a population. *Ann. Math. Statist.* **39**, 1057–1068.

Samuel, E. (1969). Comparison of sequential rules for estimation of the size of a population. *Biometrics* **25**, 517–527.

Seber, G.A.F. (1962). The multi-sample size recapture census. *Biometrika* **49**, 339–350.

Seber, G.A.F. (1965). A note on multiple-recapture census. *Biometrika* **52**, 249–259.

Seber, G.A.F. (1985). *The Estimation of Animal Abundance*, 2nd. ed., Griffin, London.