INFERÊNCIA ESTATÍSTICA EM POPULAÇÃO ANIMAL: PROCESSO DE CAPTURA-RECAPTURA

José Galvão Leite

Texto sistematizado de parte das atividades científicas apresentado ao Instituto de Matemática e Estatística da Universidade de São Paulo, para o concurso de

LIVRE-DOCÊNCIA NO DEPARTAMENTO
DE ESTATÍSTICA

São Paulo, março de 1992

ÍNDICE

INTRODUÇÃO

Após a conclusão de nosso doutorado em 1986, concentramos nossa atividade de pesquisa na área de inferência estatística em população animal. Uma parte dos resultados obtidos está contida em quatro artigos que publicamos durante os anos de 87, 88 e 90. O presente texto é uma sistematização desses trabalhos.

Discutiremos, no que segue, os principais resultados desses artigos, anexados no texto, procurando resumir as técnicas e os argumentos utilizados, bem como os resultados da área que os motivaram. Acreditamos que isto dará ao leitor uma melhor compreensão da linha de pesquisa em que estamos engajados e de nossa contribuição.

Gostaríamos de ressaltar que, todo nosso trabalho de pesquisa junto ao Departamento de Estatística do IME é uma consequência de nossa interação com os Profs. Carlos Alberto de Bragança Pereira, que orientou nosso trabalho de doutorado e com quem aprendemos a não complicar o simples, Josemar Rodrigues e Heleno Bolfarine, com quem tivemos o grato prazer de trabalhar e, em particular, com o Prof. Antonio Galves, a quem devemos grande parte de nossa formação.

CAPÍTULO I

Processo de Captura-Recaptura em População Animal

I.1 - Descrição do processo

O método de estimação do tamanho de uma população animal pelo processo de captura-recaptura consiste em selecionarmos, em uma primeira etapa (época), um número fixado ou aleatório de animais da população. Os animais capturados são marcados e devolvidos à população. Em seguida são selecionados, em cada uma de $k-1$ épocas ($k \geq 2$), um número fixado ou aleatório de animais. Em cada uma das k amostragens os animais não marcados recebem marcas e são devolvidos à população. O problema consiste em determinarmos uma estimativa do tamanho populacional, a partir do número de animais distintos observados durante todo o processo. O primeiro trabalho nesse contexto foi o do dinamarquês Petersen (1896), que estudou o fluxo migratório de peixes no mar Báltico. Em nosso trabalho subentendemos que a população em estudo é uma população fechada, isto é, não há nascimentos (imigrações) ou mortes (emigrações) durante as k épocas de amostragem.

I.2 - Estimativas de máxima verossimilhança

Denotemos por N o tamanho da população em estudo e suponhamos que na j-ésima seleção sejam capturados, sem reposição, m_j ($m_j \geq 1$) elementos dentre os quais U_j

elementos são não marcados, $j = 1, 2, \ldots, k$. A estatística $T_k = U_1 + U_2 + \cdots + U_k$ é o número de elementos distintos (não marcados) selecionados durante todo o processo. Em Leite e Pereira (1987) mostramos que T_k é uma estatística suficiente para N e que o kernel da verossimilhança é

$$K(N,t) = N! \left\{ (N-t)! \prod_{j=1}^{k} \binom{N}{m_j} \right\}^{-1} I_t(N) ,$$

onde t é o valor observado de T_k e $I_t(\cdot)$ é a função indicadora do conjunto $\mathbb{N}_t = \{n \in \mathbb{N}^* : n \geq t\}$, com $\mathbb{N}^* = \{1, 2, \ldots\}$. Notemos que $m \leq T_k \leq \min\{s, N\}$, onde $m = \max\{m_1, m_2, \ldots, m_k\}$ e $s = m_1 + m_2 + \cdots + m_k$. Darroch (1958) prova que a estimativa de máxima verossimilhança de N, $\widehat{N}$ é solução da equação

$$\prod_{j=1}^{k} (N - m_j) = (N-t)N^{k-1} .$$

No artigo **A note on the exact maximum likelihood estimation of the size of a finite and closed population** obtemos uma solução explícita dessa última equação, determinando assim um expressão exata para $\widehat{N}$. Provamos também que, para o caso de seleções um a um ($m_1 = m_2 = \cdots = m_k = 1$), a estimativa de máxima verossimilhança de N é única. A seguir apresentamos o resultado mais importante desse artigo.

Teorema. Uma estimativa de máxima verossimilhança de N, $\widehat{N}$, existe e é dada por

$$\widehat{N} = \begin{cases} t & \text{se } t = m, \\ t + n_t - 1 & \text{se } m < t < s, \\ \infty & \text{se } t = s, \end{cases}$$

onde $n_t = \min\{n \in \mathbb{N}^* : \prod_{j=1}^{k} (t + n - m_j) < n(t+n)^{k-1}\}$.

Essa estimativa é única exceto quando

$$\prod_{j=1}^{k} (t + n_t - m_j - 1) = (n_t - 1)(t + n_t - 1)^{k-1} ; \qquad (1)$$

neste caso há duas estimativas: $t + n_t - 1$ e $t + n_t - 2$. Se $m_1 = m_2 = \cdots = m_k = 1$ há uma única estimativa.

Samuel (1968) conjectura erradamente que, no caso de seleções um a um, a estimativa não é única. Esta conjectura vinha sendo aceita até a publicação do nosso trabalho. A prova desse teorema baseia-se no fato de que se $t = m$ então $K(N, t)$ é uma função decrescente de N, o que implica $\widehat{N} = t$; se $t = s$ então $K(N, t)$ é uma função crescente de N, o que implica $\widehat{N} = \infty$. Finalmente, se $m < t < s$ consideremos a função

$$f_t(x) = (1 - tx)^{-1} \prod_{j=1}^{k} (1 - m_j x), \quad x \in [0, t^{-1}) .$$

Pode-se provar que a equação $f_t(x) = 1$ admite uma única solução x_0 no intervalo $(0, t^{-1})$, com $f_t(x) < 1$ para todo $x \in (0, x_0)$ e $f_t(x) > 1$ para todo $x \in (x_0, t^{-1})$. Deste comportamento da função $f_t(\cdot)$ e do fato que

$$f_t\left(\frac{1}{n+1}\right) = \frac{K(n+1, t)}{K(n, t)} ,$$

para todo $n \in \mathbb{N}_t$ segue o resultado, exceto a unicidade para o caso de seleções um a um. No caso de $m_1 = m_2 = \cdots = m_k = 1$, a relação (1) torna-se

$$(t + n_t - 2)^k = (n_t - 1)(t + n_t - 1)^{k-1} . \quad (2)$$

Desde que a equação $(x - 1)^k = (x - t)x^{k-1}$ não tem solução em $\mathbb{N}^*$ segue que (2) não se verifica e, portanto, a estimativa de máxima verossimilhança é única.

CAPÍTULO II

Inferência Bayesiana em População Animal

II.1 - Introdução

Observemos que, com relação à estimativa de máxima verossimilhança de N discutida na seção anterior,

1º) o espaço paramétrico $\mathbb{N}_t$ muda com o valor observado t de T_k;

2º) o estimador definido a partir de $\widehat{N}$ não tem momentos finitos;

3º) as variáveis aleatórias $U_1, U_2, \ldots, U_k$, que constituem os dados, não são independentes e nem identicamente distribuídas. Tais fatos nos impedem de usar procedimentos estatísticos padrões no estudo do estimador de máxima verossimilhança de N. O uso de procedimentos Bayesianos, assunto desse capítulo, é um meio de contornar problemas deste tipo.

II.2 - A moda como estimativa do tamanho populacional

Consideremos, nesta seção, um modelo de captura-recaptura em que cada elemento da população tem, independentemente dos demais, uma probabilidade desconhecida p_j de ser capturado na j-ésima amostra, $j = 1, 2, \ldots, k$. Denotemos por X_j o número de elementos não marcados na j-ésima amostra, Y_j o número de elementos marcados na j-ésima amostra e por M_j o número de elementos marcados na população na época da j-ésima amostragem, $j = 1, 2, \ldots, k$. Notamos que $M_1 = 0$, $M_{j+1} = M_j + X_j =$

$X_1 + X_2 + \cdots + X_j$, $j = 1, 2, \ldots, k$ onde $M_{k+1} = T_k$. Para completar a notação seja $\mathbf{p} = (p_1, p_2, \ldots, p_k)$ e

$$\mathcal{D} = \{(x_j, y_j),\ \ j = 1, 2, \ldots, k\}$$

os dados observados. Logo,

$$X_j | M_j \sim B(N - M_j, p_j) \quad \text{e} \quad Y_j | M_j \sim B(M_j, p_j) \ ,$$

$j = 1, 2, \ldots, k$, o que implica que a função de verossimilhança é

$$\begin{aligned} L(N, \mathbf{p}|\mathcal{D}) &= \prod_{j=1}^{k} \binom{N - M_j}{x_j} \binom{M_j}{y_j} p_j^{n_j} (1 - p_j)^{N - n_j} \\ &\propto \binom{N}{t} \prod_{j=1}^{k} p_j^{n_j} (1 - p_j)^{N - n_j} \ , \end{aligned}$$

onde $n_j = x_j + y_j$ e t é o valor observado de T_k. Consideremos uma distribuição a priori para $(N, \mathbf{p})$ da forma

$$\pi(N, \mathbf{p}) = \pi(N) \prod_{j=1}^{k} \pi(p_j) \ ,$$

onde $\pi(N)$ e $\pi(p_j)$, $j = 1, 2, \ldots, k$ são distribuições não informativas. Logo, a distribuição a posteriori de $(N, \mathbf{p})$ é tal que $\pi(N, \mathbf{p}|\mathcal{D}) \propto L(N, \mathbf{p}|\mathcal{D})$, o que implica que a distribuição a posteriori marginal de N é tal que

$$\pi(N|\mathcal{D}) \propto \binom{N}{t} [(N + 1)!]^{-k} \prod_{j=1}^{k} (N - n_j)! I_t(N) \ .$$

No artigo **Exact expressions for the posterior mode of a finite population size: capture-recapture sequential sampling** determinamos a moda da distribuição a posteriori de N e mostramos que ela é sempre finita, ao contrário, como vimos na seção anterior, do que ocorre com a estimativa de máxima verossimilhança de N.

O principal resultado obtido é o seguinte:

Teorema. Existe uma única moda de $\pi(N|\mathcal{D})$, $\overline{N}$, dada por

$$\overline{N} = \begin{cases} t & \text{se } t = m, \\ t + n_t^* - 1 & \text{se } t > m, \end{cases}$$

onde $n_t^* = \min\{n \in \mathbb{N}^* : \prod_{j=1}^{k}(t + n - n_j) < \frac{n}{t+n}(t + n + 1)^k\}$.

Segue, deste teorema, que se $t > m$, então $\overline{N} = t$ se e somente se

$$\prod_{j=1}^{k}(t + 1 - n_j) < \frac{1}{t+1}(t + 2)^k \ .$$

A prova desse teorema segue do fato que, se $t = m$ então, supondo sem perda de generalidade que $t = n_k$, temos

$$\pi(N|\mathcal{D}) = \left\{(N+1)\prod_{j=1}^{k-1}\binom{N+1}{n_j+1}\right\}^{-1} I_t(N) \ ,$$

o que implica $\overline{N} = t$; se $t > m$ consideremos a função

$$g_t(x) = (1 - tx)^{-1}(1 + x)^{-k}\prod_{j=1}^{k}(1 - n_j x) \ , x \in [0, t^{-1}) \ .$$

Provamos que a equação $g_t(x) = 1$ admite uma única solução não nula x_0 no intervalo $[0, t^{-1})$, com $x_0 \neq \frac{1}{n}$ para todo $n \in \mathbb{N}_t$, $g_t(x) < 1$ para todo $x \in (0, x_0)$ e $g_t(x) > 1$ para todo $x \in (x_0, t^{-1})$. Tal comportamento da função $g_t(\cdot)$ e o fato que

$$g_t\left(\frac{1}{n+1}\right) = \frac{\pi(n+1|\mathcal{D})}{\pi(n|\mathcal{D})} \ ,$$

para todo $n \in \mathbb{N}_t$ implicam o resultado.

Assim, podemos considerar $\overline{N}$ como uma estimativa de N e, a título de ilustração, apresentamos a tabela abaixo que nos permite comparar $\overline{N}$ com $\widehat{N}$, para alguns valores de $n_1, n_2, \ldots, n_k$.

$(n_1, n_2, \ldots, n_k)$	t	$\overline{N}$	$\widehat{N}$
	62	63	63
(40, 60)	80	116	119 e 120
	100	1299	∞
	90	92	92
	120	150	152
(40, 60, 80)	140	232	239 e 240
	179	2715	10381
	180	3628	∞
	60	61	61
	80	93	95
	98	143	149 e 150
(15, 20, 25, 30, 50)	120	298	347
	139	1336	7449
	140	1609	∞

Se tomarmos uma distribuição a priori (imprópria) para $(N, \mathbf{p})$ da forma

$$\pi(N, \mathbf{p}) = \pi(N)\pi(\mathbf{p}) \propto \prod_{j=1}^{k} p_j^{-1} , \qquad (3)$$

segue que a distribuição a posteriori marginal de N é tal que

$$\pi(N|\mathcal{D}) \propto \binom{N}{t} (N!)^{-k} \prod_{j=1}^{k} [(N - n_j)!] I_t(N) .$$

O comportamento da moda desta última distribuição a posteriori é exatamente o mesmo de $\widehat{N}$. Se considerarmos a distribuição a priori imprópria para $(N, \mathbf{p})$ como sendo aquela dada em (3), teremos a versão Bayesiana do trabalho de Darroch (1958). Veja também Castledine (1981).

Evidentemente, por ser imprópria a distribuição a priori, não podemos garantir que a posteriori tenhamos uma distribuição própria e assim, como anteriormente, a moda pode não ser sempre finita.

II.3 - O estimador de Bayes do tamanho populacional

Suponhamos, como na seção I.2, que na j-ésima etapa do processo sejam selecionados, sem reposição, m_j elementos da população dentre os quais U_j elementos são não marcados, $j = 1, 2, \ldots, k$.

No artigo **Bayes estimation of the size of a finite population: capture-recapture sequential sample data** introduzimos o estimador de Bayes do tamanho populacional. Estudamos suas propriedades e, no caso de grandes amostras, algumas dessas propriedades são obtidas via Teoria dos Martingais e Supermartingais. Freeman (1972) e Zacks (1984) também trataram do problema. Contudo, ambos consideraram o caso de seleções um a um. Nossos resultados dependem da distribuição a priori de N ter segundo momento finito. Os resultados mais relevantes são os de convergência, com probabilidade um, do estimador de Bayes, do risco de Bayes e de T_k, quando k tende ao infinito. No que segue enunciaremos os principais resultados desse artigo.

Observadas as k amostras o vetor $\mathcal{D}_k = (u_1, u_2, \ldots, u_k)$, onde $u_1 = m_1$ e $u_j \in \{0, 1, \ldots, m_j\}$ para $j = 2, 3, \ldots, k$ é uma observação do vetor aleatório $\mathbb{D}_k = (U_1, U_2, \ldots, U_k)$. Sejam $S_j = m_1 + m_2 + \cdots + m_j$ e $M_j = \max\{m_1, m_2, \ldots, m_j\}$ para $j = 1, 2, \ldots, k$. Seja π uma distribuição de probabilidade a priori para N e seja $\mathbb{N}_t^\pi = \{x \in \mathbb{N}^* : x \geq t, \pi(x) > 0\}$ onde t, como antes, é o valor observado de T_k. Para todo $t \in \mathbb{N}^*$ tal que $M_k \leq t \leq S_k$ e $\mathbb{N}_t^\pi \neq \emptyset$, a distribuição de probabilidade a posteriori de N é dada por

$$\pi(n|k, t) = \lambda(n, k, t)K(k, t)\pi(n)I_t^\pi(n) ,$$

onde $\lambda(n,k,t)$ é o kernel da verossimilhança (veja Leite e Pereira (1987)),

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

e $I_t^{\pi}(\cdot)$ é a função indicadora do conjunto $\mathbb{N}_t^{\pi}$. Como $\lambda(n,k,t)$ é limitado (relação 3.3, p. 203 do artigo), então $K(k,t)$ é limitado. A restrição $M_k \leq t \leq S_k$ é natural pois, se $t < m_j$, para algum j, então as seleções seriam com reposição e se $t > S_k$ então, antes do início do processo de seleção já existiriam elementos marcados na população. É praticamente impossível definir-se uma classe conjugada de distribuições para o processo pois, para alguns pontos amostrais, a soma da verossimilhança para todos os possíveis valores de N pode divergir. Por exemplo, no caso dę seleções um a um, se considerarmos que a distribuição a priori de N é uma medida uniforme em $\mathbb{N}^*$, então $\pi(n|k,t)$ não está definida para $t = k-1$ e $t = k$. Com efeito,

$$[K(k,t)]^{-1} = \sum_{n=t}^{\infty} \lambda(n,k,t) = \sum_{n=t}^{\infty} \frac{n!}{(n-t)!n^k}$$

converge somente se $t \leq k-2$.

Seja $\pi(n)$ uma distribuição de probabilidade a priori para N com segundo momento finito. Para todo $t \in \mathbb{N}^*$ com $M_k \leq t \leq S_k$ e $\mathbb{N}_t^{\pi} \neq \emptyset$, a estimativa de Bayes de N é

$$\begin{aligned} \beta(k,t) &= E(N|T_k = t) = E(N|\mathbb{D}_k = \mathcal{D}_k) = \\ &= K(k,t) \sum_{n=t}^{\infty} n\lambda(n,k,t)\pi(n) \ . \end{aligned}$$

Como $\lambda(n,k,t)$ e $K(k,t)$ são limitados e π tem segundo momento finito, então $\beta(k,t)$ é finito. Para o caso de seleções um a um, com $t \leq k$ e $\mathbb{N}_t^{\pi} \neq \emptyset$, temos

$$\beta(k,t) = \frac{K(k,t)}{K(k-1,t)} \ .$$

Em particular, se π for a distribuição de Poisson com parâmetro θ, então

$$\beta(k,t) = \frac{E\left[(N+t)^{-k+1}\right]}{E\left[(N+t)^{-k}\right]} .$$

onde N é uma variável aleatória com distribuição de Poisson de parâmetro θ.

Na sequência a distribuição de probabilidade a priori de N tem segundo momento finito. Isto garante não só que $\beta(k,t)$ é finito, como vimos, mas também que o risco de Bayes

$$\rho(k,t) = \text{Var}\,(N|T_k = t) = \text{Var}\,(N|\mathbb{D}_k = \mathcal{D}_k)$$

é finito. Os seguintes teoremas mostram que $\beta(\cdot,\cdot)$ é uma função não crescente de k e uma função não decrescente de t.

Teorema 1. Para todo $k \geq 2$ e $t \in \mathbb{N}^*$ suponhamos $M_k \leq t \leq S_k$ e $\mathbb{N}_t^\pi \neq \emptyset$. Se $m_{k+1} \leq S_k$, então

$$\beta(k,t) \geq \beta(k+1,t) .$$

Vale a igualdade na relação acima se a distribuição de probabilidade a priori de N, π, for degenerada.

Teorema 2. Para todo $k \geq 2$ e $t \in \mathbb{N}^*$ suponhamos $M_k \leq t \leq S_k - 1$ e $\mathbb{N}_t^\pi \neq \emptyset$. Se $\pi(\cdot|k,t)$ não for degenerada no ponto t, então

$$\beta(k,t) \leq \beta(k,t+1) .$$

Vale a igualdade na relação acima se a distribuição a priori de N, π, for degenerada em qualquer ponto distinto de t.

A prova do Teorema 1 segue do fato que, a função

$$h(n) = \left\{ \binom{n}{m_{k+1}} \right\}^{-1} I_t(n) \ , \quad n \in \mathbb{N}^* \ ,$$

restrita ao conjunto $\mathbb{N}_t^\pi$ é decrescente.

Com efeito, esta propriedade da função $h(\cdot)$ implica (Lehmann (1966)) que

$$E(h(N)|T_k = t)E(N|T_k = t) \geq E(Nh(N)|T_k = t) \ ,$$

que por sua vez implica o resultado.

A prova do teorema 2 segue do fato que

$$\rho(k,t) = \text{Cov}\,(N, N - t|T_k = t) \ .$$

De fato, desta última relação segue que

$$\rho(k,t) = \frac{K(k,t)}{K(k,t+1)}\{\beta(k,t+1) - \beta(k,t)\} \ , \tag{4}$$

que implica o resultado.

Para estudar o comportamento de $\beta(k,t)$ e $\rho(k,t)$ no caso de grandes amostras suponhamos, como anteriormente, que m_j seja o tamanho da j-ésima amostra. Desde que N é finito, $\{m_j\}_{j\geq 1}$ é uma sequência limitada de elementos de $\mathbb{N}^*$ com $M = \max\{m_j : j \geq 1\}$. Para todo $t \in \mathbb{N}^*$ tal que $t \geq M$ e $\mathbb{N}_t^\pi \neq \emptyset$, seja $s = \min\{j \in \mathbb{N}^* : j \geq 2 \text{ e } S_j \geq t\}$. Para todo $k \geq s$, temos que $M_k \leq t \leq S_k$ e, consequentemente, $\beta(k,t)$ e $\rho(k,t)$ estão bem definidos. Do teorema 1 e do fato que $\beta(k,t) \geq 1$, para todo $k \geq s$, segue que a $\{\beta(k,t)\}_{k\geq s}$ (para t fixado) tem limite finito, quando $k \to \infty$.

O valor deste limite é dado pelo seguinte teorema.

Teorema 3. Para t fixado, $t \in \mathbb{N}^*$, se $t \geq M$, $\mathbb{N}_t^\pi \neq \emptyset$ e $\tau = \min N_t^\pi$, então

$$\beta(k,t) \underset{k\to\infty}{\longrightarrow} \tau \ .$$

Como uma consequência deste teorema temos o seguinte resultado.

Teorema 4. Para t fixado, $t \in \mathbb{N}^*$, se $t \geq M$ e $N_t^\tau \neq \emptyset$, então

$$\rho(k,t) \xrightarrow[k\to\infty]{} 0 \,.$$

A prova do teorema 3 baseia-se no fato que, para todo $k \geq s$, pode-se escrever

$$\beta(k,t) = \left\{ A\frac{(\tau-t)!}{\pi(\tau)\tau!} + \tau \right\} \left\{ B\frac{(\tau-t)!}{\pi(\tau)\tau!} + 1 \right\}^{-1} ,$$

onde

$$A = A(\tau,k) \to 0 \quad \text{e} \quad B = B(\tau,k) \to 0 \,,$$

quando k tende ao infinito.

Com relação à prova do teorema 4 observemos que, se $\Pi(\cdot|s,t)$ for degenerada, então $\Pi(\cdot|k,t)$ também será degenerada, para todo $k \geq s$, e o resultado vale; se $\Pi(\cdot|s,t)$ não for degenerada, então $\Pi(\cdot|k,t)$ não será degenerada no ponto t, para todo $k \geq s$. Por outro lado, da relação (4) temos que

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

e o resultado segue do teorema 3.

Para provar a consistência do estimador de Bayes e as convergências quase certa do risco de Bayes de T_k, construímos no item 6 do artigo uma família de espaços de probabilidades $(\Omega, \mathcal{F}, \{P_n : n \geq M\})$, onde P_n é uma probabilidade definida pela distribuição de $\mathbb{D}_k$ dado $N = n$, e um outro espaço de probabilidade $(\Omega^*, \mathcal{F}^*, \Pi)$, que é o chamado modelo Bayesiano. Interpretando N e T_k como variáveis aleatórias definidas em $(\Omega^*, \mathcal{F}^*, \Pi)$ temos o seguinte resultado.

Teorema 4. (i) $T_k \xrightarrow[k\to\infty]{} n$ quase certamente $[P_n]$, para todo $n \geq M$ e

(ii) $T_k \xrightarrow[k\to\infty]{} N$ quase certamente $[\Pi]$.

A prova deste teorema segue do fato que T_k converge em P_n - probabilidade para n, para todo $n \geq M$, e T_k converge em Π - probabilidade para N.

Finalmente, definimos uma família $\{\mathcal{F}_k\}_{k\geq 2}$ não decrescente de sub-σ-álgebras de $\mathcal{F}^*$ tal que, com relação a esta família, a sequência dos estimadores de Bayes de N, $\{\beta_k\}_{k\geq s}$, é um martingal e a sequência do risco de Bayes, $\{\rho_k\}_{k\geq s}$, é um supermartingal. Estabelecida esta identificação obtemos o principal resultado do artigo, dado pelo seguinte teorema.

Teorema 5. (i) $\beta_k \xrightarrow[k\to\infty]{} N$ quase certamente $[\Pi]$ e

(ii) $\rho_k \xrightarrow[k\to\infty]{} 0$ quase certamente $[\Pi]$.

Observamos que, segue imediatamente deste teorema o

Corolário 1. $\mathrm{Var}\,(\beta_k) \uparrow \mathrm{Var}\,(N)$ quando $k \to \infty$.

II.4 - Modelos de captura-recaptura com ou sem resposta a armadilha

Consideramos nesta seção modelos de captura-recaptura que levam em conta o efeito da armadilha sobre o comportamento futuro do animal. Se um animal capturado em alguma ocasião for penalizado, o mesmo se tornará arredio e espera-se que sua recaptura seja mais difícil; caso contrário, se não for penalizado (ou sentir-se "feliz" durante a captura), espera-se que sua recaptura seja fácil. Assim, incorporamos nos modelos as probabilidades de recaptura para as diversas ocasiões.

No artigo **A Bayesian analysis in closed animal populations from capture-recapture experiments with trap response**, estudamos na perspectiva Bayesiana alguns desses modelos. Consideramos o problema da escolha do modelo e determinamos a correspondente estimativa de N.

Várias distribuições a priori foram estudadas com o objetivo de estudar suas influências sobre a distribuição a posteriori. Nossa contribuição foi a extensão dos resultados de Castledine (1981), que trata do assunto sem levar em conta o efeito da armadilha sobre o comportamento do animal. Antes de discutirmos os principais resultados do artigo introduzimos a seguinte notação.

Sejam p_j a probabilidade de que um elemento da população seja capturado pela primeira vez, independentemente dos demais, na j-ésima amostra e c_j a probabilidade de que um elemento da população seja recapturado, independentemente dos demais, na j-ésima amostra, $j = 1, 2, \ldots, k$. Denotando por $\mathbf{c}$ o vetor $(c_1, c_2, \ldots, c_k)$ e utilizando a mesma notação da seção II.2, consideremos os seguintes modelos.

M_{1h}^* : "heterogeneidade com resposta a armadilha" - caso onde $p_j \neq c_j$ para algum j, $1 \leq j \leq k$;

M_{0h}^* : "heterogeneidade com não resposta a armadilha" - caso onde $p_j = c_j$ para todo j, $1 \leq j \leq k$;

M_1^* : "homogeneidade com resposta à armadilha" - caso onde $p_j = p$, $c_j = c$ e $p \neq c$, $1 \leq j \leq k$, e

M_0^* : "homogeneidade com não resposta à armadilha" - caso onde $p_j = p$, $c_j = c$ e $p = c$, $1 \leq j \leq k$.

Observemos que, os modelos M_{0h}^* e M_0^* foram estudados por Castledine (1981);

Pollock (1975) mostrou que, para o modelo M_{1h}^* no caso $k = 2$, não é possível se determinar o estimador de máxima verossimilhança de N, a menos de algumas restrições sobre o espaço paramétrico.

No caso do modelo M_1^*, a estimativa de máxima verossimilhança de N, $\widehat{N}$, é dada pela solução da equação

$$\widehat{N} = t\{1 - (1 - t(k\widehat{N} - \sum_{j=1}^{h} M_j))^k\}^{-1} .$$

Para comparar os modelos M_{0h}^* e M_{1h}^* sob a ótica Bayesiana, definimos o fator de Bayes

$$K = \frac{\pi_0 p(\mathcal{D}|M_{0h}^*)}{\pi_1 p(\mathcal{D}|M_{1h}^*)} , \tag{5}$$

onde π_i denota o peso a priori associado ao modelo M_{ih}^*, $i = 0, 1$ e $p(\mathcal{D}|M_{ih}^*)$ a densidade preditiva sob o modelo M_{ih}^*, $i = 0, 1$.

Considerando a distribuição a priori para $(N, \mathbf{p}, \mathbf{c})$ do tipo

$$\pi(N, \mathbf{p}, \mathbf{c}) = \pi(N)\pi(\mathbf{p}, \mathbf{c}) , \tag{6}$$

onde $\pi(N) = 1$, $N \in \mathbb{N}^*$ e $\pi(\mathbf{p}, \mathbf{c}) = 1$, $0 \leq p_j \leq 1$ e $0 \leq c_j \leq 1$, $j = 1, 2, \ldots, k$ e $\pi_0 = \pi_1 = 0, 5$, mostramos que K é aproximadamente igual à expressão 9 (p. 414) do artigo e, para k suficientemente grande, aproximadamente igual à expressão 10 (p. 415). A distribuição a posteriori marginal de N é tal que, sob o modelo M_{1h}^*,

$$\pi(N|\mathcal{D}) \propto \prod_{j=1}^{k} \{N + 1 - M_j\}^{-1} I_t(N)$$

e, sob o modelo M_{0h}^*,

$$\pi(N|\mathcal{D}) \propto \binom{N}{t} [(N+1)!]^{-k} \prod_{j=1}^{k} (N - n_j)! I_t(N) .$$

Logo, tomando como estimativa de N a moda da distribuição a posteriori de N, $\overline{N}$, temos que, sob o modelo M_{1h}^*, $\overline{N} = t$ e, sob o modelo M_{0h}^*, $\overline{N}$ é dada na seção II.2. Observamos que, sob o modelo M_{1h}^* e com a distribuição a priori não informativa para $(N, \mathbf{p}, \mathbf{c})$, t é a solução Bayesiana do problema da estimação de N discutido por Pollock (1975). Assim, o problema da estimação de N fica contornado com a adoção da priori não informativa. A título de ilustração reanalisamos o exemplo da seção 5 de Castledine (1981). Neste caso $K \cong 3,498 \times 10^{-5}$. Logo, $p(M_{0h}^*|\mathcal{D}) = 3,489 \times 10^{-5}$. Isto mostra que M_{0h}^* não é um modelo adequado, desde que com base nos dados ele tem suporte muito pequeno. Então, é razoável a escolha do modelo M_{1h}^* e $\overline{N} = 135$.

Para comparar os modelos M_0^* e M_1^* poderíamos usar o fator de Bayes dado em (5). Contudo, na seção 4.1 do artigo propomos outro método, que consiste em considerarmos a distribuição a priori para $(N, \mathbf{p}, \mathbf{c})$ como em (6) e a distribuição a posteriori de $q = \frac{W}{Z}$, onde $W = \log\left(\frac{p}{1-p}\right)$ e $Z = \log\left(\frac{c}{1-c}\right)$. A distribuição a posteriori de q é proporcional à expressão 16 (p. 419) do artigo. A decisão por M_0^* ou M_1^* será baseada no menor intervalo de credibilidade de probabilidade $(1-\alpha)$ 100% . Se o intervalo contiver a unidade então M_0^* deve ser usado. Caso contrário M_1^* deve ser usado. Para estimar N observamos que, segundo o modelo M_1^*, a distribuição a posteriori ou marginal de N é proporcional à expressão 18 (p. 420) do artigo e, sob o modelo M_0^*, à expressão 19 (p. 421). Neste contexto, para o exemplo da seção 5 de Castledine (1981) determinamos que o intervalo de 95% para q é $(0,59; 0,85)$. Então M_1^* é o modelo escolhido e $\overline{N} = 238$.

Como distribuição a priori informativa para $(\mathbf{p}, \mathbf{c})$ tomamos, sob o modelo M_1^*, $p_j = p \sim \text{Beta}(a_1, b_1)$ e $c_j = c \sim \text{Beta}(a_2, b_2)$, $j = 1, 2, \ldots, k$, com p e c independentes e sob o modelo M_{1h}^*, $p_j \sim \text{Beta}(a_1, b_1)$ e $c_j \sim \text{Beta}(a_2, b_2)$ com $p_1, c_1, p_2, c_2, \ldots, p_k, c_k$

independentes. As distribuições a posteriori marginal de N para os modelos M_0^* e M_{1h}^* são dadas pelas expressões 22 e 23 (p. 425) do artigo. Como em Castledine (1981), para analisar a sensibilidade da distribuição a posteriori de N consideramos o quociente

$$\lambda = \frac{\Pi(N|\mathcal{D})}{\Pi(N)}$$

e concluímos que, quando k cresce a influência dos parâmetros a_1 e b_1 é muito maior no modelo M_{1h}^* do que no modelo M_0^*.

Finalmente, sob o modelo M_{1h}^* tomamos a distribuição de Dirichlet como distribuição a priori para $(\mathbf{p}, \mathbf{c})$. Nesse caso a distribuição a posteriori marginal de N é dada pela expressão 26 (p. 428) do artigo, que pode ser simplificada quando os c_j forem suficientemente pequenos. Para alguns valores dos parâmetros da distribuição de Dirichlet, deteminamos a moda da distribuição a posteriori de N.

REFERÊNCIAS

Castledine, B.J. (1981). A Bayesian analysis of multiple-recapture sampling for a closed population. *Biometrika* 67(1), 197-210.

Darroch, J.N. (1958). The multiple-recapture census I. Estimation of a closed population. *Biometrika* 45, 343-59.

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, 1137-53.

Leite, J.G. & Pereira, C.A.B. (1987). An urn model for the multi-sample capture-recapture sequential tagging process. *Sequential Anal.* 6(2), 179-86.

Petersen, G.G.J. (1896). The yearly immigration of young plaice into the Limfjord from the German Sea. *Rept. Danish Biol. Statist.* 6, 1-48.

Pollock, K. (1975). Building models of capture-recapture experiments. *The Statistician* 25, 253-59.

Samuel, E. (1968). Sequential maximum likelihood estimation of the size of a population. *Ann. Statist.* 39, 1057-68.

Seber, G.A.F. (1982). The estimation of animal abundance, *2nd ed. London: Griffin.*

Seber, G.A.F. (1986). A review of estimating animal abundance. *Biometrics* 42, 267-92.

Zacks, S. (1984). Bayes sequential estimation of the size of a finite population. *University of São Paulo, Brazil. RT-MAE-8404.*

Biometrika (1988), 75, 1, *pp.* 178–80
Printed in Great Britain

A note on the exact maximum likelihood estimation of the size of a finite and closed population

BY JOSÉ GALVÃO LEITE, JORGE OISHI
AND CARLOS ALBERTO DE BRAGANÇA PEREIRA
Instituto de Matemática e Estatística, Universidade de São Paulo,
C.P. 20570, São Paulo, SP 01498, Brazil

SUMMARY

Using data obtained by the general capture-recapture sequential sampling process, an analytical expression for the maximum likelihood estimate of the population size is introduced. It is shown that the bounded likelihood functions have at most two maxima. For the simple one-to-one case the estimate is unique.

Some key words: Capture-recapture sequential sampling process; Maximum likelihood estimate; Sufficient statistic.

1. INTRODUCTION

The objective of this note is to present a closed analytical expression for the maximum likelihood estimate to the size, N, of a finite and closed population when the data are obtained by capture-recapture sequential sampling. Inferences about N based on data obtained in special cases have been considered by many authors; see, for instance, Seber (1982, Ch. 3, 4; 1986) for a complete reference list.

Consider a population of size N that changes neither in size nor in form; that is, the population is closed. From this population, $k>1$, random samples without replacement are sequentially selected from the population. Each of these samples is returned to the population before the next is selected. For the jth ($j=1,\ldots,k$) sample, the scientist records the sample size $m_j \geq 1$ and the number U_j of units selected for the first time; that is, units that were not selected in samples 1 to $j-1$. 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 and that the smallest factor of the likelihood function that depends on the value of N, the likelihood kernel, is

$$K(N,t) = N!\left\{(N-t)!\prod_{j=1}^{k}\binom{N}{m_j}\right\}^{-1} I_t(N),$$

where t is the observed value of T_k and $I_t(.)$ is the indicator function of $N_t=\{n \geq t\}$. The probability distribution of T_k (Leite & Pereira, 1987) is

$$\operatorname{pr}\{T_k = t \mid N\} = K(N,t)\,t!\sum_{i=0}^{t}(-1)^{t-i}\binom{t}{i}\binom{i}{m_1}\cdots\binom{i}{m_k} I^*(t),$$

where $m=\max\{m_1,\ldots,m_k\}$, $s=m_1+\ldots+m_k$ and $I^*(t)$ is the indicator function of

$$\{m \leq t \leq \min(N,s)\}.$$

2. MAIN RESULTS

If $t=m$, its smallest possible value, $K(N,t)$ is a decreasing function of N so that the maximum likelihood estimate is $\hat{N}=t$. At the other extreme, when $t=s$, $K(N,t)$ is an increasing function of N, so that $\hat{N}=\infty$.

Let now $m<t<s$, and consider the function

$$f_t(x)=(1-xt)^{-1}\prod_{j=1}^{k}(1-xm_j)$$

defined in $0\leq x\leq t^{-1}$. This function is continuous in $[0, t^{-1})$, is equal to unity at $x=0$, goes to ∞ as x increases to t^{-1}, and, if $n\geq t$,

$$f_t\left(\frac{1}{n+1}\right)=\frac{K(n+1, t)}{K(n, t)}.$$

The behaviour of f_t is described next.

LEMMA. *For $m<t<s$, the equation $f_t(x)=1$ has a unique positive solution x_0 in the open interval $(0, t^{-1})$. Also, $f_t(x)<1$ if $0<x<x_0$ and $f_t(x)>1$ if $x_0<x<t^{-1}$.*

Proof. Note that $f_t=g/h_t$, where g and h_t are defined as

$$g(x)=\prod_{j=1}^{k}(1-xm_j),\quad h_t(x)=1-xt.$$

The first and second derivatives of g are, respectively, negative and positive. Hence, g is a decreasing convex continuous function. Also, h_t is a decreasing linear function, $g(0)=h_t(0)=1$, $g(t^{-1})>0$ and $h_t(t^{-1})=0$. The derivative of (h_t-g) evaluated at $x=0$ is $m_1+\ldots+m_k-t>0$. Consequently, there is a unique point, x_0, in the open interval $(0, t^{-1})$, such that $g(x_0)=h_t(x_0)$, $g(x)<h_t(x)$ if $0<x<x_0$, and $g(x)>h_t(x)$ if $x_0<x<t^{-1}$. □

THEOREM. *A maximum likelihood estimate of N exists and is defined as*

$$\hat{N}=\begin{cases}t & (t=m),\\ t+n_t-1 & (m<t<s),\\ \infty & (t=s),\end{cases}$$

where $n_t=\min\{n; (t+n-m_1)\ldots(t+n-m_j)<n(t+n)^{k-1}\}$. Also this estimate is unique except when

$$\prod_{j=1}^{k}(t+n_t-m_j-1)=(n_t-1)(t+n_t-1)^{k-1}. \tag{1}$$

In this case the only two possible estimates are $(t+n_t-1)$ and $(t+n_t-2)$. If $m_1=\ldots=m_k=1$ it is always unique.

The proof follows directly from the Lemma. The uniqueness in the one-by-one case is not in agreement with Samuel (1968). To see this, for $1<t<k$, note that (1) simplifies to

$$(t+n_t-2)^k=(n_t-1)(t+n_t-1)^{k-1}$$

and, defining x to be the integer $t+n_t-1$, we can write $(x-1)^k=(x-t)x^{k-1}$. This last equation is equivalent to

$$(t-k)x^{k-2}+\sum_{i=1}^{k}\binom{k}{i}(-1)^i x^{k-i-1}=x^{-1}(-1)^{k-1}.$$

Since $x>t>1$ is an integer, the left-hand side of this equation must also be an integer but the right-hand side cannot be an integer. Hence the equation has no integer solution.

3. COMMENTS

Most recent publications on inference about N are restricted to the case on $k=2$. See for instance Isaki (1986) and Pollock, Hines & Nichols (1985). However, with the simple expression of the maximum likelihood estimator obtained from N, some of their analysis can be extended to the general case of $k\geq 2$.

Before using $\hat{N}$ one needs to observe the following limitations.

(i) The parameter space N_t changes with the observed value t of T_k.

(ii) the random variables U_i, for $i = 1, \ldots, k$, that form the data are not independently and identically distributed. In fact they are not even exchangeable.

(iii) The estimator defined from $\hat{N}$ has no finite moments.

(iv) When either event $\{T_k = m\}$ or $\{T_k = s\}$ occurs the value of $\hat{N}$ is not related to the value of k.

Facts (i), (ii) and (iii) restrict the use of standard statistical procedures. The use of Bayesian procedures may be the way to counter these problems since they rely on the observed data rather than on the distributional properties of T_k. To the best of our knowledge Freeman (1972) is the only available reference for the Bayesian estimation of N.

The strongest limitation on $\hat{N}$ is, in our opinion, introduced by (iv). To make the number of samples, k, relevant when $T_k = m$ or $T_k = s$, one needs to use prior knowledge.

ACKNOWLEDGEMENTS

The authors are grateful to Luis Pericchi for his comments on earlier drafts and the referee for useful suggestions.

REFERENCES

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

ISAKI, C. T. (1986). Bias of the dual system estimator and some alternatives. *Comm. Statist.* A **15**, 1435–50.

LEITE, J. G. & PEREIRA, C. A. DE B. (1987). An urn model for the multi-sample capture/recapture sequential tagging process. *Sequential Anal.* **6** (2), 179–86.

POLLOCK, K. H., HINES, J. E. & NICHOLS, J. D. (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. Statist.* **39**, 1057–68.

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

SEBER, G. A. F. (1986). A review of estimating animal abundance. *Biometrics* **42**, 267–92.

[*Received September* 1986. *Revised April* 1987]

Revista Brasileira de Probabilidade e Estatística (1987), 1, pp. 91-100

EXACT EXPRESSIONS FOR THE POSTERIOR MODE OF A FINITE POPULATION SIZE: CAPTURE-RECAPTURE SEQUENTIAL SAMPLING

José Galvão Leite, Heleno Bolfarine and Josemar Rodrigues

Departamento de Estatística
Universidade de São Paulo
Caixa Postal 20570, 01498 São Paulo, SP, Brasil

Summary

Using data obtained by a multiple capture-recapture experiment that depends on the catchability of the individuals in a closed population, an exact analytical expression for the mode of the posterior distribution of the population size is derived. The prior structure considered is noninformative. It is shown that there are cases where the posterior mode is always finite, a result that was conjectured by Leite et al. (1986).

Key words: Capture-recapture experiment; sequential sampling; posterior distribution mode; maximum likelihood estimator.

1. Introduction

The objective of this paper is to derive a closed analytical expression for the posterior mode of the population size, N, of a closed population, when a capture-recapture model that depends on the "catchability" of the individuals is considered. It is shown that there are situations where the posterior mode is always finite, a result that was conjectured by Leite et al. (1986), when deriving exact expressions for the ML estimate of N. But unfortunately there are situations where the ML estimate is infinite, what also may happen with the posterior mode, under a particular class of improper priors.

As in Castledine (1981), the capture probabilities are assumed to be the same for all animals, but with the possibility of changing over time. A review of the literature is given in Seber (1985).

In Section 2, the basic notation is presented and the posterior distribution of N is derived. Section 3 presents the main results of the paper. The behavior of the posterior mode is illustrated for various practical situations.

2. Basic notation and the posterior distribution of N

As in Castledine (1981) let N denote the unknown population size, s ($s \geq 2$) the number of samples taken, p_i , $1 \leq i \leq s$, the unknown probability of each animal to be captured in the i-th sample, X_i the unmarked number of animals in the i-th sample, Y_i the number of marked animals in the i-th sample ($Y_1 = 0$) and M_i the number of marked animals in the population just before the i-th sample ($M_1 = 0$). Note that $M_{i+1} = M_i + X_i = \sum_{j=1}^{i} X_j$. To complete the notation, let $\underset{\sim}{p} = (p_1, p_2, \ldots, p_s)$ and

$$\mathcal{D} = \{(x_i, y_i),\ i = 1,2,\ldots,s,\ y_1 = 0\} ,$$

the observed data. X_i and Y_i, $i = 1,2,\ldots,s$ are assumed to be independent (conditionally on the M_i). Therefore, under the assumptions made above,

$$X_i | p_i \sim \mathcal{B}(N - M_i, p_i, x_i) \ , \quad i = 1,2,\ldots,s$$

$$Y_i | p_i \sim \mathcal{B}(M_i, p_i, y_i) \ , \qquad i = 2,3,\ldots,s ,$$

from where it follows that the likelihood function of N and $\underset{\sim}{p}$ is

$$L(N,\underset{\sim}{p},\mathcal{D}) = \prod_{i=1}^{s} \binom{N-M_i}{x_i}\binom{M_i}{y_i} p_i^{n_i}(1-p_i)^{N-n_i} \ \alpha \ \binom{N}{r} \prod_{i=1}^{s} p_i^{n_i} \ (1-p_i)^{N-n_i} ,$$

where $n_i = x_i + y_i$ and $r = \sum_{i=1}^{s} x_i$, the total number of different animals captured.

Considering a prior distribution for $(N,\underset{\sim}{p})$ such that

$$\pi(N,\underset{\sim}{p}) = \pi(N)\pi(\underset{\sim}{p}) \quad , \tag{2.1}$$

where $\pi(N)$ and $\pi(\underset{\sim}{p})$ are noninformative distributions, it follows that the joint posterior density of $(N,\underset{\sim}{p})$ is given by

$$\pi(N,\underset{\sim}{p}|\mathcal{D}) \;\alpha\; L(N,\underset{\sim}{p}|\mathcal{D}). \tag{2.2}$$

By integrating out $\underset{\sim}{p}$ in (2.2), it follows that

$$\pi(N|\mathcal{D}) \;\alpha\; \frac{\binom{N}{r}}{[(N+1)!]^s} \prod_{i=1}^{s} (N-n_i)! \;, \tag{2.3}$$

where $N \geq r$ and $\max\{n_1,n_2,\ldots,n_s\} \leq r \leq \sum_{j=1}^{s} n_j$.

If the improper prior

$$\pi(N,\underset{\sim}{p}) = \pi(N)\pi(\underset{\sim}{p}) \;\alpha\; \prod_{i=1}^{s} p_i^{-1} \tag{2.4}$$

is taken for N and $\underset{\sim}{p}$, then, it follows that the posterior probability function of N is given by

$$\pi(N|\mathcal{D}) \;\alpha\; \frac{\binom{N}{r}}{[N!]^s} \prod_{i=1}^{s} (N-n_i)! \;. \tag{2.5}$$

Improper prior (2.4) is used for relating the Bayesian approach to the work of Darroch (1958). See also Castledine (1981).

The behavior of the mode of the posterior distribution (2.5) is exactly the same as the behavior of the ML estimate considered in Leite et al. (1986).

3. Analytical expression for the mode of $\pi(N/D)$ with noninformative prior structure

For the observed data $\mathcal{D}$, the posterior mode of N is a point $\hat{N} \in \{n \in \mathbb{N}^*;\ n \geq r\}$ ($\mathbb{N}^* = \{1,2,\ldots\}$) such that it maximizes the posterior probability function of N. After simple algebraic manipulations, (2.3) may be written as

$$\pi(N|\mathcal{D}) \propto \frac{N!}{(N-r)!\prod_{j=1}^{s}\binom{N+1}{n_j+1}} . \tag{3.1}$$

The following result introduces the mode of (3.1) for the extreme case where $r = \max\{n_1,n_2,\dots,n_s\}$.

Lemma 3.1. If $r = \max\{n_1,n_2,\dots,n_s\}$, then $\hat{N} = r$ is the unique mode of (2.3).

Proof. After simple algebraic manipulations, (3.1) may be written as

$$\pi(N|\mathcal{D}) \propto \frac{\binom{N+1}{r+1}}{(N+1)\prod_{j=1}^{s}\binom{N+1}{n_j+1}}, \quad N \geq r. \tag{3.2}$$

Supposing, without any loss of generality, that $r = n_s$, it follows from (3.2) that

$$\pi(N|\mathcal{D}) \propto \frac{1}{(N+1)\prod_{j=1}^{s-1}\binom{N+1}{n_j+1}}, \quad N \geq r ,$$

from where the result follows. □

Considering the case where $r > \max\{n_1,n_2,\dots,n_s\}$, let

$$K(N) = \frac{N!}{(N-r)!\prod_{j=1}^{s}\binom{N+1}{n_j+1}}, \quad N \geq r .$$

It follows from (3.1) that $\hat{N}$ is the point which maximizes K(N) and, for all $N \geq r$,

$$\frac{K(N+1)}{K(N)} = \left(1 - \frac{r}{N+1}\right)^{-1}\left(1 + \frac{1}{N+1}\right)^{-s}\prod_{j=1}^{s}\left(1 - \frac{n_j}{N+1}\right).$$

Define the function

$$g(x) = (1-rx)^{-1}(1+x)^{-s} \prod_{j=1}^{s} (1-n_j x) \ , \ x \in \left[0,\tfrac{1}{r}\right).$$

For all $N \geq r$,

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

In the theorem that follows, the behavior of the function g is studied.

Theorem 3.1. If $r > \max\{n_1,n_2,\ldots,n_s\}$, then the equation $g(x) = 1$ has only one non null root x_o in the interval $[0,\frac{1}{r})$. For $x \in (0,x_o), g(x)<1$ and, for $x \in (x_o,\frac{1}{r})$, $g(x) > 1$.

Proof. Consider the functions $g_1(x) = \prod_{j=1}^{s} (1-n_j x)$ and $g_2(x)=(1-rx)(1+x)^s$, for all real x. The first and second derivatives of g_1 are such that

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

and

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

for all x in the interval $(0,\frac{1}{r})$.

Therefore, g_1 is a decreasing convex continuous function in the interval $[0,\frac{1}{r}]$. The function g_2 is continuous and its first and second derivatives are

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

and

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

respectively, for all real x. Thus

(a) if $r \geq s$, $g_2'(x) < 0$ and $g_2''(x) < 0$, for all $x \in (0,\frac{1}{r})$, that is, g_2 is a decreasing and concave function in the interval $[0,\frac{1}{r}]$. At the origin, $g_1(0) = g_2(0) = 1$, and at the point $\frac{1}{r}$, $g_1(\frac{1}{r}) > 0$ and $g_2(\frac{1}{r})=0$. The derivative of $g_2 - g_1$ evaluated at the origin is

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

Therefore, there is a positive real number δ, such that $g_2'(x)-g_1'(x)>0$, for all x in the interval $[0,\delta)$. It follows, from the Mean Value Theorem, that $g_2(x) > g_1(x)$, for all x in the interval $(0,\delta)$. Consequently, there is a unique point $x_o \in (0,\frac{1}{r})$, such that

$$g_1(x_o) = g_2(x_o) ,$$

$$g_1(x) < g_2(x) , \text{ for all } x \in (0,x_o), \text{ and}$$

$$g_1(x) > g_2(x) , \text{ for all } x \in (x_o,\tfrac{1}{r}).$$

Since g is the restriction of $\frac{g_1}{g_2}$ to the interval $[0,\frac{1}{r})$, the result follows.

(b) If $r < s$, $g_2'(x) > 0$ for all $x \in \left(0, \frac{s-r}{r(s+1)}\right)$;

$g_2'(x) < 0$ and $g_2''(x) < 0$ for all $x \in \left(\frac{s-r}{r(s+1)},\frac{1}{r}\right)$, implying that g_2 is increasing in the interval $\left[0,\frac{s-r}{r(s+1)}\right]$ and decreasing and concave in $\left[\frac{s-r}{r(s+1)},\frac{1}{r}\right]$. So, there is a unique point $x_o \in \left(\frac{s-r}{r(s+1)},\frac{1}{r}\right)$ such that

$$g_1(x_o) = g_2(x_o) ,$$

$$g_1(x) < g_2(x), \text{ for all } x \in (0,x_o), \text{ and}$$

$$g_1(x) > g_2(x), \text{ for all } x \in (x_o,\tfrac{1}{r}).$$

Since $g = \frac{g_1}{g_2}$ in the interval $[0,\frac{1}{r}]$, the proof is completed. □

Figure 1 below illustrates the behavior of the functions g_1 and g_2

in the interval $[0,\frac{1}{r}]$, as described in Theorem 3.1, in a general situation.

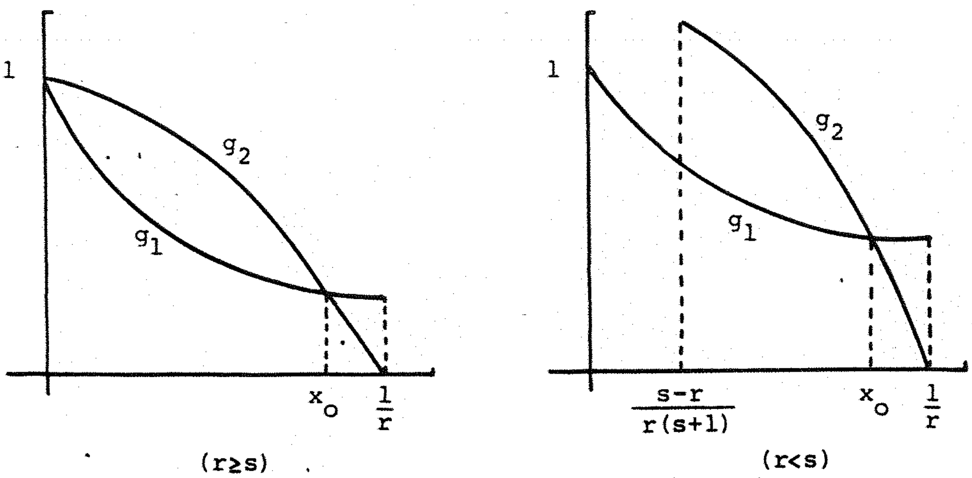

Figure 1

Functions g_1 and g_2 in a general situation.

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

$$\sum_{j=0}^{s} \binom{s}{j} n^j - \prod_{j=1}^{s} (n-n_j) - r \sum_{j=1}^{s} \binom{s}{j} n^{j-1} = \frac{r}{n} .$$

The left-hand side of the above equation is an integer. Since the right hand side can not be an integer, it follows that $x_o \neq \frac{1}{n}$, for all $n > r$.

The main result of the paper is stated next. Let $m = \max\{n_1,n_2,\ldots,n_s\}$ and

$$n_r = \min\{n \in \mathbb{N}^*;\ \prod_{j=1}^{s} (r+n-n_j) < \frac{n}{r+n} (r+n+1)^{r}\} .$$

Theorem 3.2. For all $s \geq 2$ there exists a unique mode of $\pi(N|\mathcal{D})$, $\hat{N}$, defined by

$$\hat{N} = \begin{cases} m & \text{if } r = m \\ r+n_r-1 & \text{if } r > m. \end{cases}$$

Proof. From Lemma 3.1, $\hat{N} = r$ if $r = m$. If $m < r$, it follows from Theorem 3.1 that there exists $n_o \in \mathbb{N}^*$, $n_o \geq 1$, such that, for $n \in \mathbb{N}^*$

$$g\left(\frac{1}{r+n}\right) \begin{cases} < 1 & \text{for } n \in \mathbb{N}^*,\ n \geq n_o \\ > 1 & \text{for } n \in \mathbb{N}^*,\ n \leq n_o-1. \end{cases}$$

Therefore, $n_o = \min\{n \in \mathbb{N}^*;\ g\left(\frac{1}{r+n}\right) < 1\}$

$$= \min\{n \in \mathbb{N}^*;\ \prod_{j=1}^{s} (r+n-n_j) < \frac{n}{r+n}(r+n+1)^s\};$$

that is, $n_o = n_r$ and from $g\left(\frac{1}{r+n}\right) = \frac{K(r+n)}{K(r+n-1)}$, $n \in N^*$, it follows that

$$K(r+n_r-1) > K(r+n_r-2) > \dots > K(r)$$

and

$$K(r+n_r-1) > K(r+n_r) > K(r+n_r+1) > \dots .$$

So, $\hat{N} = r+n_r-1$ is the only mode of $\pi(N|\mathcal{D})$, given by (2.3). □

The next result is a direct consequence of Theorem 3.2.

Corollary. If $r > m$, then $\hat{N} = r$ is the mode of (2.3), if and only if,

$$\prod_{j=1}^{s} (r+1-n_j) < \frac{1}{r+1}(r+2)^s. \qquad \square$$

From the posterior probability function (2.5), it follows that if an improper prior distribution is taken for $\underset{\sim}{p}$, then the posterior mode may be infinite, as it happens with the ML estimate. On the other hand, if a proper prior is taken for $\underset{\sim}{p}$, it follows from Theorem 3.2 that the posterior mode of N is always finite. Therefore, the conjecture made by Leite et al. (1986) is right, as long as a proper prior distribution is taken for $\underset{\sim}{p}$.

Table 1 bellow illustrates the behavior of the posterior mode given in Theorem 3.2, for some numerical examples.

For the sunfish data in Castledine (1981), it follows from Theorem 3.2 that $n_r = 160$ and then $\hat{N} = 295$.

Table 1

Examples of posterior modes of (2.3)

$n_1, n_2, \ldots, n_s$	r	ML estimate (mode of (2.5))	Posterior mode of (2.3) (Theorem 3.2)
(40,60)	62	63	63
	80	119 & 120	116
	100	∞	1299
(1,5,8)	10	12	11
	11	16	13
	12	25	17
	14	∞	30
(40,60,80)	90	92	92
	120	152	150
	140	239 & 240	232
	179	10381	2715
	180	∞	3628
(3,3,4,4,5)	6	6	6
	7	7	7
	10	11	10
	17	67	29
	18	139	35
	19	∞	44
(15,20,25,30,50)	60	61	61
	80	95	93
	98	149 & 150	143
	120	347	298
	139	7449	1336
	140	∞	1609

(Received March 1987. Revised August 1987)

References

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

Darroch, J.N. (1958). The multiple recapture census I: Estimation of a closed population. *Biometrika* 45, 343-59.

Leite, J.G., Oishi, J. & Pereira, C.A.B. (1987). Exact ML estimate of a finite population size: capture-recapture sequential sample data. *Probability in the Engineering and Informational Sciences* 1, 225-236.

Seber, G.A.F. (1986). A Review of estimating animal abundance. *Biometrics* 42, 267-292.

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

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

Note that any reasonable person should try to decrease the posterior variance, and consequently its expectation $E\{\mathrm{Var}\,\{N \mid T_k\}\}$, to zero. This corresponds to increasing the variance of the Bayes' estimator to $\mathrm{Var}\,\{N\}$, 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 jth $(j > 1)$ 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

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

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

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

where t is the observed value of T_k, $\mathscr{I}_t(n)$ is the indicator function of $\mathbb{N}_t = \{x \in \mathbb{N}, x \geqslant t\}$ evaluated at point n, and $\mathscr{M}_k$ 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

$$P\{\mathbb{D}_k = \mathscr{D}_k \mid N = n\} = \lambda(n, \mathscr{M}_k, t)\mathscr{I}_A(t)\left\{\prod_{j=2}^{k}\binom{t_{j-1}}{m_j - u_j}\right\} \div \left\{\prod_{j=1}^{k}(u_j)!\right\}, \tag{2.1}$$

$$P\{T_k = t \mid N = n\} = \lambda(n, \mathscr{M}_k, t)\mathscr{I}_A(t)\sum_{i=0}^{t}\frac{(-1)^{t-i}}{i!\,(t-i)!}\prod_{j=1}^{k}\binom{i}{m_j},$$

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

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

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,

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

of the posterior mean or Bayes' estimator,

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

and of the posterior variance or Bayes' risk,

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

Note that π, β and ρ also depend on $\mathcal{M}_k$. In the sequel, all the functions depending on k also should depend on $\mathcal{M}_k$ and, following this rule, we write $\lambda(n, k, t)$ for $\lambda(n, \mathcal{M}_k, t)$. 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 π be a prior probability function for N and let

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

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

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

where $\mathcal{I}_t^\pi(n)$ 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)

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

one may easily prove that $\kappa(k, t)$ is positive and bounded. Note that $M_k \leq t \leq S_k$ 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, $\{n \in \mathbb{N}; n \geq t\}$, 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 $t \leq k - 2$.

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 $\mathbb{N}_t^\pi = \{x \in \mathbb{N}; x \geq t, \pi(x) > 0\}$ 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 $M_k \leq t \leq S_k$ 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 π has a second moment, $\beta(k, t)$ is finite. For this one-by-one sampling, with $t \leq k$ 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 θ = 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

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

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 π is said to be degenerate if its support has only one point. The following results show that β 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, π, 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

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

which completes the proof. □

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

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

Equality holds if the prior probability function, π, 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 ρ 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 ρ, unlike β, 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 ρ and β 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, $\{m_j\}_{j\geq 1}$ is a bounded sequence of elements of $\mathbb{N}$ with $M = \max\{m_j; j \geq 1\}$. As before, m_j is the size of the jth sample and, for all $t \in \mathbb{N}$ such that $t \geq M$ and $\mathbb{N}_t^\pi \neq \varnothing$, define $s = \min\{j \in \mathbb{N}, j \geq 2 \text{ and } S_j \geq t\}$. For all $k \geq s$, we have that $M_k \leq t \leq S_k$ and consequently both $\beta(k, t)$ and $\rho(k, t)$ are well defined. From Theorem 1 and the fact that $\beta(k, t) \geq 1$ for all $k \geq s$, the sequence $\{\beta(k, t)\}_{k\geq s}$ 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 $t \geqslant M$, $\mathbb{N}_t^\pi \neq \emptyset$, 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, $\{\rho(k, t)\}_{k \geqslant s}$, to zero is stated next.

THEOREM 4. *For a fixed $t \in \mathbb{N}$, if $t \geqslant M$, and $\mathbb{N}_t^\pi \neq \emptyset$, 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 $k \geqslant s$ and the result holds. If $\pi(n \mid s, t)$ is not degenerate then, for all $k \geqslant s$, $\pi(n \mid k, t)$ 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 $M_k \leqslant t \leqslant S_k$, $\beta(k, t+1)$ 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 $\{P\{T_k = t \mid N = n\}\}_{k \geqslant 1}$, 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 $\{P\{T_k = t \mid N = n\}\}_{k \geqslant 1}$, 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 ∞ 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 $\mathcal{F}$ be the σ-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 $(\Omega, \mathcal{F})$ 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 $(\Omega, \mathcal{F})$, defined for all positive integers as

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

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

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

The triplet $(\Omega, \mathcal{F}, \{P_n; n \in \mathbb{N}_M\})$ is the *Statistical Space* or *Statistical Model.* Consider the σ-algebra $\mathcal{E}$ of subsets of $\mathbb{N}_M$ and a probability distribution π on $(\mathbb{N}_M, \mathcal{E})$. The probability space $(\mathbb{N}_M, \mathcal{E}, \pi)$ 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 σ-algebra, $\mathcal{F}^*$, containing the set of cartesian products of elements of $\mathcal{E}$ times the elements of $\mathcal{F}$, that is

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

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

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

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

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

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

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 $\mathscr{F}_k$ and we write, for all $k>1$,

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

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 β_k and ρ_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 Π.

LEMMA 2. *We have that:*

(vi) $\{\beta_k\}_{k\geq 1}$ *is a martingale relative to* $\{\mathscr{F}_k\}_{k\geq 1}$;
(vii) $\{\rho_k\}_{k\geq 1}$ *is a supermartingale relative to* $\{\mathscr{F}_k\}_{k\geq 1}$;
(viii) $\{\beta_k\}_{k\geq 1}$ *converges almost surely* $[\Pi]$ *to a random variable defined on* $(\Omega^*, \mathscr{F}^*, \Pi)$;
(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* $\text{Var}\{\beta_k\} \uparrow \text{Var}\{N\}$ *as* $k \to \infty$.

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

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

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 $\lim \text{Var}\{\beta_k\} \geq \text{Var}\{N\}$ 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

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

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 $k \geq t+1$,

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

Consequently,

$$a \leq \tau^t \sum_{n=\tau+1}^{\infty} (\tau/n)^{k-t} \frac{(n!)\pi(n)}{(n)^t(n-t)!} \leq \tau^t \sum_{n=\tau+1}^{\infty} \pi(n)(\tau/n)^{k-t}$$

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

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 $\{m_j\}_{j\geq 1}$), such that $n \leq S_k$ for every $k \geq k_1$. Then, for all $k \geq k_1$

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

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

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

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

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

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

where $\mathcal{F}_\infty$ is the smallest σ-algebra containing

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

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

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

Then

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

concluding the proof of item (x).

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

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

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

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

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

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

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

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

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]

COMMUN. STATIST. -SIMULA., 17(2), 407-430 (1988)

A BAYESIAN ANALYSIS IN CLOSED ANIMAL POPULATIONS FROM CAPTURE RECAPTURE EXPERIMENTS WITH TRAP RESPONSE

Josemar Rodrigues, Heleno Bolfarine
and José Galvão Leite

Instituto de Matemática e Estatística
Universidade de São Paulo
Caixa Postal 20570 - Agência iguatemi
01051 São Paulo, SP. - Brasil

Key Words and Phrases: Capture-recapture process, closed animal population, Bayes factor, Bayes estimator.

ABSTRACT

Using data obtained by a multiple capture-recapture process with trap response, the problem of model choice and inference for the size of a closed animal population is considered from a Bayesian viewpoint. Four different models are discussed and for some estimators and tests developed there are no competitors from the classical sampling approach. A variety of prior structures are considered with the purpose of studying the influence of the priors chosen on the posterior distribution. A special prior structure which takes into consideration the possible correlation between capture and recapture probabilities is also analyzed

1. INTRODUCTION

The problem of estimating the size of a closed animal population has recently been considered by Pollock and Otto (1983) from a classical sampling viewpoint and by Castledine (1981) using a Bayesian approach. Pollock and Otto (1983) consider the capture probabilities to be constant over sampling times, except as influenced by trap response. Castledine (1981) considers the capture probabilities to be the same for all animals but with the possibility of changing over time. However, he has not considered the possibility of trap response. In this paper we follow Castledine's approach but allowing for trap response. According to Pollock and Otto (1983), this problem is very difficult to deal with statistically. Two basic models are considered and Bayesian tests for checking their adequacy are proposed. In each case, Bayesian estimators for the population size are considered. A sensitive study of the posterior distribution for a special class of prior distributions is performed. It is also considered a prior structure which allows the possibility of dependency between capture and recapture probabilities.

2. THE MODELS AND BASIC NOTATION

As in Castledine (1981), we write N for the unknown population size, $s(\geq 2)$ for number of samples taken, p_i, $1 \leq i \leq s$, for the probability of each animal to be captured in the ith sample, c_i, $2 \leq i \leq s$, for the probability of an animal to be recaptured in the ith sample, X_i for the number of unmarked animals in the ith sample, Y_i, for the number of marked animals in the ith sample ($Y_1 = 0$), and M_i for the number of marked animals just before the ith sample ($M_1 = 0$). We note that $M_{i+1} = M_i + X_i = \sum_{j=1}^{i} X_j$, $i = 1, \ldots, s$. Let $n_i = x_i + y_i$, $i = 1, \ldots, s$. It is assumed that the population remains closed throughout the realization of the experiment. The case where $p_i \neq c_i$, for some i, was called by Pollock (1975), the "heterogeneity with trap response" model, and it is here denoted by M^*_{1h}; if $p_i = c_i$, then the "heterogeneity with nontrap response" model, M^*_{oh}, follows. This situation is considered in Castledine (1981). The case where $p_i = p$, $c_i = c$ and $p \neq c$, was called by Pollock and Otto (1983),the "no heterogeneity with trap response" model, here denoted by M^*_1, and if $p = c$, the "no heterogeneity, no trap response" model, M^*_o, follows.

Under the assumptions made above, it follows that

$$X_i \mid p_i \sim B(N - M_i, p_i;\ x_i),\ i=1,\ldots,s \tag{1}$$

and

$$Y_i \mid c_i \sim B(M_i, c_i;\ y_i),\ i= 2,\ldots,s \tag{2}$$

where $B(n,p;x)$ stands for the probability of x successes in n trials of a binomial experiment with success probability p. Similar results hold for the case where $p_i = p$ and $c_i = c$. It is also assumed that X_i is independent of Y_i (conditional on M_i).

To complete the notation let $\underset{\sim}{p} = (p_1,\ldots,p_s)$, $\underset{\sim}{c} = (c_2,\ldots,c_s)$ and $D = \{(x_i,y_i),\ i = 1,\ldots,s;\ y_1 = 0\}$, the observed data. From (1) and (2) and the assumptions made above, it follows easily that the likelihood function is

$$L(N,\underset{\sim}{p},\underset{\sim}{c} \mid D)\alpha\binom{N}{x_1}p_1^{x_1}(1-p_1)^{N-x_1}\prod_{i=2}^{s}\binom{N-M_i}{x_i}\binom{M_i}{y_i}\ .$$

$$\cdot\ p_i^{x_i}(1-p_i)^{N-M_i-x_i}c_i^{y_i}(1-c_i)^{M_i-y_i}\ , \qquad (3)$$

with the restriction that $N \geq M_{s+1}$. Note that if $p_i = c_i$, for all i, then $L(N,\underset{\sim}{p},\underset{\sim}{c} \mid D)$ reduces to (2) in Castledine (1981). If $s = 2$, then $L(N,\underset{\sim}{p},\underset{\sim}{c} \mid D)$ is the model considered by Pollock (1975). It is easily seen that the above maximum of the likelihood function (3) occurs at the solution of the equations

$$\hat{p}_i = x_i/(\hat{N} - M_i),\ i = 1,\ldots,s\ ,$$

$$\hat{c}_i = y_i/M_i\ ,\ i = 2,\ldots,s\ ,$$

and

$$\hat{N} = M_{s+1}/(1 - \prod_{i=1}^{s}(1 - \hat{p}_i)).$$

The equations above were found to be redundant by Pollock (1975) and the reason is that N is not an identifiable parameter for this model. In order to obtain the maximum likelihood estimator, Pollock had to make some restrictions on the parametric space. As will be seen, a Bayesian estimator for N can be obtained without any restriction and with minimal prior information.

In the case of model M_1^*, it follows from (3) that

$$L(N,p,c \mid D) \alpha \binom{N}{M_{s+1}} p^{M_{s+1}} (1-p)^{N_s - \sum_{i=1}^{s+1} M_i} c^{\sum_{i=1}^{s} y_i} .$$

$$. (1-c)^{\sum_{i=1}^{s} M_i - \sum_{i=1}^{s} y_i} \text{, where } N \geq M_{s+1}, \quad (4)$$

and the maximum likelihood estimators of p,c and N are given by the solution of

$$\hat{p} = M_{s+1} (s\hat{N} - \sum_{i=1}^{s} M_i) ,$$

$$\hat{c} = \sum_{i=1}^{s} y_i / (\sum_{i=1}^{s} y_i + \sum_{i=1}^{s} M_i)$$

and

$$\hat{N} = M_{s+1} / (1-(1-\hat{p})^s).$$

These equations have also been obtained by Pollock and Otto (1983).

3. A BAYESIAN ANALYSIS FOR THE "HETEROGENEITY WITH TRAP RESPONSE" MODEL

In this section, we first consider, from a Bayesian view point, a method for comparing the two alternative models M_{oh}^* and M_{1h}^*. Then, having decided for one of M_{oh}^*

or M^*_{1h}, we consider the problem of deriving Bayesian estimators for N.

3.1. Deciding between M^*_{oh} and M^*_{1h}

To decide between M^*_{oh} or M^*_{1h} we make use of the Bayesian factor approach (see Smith and Spiegelhalter (1980), for example). Denote by

$$K = \frac{\pi_o p(D|M^*_{oh})}{\pi_1 p(D|M^*_{1h})} \qquad (5)$$

the Bayes factor in favor of M^*_{oh}, where π_i denotes the prior weight attached to model M^*_{ih}, i = 0,1 and $p(D|M^*_{ih})$ is the predictive density under model M^*_{ih}, i = 0, 1. In order to provide a Bayesian solution to the estimation of N and a model choice procedure for selecting between M^*_{1h} and M^*_{oh}, it is considered in this and in the next section, the following prior structure:

$$\pi(N,\underset{\sim}{p},\underset{\sim}{c}) = \pi(N)\,\pi(\underset{\sim}{p},\underset{\sim}{c}),$$

where $\pi(N)$ is a constant; $\pi(\underset{\sim}{p},\underset{\sim}{c}) = 1$, $0 \leq p_i \leq 1$, $0 \leq c_i \leq 1$, $i = 2,\ldots,s$; $\pi(p_1) = 1$, $0 \leq p_1 \leq 1$ and $\pi_i = 0{,}5$, $i = 0,1$. The above prior structure is reasonable when little prior information is available about the parameters of the model. It also makes it possible to find a Bayesian solution (with non-informative priors) for the problem of estimating N, not completely solved by Pollock (1975). Informative priors which also avoids the ridge between N, $\underset{\sim}{p}$ and $\underset{\sim}{c}$ in (3) and the question of influence on the posterior distribution of a special class of prior distributions, are studied in detail in Section 5.

The predictive density under model M^*_{oh}, $p(D|M^*_{oh})$, is given by

$$p(D|M^*_{oh}) = \sum_{N\geq 1} \int_0^1\int_0^1 L(N,\underset{\sim}{p},\underset{\sim}{c})\Pi(N)\Pi(\underset{\sim}{p},\underset{\sim}{c})\; d\underset{\sim}{p}\; d\underset{\sim}{c}$$

$$= \prod_{i=1}^{s} \int_0^1\int_0^1 c_i^{y_i} (1-c_i)^{M_i-y_i}\; p_i^{x_i} (1-p_i)^{M_{s+1}-M_i-x_i} \cdot$$

$$\cdot \left[1 - \prod_{i=1}^{s} (1-p_i)\right]^{-(M_{s+1}+1)} dp_i dc_i$$

$$\doteq \prod_{i=1}^{s} \int_0^1 p_i^{n_i} (1-p_i)^{M_{s+1}-n_i} \sum_{j=0}^{\infty} \frac{(M_{s+1})^j}{j!} \prod_{i=1}^{s} (1-p_i)^j dp_i$$

$$= \sum_{j=0}^{\infty} \frac{(M_{s+1}+1)^j}{j!} \prod_{i=1}^{s} \frac{n_i!(M_{s+1}-n_i+j)!}{(M_{s+1}+j+1)!}$$

$$= \sum_{j=0}^{\infty} \frac{(M_{s+1}+1)^j}{j!} \prod_{i=1}^{s} \frac{1}{\binom{M_{s+1}+j+1}{n_i+1}(n_i+1)},$$

where $\doteq$ means that the following aproximation was used:

$$\left(1 - \prod_{i=1}^{s} (1-p_i)\right)^{-(M_{s+1}+1)} \doteq \sum_{j=0}^{\infty} \frac{(M_{s+1}+1)^j}{j!} \prod_{i=1}^{s} (1-p_i)^j,$$

which is justified by the fact that, for small t, $\ln(1-t) \doteq -t$.

Similarly, under model M^*_{1h}, it can be shown that the predictive density is

$$p(D|M^*_{1h}) \doteq \sum_{j=0}^{\infty} \frac{(M_{s+1}+1)^j}{j!} \prod_{i=1}^{s} \frac{1}{\binom{M_i}{y_i}\binom{M_{s+1}-M_i+j+1}{y_i+1}(x_i+1)} .$$

Accordingly, the Bayes factor in favor of M^*_{oh} is:

$$K \doteq \frac{\displaystyle\sum_{j=0}^{\infty} \frac{(M_{s+1}+1)^j}{j!} \prod_{i=1}^{s} \frac{1}{\binom{M_{s+1}+j+1}{n_i+1}(n_i+1)}}{\displaystyle\sum_{j=0}^{\infty} \frac{(M_{s+1}+1)^j}{j!} \prod_{i=1}^{s} \frac{1}{\binom{M_i}{y_i}\binom{M_{s+1}+j+1-M_i}{x_1+1}(x_i+1)}}$$

$$= \frac{\displaystyle\sum_{j=0}^{\infty} \frac{(M_{s+1}+1)^j}{j!} \prod_{i=1}^{s} \frac{1}{\binom{M_{s+1}+j+1}{n_i+1}}}{\displaystyle\sum_{j=0}^{\infty} \frac{(M_{s+1}+1)^j}{j!} \prod_{i=1}^{s} \frac{1}{\binom{M_i}{y_i}\binom{M_{s+1}-M_i+j+1}{x_i+1}}} . \tag{9}$$

If s is reasonably large, it is enough to consider $j=0$ in (9), since the other terms in the summation will be extremely small. In this case, we may write (9) as approximately equal ($\doteq$) to

$$K \doteq \prod_{i=1}^{s} \frac{\binom{M_i}{y_i}\binom{M_{s+1}-M_i+1}{x_i+1}}{\binom{M_{s+1}+1}{n_i+1}} \cdot \frac{(x_i+1)(M_i+1)}{(n_i+1)} . \qquad (10)$$

The first term on the right hand side of (10) appears to be an extension of the Fisher exact test statistic for comparing two binomials (see Lehmann (1986), pp.143). The second term would then correspond to a contribution from the Bayesian approach. If a binomial approximation is used for the hypergeometric distribution, an approximation for K in (10) is

$$K \doteq \prod_{i=1}^{s} B\left(n_i+1, \frac{M_{s+1}-M_i+1}{M_{s+1}+1} ; x_i+1\right) . \frac{(x_i+1)(M_i+1)}{(n_i+1)} . \qquad (11)$$

3.2. Making inferences about N

In this section, the posterior distribution of N is found under models M^*_{oh} and M^*_{1h} and the prior structure considered in Section 3.1.

Under M^*_{1h} we may write, according to (3),

$$\Pi(N,\underset{\sim}{p},\underset{\sim}{c}|D) \propto \binom{N}{M_{s+1}} p_1^{x_1} (1-p_1)^{N-x_1} . \qquad (12)$$
$$\cdot \prod_{i=2}^{s} p_i^{x_i} (1-p_i)^{N-M_i-x_i} c_i^{y_i} (1-c_i)^{M_i-y_i} .$$

Integrating out $\underset{\sim}{p}$ and $\underset{\sim}{c}$ in (12) we obtain, after making use of some properties of the beta distribution, that the marginal posterior distribution of N is

$$\pi(N|D) \propto \prod_{i=1}^{s} \frac{1}{(N+1-M_i)} \tag{13}$$

with $N \geq M_{s+1}$. It is easily seen that the mode of the posterior distribution given by (13) is M_{s+1}. As mentioned in Section 2, under M^*_{1h} and the noninformative prior structure considered in Section 3.1, M_{s+1} is the Bayesian solution for the unsolved estimation problem discussed in Pollock (1975). It is also worth remarking that if one considers the improper prior distribution $\pi(p,c) \propto \prod_{i=1}^{s} p_i^{-1}$, then $\pi(N|D) = \pi(N)$, that is, the data is noninformative for making inferences about N. This gives a formal justification for why it is not possible to estimate N without a proper prior distribution for $\underset{\sim}{p}$ and $\underset{\sim}{c}$. Castledine (1981), used the improper prior above to show the correspondence between his and Darroch's (1958) approach.

According to the prior structure considered in Section 3.1, we obtain under model M^*_{oh}

$$\pi(N,p|D) \propto (M_{s+1})^N \prod_{i=1}^{s} p_i^{n_i}(1-p_i)^{N-n_i}. \tag{14}$$

where, as before, $n_i = x_i + y_i$, $i = 1,\ldots,s$, and $N \geq N_{s+1}$. So, it is easily checked that

$$\pi(N|D) \propto \frac{(M_{s+1})^N}{(N+1)^s} \prod_{i=1}^{s} (N-n_i)! \quad ,$$

where $N > M_{s+1}$. Expression (15) can also be obtained using (7) in Castledine (1981), although he did not consider the possibility of using noninformative, but proper priors for $\underset{\sim}{p}$ and $\underset{\sim}{c}$.

Analytical expressions for the posterior mode of the probability function (15) are considered in Leite et al. (1987).

Application 1. In this application the sunfish data considered in Castledine (1981) is reanalyzed. Using the binomial approximation (11) it can be shown, after some numerical computations, that $K = 3.498 \times 10^{-5}$. So, $P(M^*_{oh}|D) = 3.489 \times 10^{-5}$. This shows that M^*_{oh} is not an adequate model, since it has extremely low support from the data. It is reasonable to base inferences about N on M^*_{1h}. Under M^*_{1h}, the posterior distribution (13) practically degenerates into $M_{s+1} = 135$. So, clearly, $\hat{N} = 135$ is the Bayesian estimator for N under the prior structure of Section 3.1.

Application 2. In this application we consider the data of Example 4.8 in Seber (1973), pp. 146. After some numerical computations, it can be shown that $K = 0.98$ and that $P(M^*_{oh}|D) = 0.489$. Thus, inferences about N could be based either on M^*_{oh} or M^*_{1h}. It follows that the mode of the posterior probability function (15) is $\hat{N} = 58$. The 95% credibility interval for N is (56;64).

4. A BAYESIAN ANALYSIS OF THE "NON HETEROGENEITY WITH TRAP RESPONSE" MODEL

In this section a method for comparing the two alternative models, M_o^*, the "non heterogeneity with non-trap response" model, with M_1^*, the "non heterogeneity with trap response" model is considered. The Bayesian factor approach considered in Section 3.1 could be used, but we proposed another method that provides, in this case, a more elaborated solution.

4.1. Deciding between M_o^* and M_1^*

From (3), considering the prior structure of Section 3.1, it is not hard to show, by using (7), that the marginal posterior distributions of c and p are given by

$$\Pi(c|D) \; \alpha \; c^{\sum_{i=2}^{s} y_i} \, (1-c)^{\sum_{i=2}^{s} M_i - \sum_{i=1}^{s} y_i}$$

and

$$\Pi(p|D) \; \alpha \; p^{\sum_{i=1}^{s} x_i} (1-p)^{sM_{s+1} - \sum_{i=1} M_i + 1} (1-(1-p)^s)^{-(M_{s+1}+1)},$$

respectively.

It is well known that the posterior distribution of $Z = \ell og\,(c/1-c))$, given D, is approximately normal with mean

$$\ell_1 = \ell og((\sum_{i=2}^{s} y_i + 1)/(\sum_{i=1}^{s} M_i - \sum_{i=2}^{s} y_i + 1)), \text{ and}$$

variance

$$v_1 = 1/(\sum_{i=2}^{s} y_i + 1) + 1/(\sum_{i=1}^{s} M_i - \sum_{i=2}^{s} y_i + 1).$$

It is also assumed that given D, $W = \log(\frac{p}{(1-p)})$ is approximately normal with mean ℓ_2, which is obtainable by solving the equation

$$(M_{s+1}+1)\frac{1}{t} + \frac{sM_{s+1}}{(1+t)^s - 1} - sM_{s+1} - \sum_{i=1}^{s} M_{i+1} = 0,$$

for $t.W = \log t$. The variance of the posterior distribution is

$$v_2 = \frac{1}{((s+1)M_{s+1} - \sum_{i=1}^{s} M_{i+1})\frac{\hat{t}}{(1+\hat{t})^2} + \frac{\hat{t}s(M_{s+1}+1)\left[(1+\hat{t})^s(1-s\hat{t})-1\right]}{\{\left[(1+\hat{t})^s-1\right](1+\hat{t})\}^2}}$$

the inverse of the second derivative of the logarithm of the posterior density of W at the mode ℓ_2, where $\hat{t} = e^{\ell_2}$. As will be seen in the applications, these approximations seem are quite reasonable. Using Hinkley (1969), it follows that the posterior distribution of $q = W/Z$ is proportional to

$$e^{-\frac{\ell_1/v_1+\ell_2/v_2}{2} + \frac{u^2}{\sigma^2}} \{2\Pi\ \mu\ \sigma\left[1-Z\Phi(\frac{\mu}{\sigma})\right] + 2\sigma^2 e^{-\mu^2/2\sigma^2}\}, \tag{16}$$

where Φ is the distribution function of the standartized normal distribution,

$$\mu = \frac{\ell_1 q/v_1 + \ell_2/v_2}{q^2/\ell_1 + 1/v_2} \quad \text{and} \quad \sigma^2 = \frac{1}{q^2/v_1 + 1/v_2} .$$

So, the decision for one of M_o^* or M_1^* might be based on a (1-α) 100% HPD interval for q; if q = 1, is in the interval then M_o^* is the model to use for making inferences about N. Otherwise, M_1^* should be used.

4.2. Making inferences about N

Under the prior structure considered in Section 3.1, it follows from (4) that the posterior distribution for N,p and c, under M_1 is

$$\Pi(N,p,c|D) \; \alpha \; L(N,P,c|D). \tag{17}$$

By intergrating out p and c in (17) it follows that the marginal posterior distribution of N is

$$\Pi(N|D) \; \alpha \; \binom{N}{M_{s+1}} \frac{(Ns - \sum_{i=1}^{s} M_{i+1})!}{(Ns - \sum_{i=1}^{s} M_i + 1)!} \tag{18}$$

$N \geq M_{s+1}$. All the information on N can be obtained from this posterior p.d.f.

On the other hand, under M_o^*, the posterior distribution of N is

$$\Pi(N|D) \; \alpha \; \binom{N}{M_{s+1}} \frac{(Ns - M_{s+1} - \sum_{i=1}^{s} y_i)!}{(Ns-1)!} \qquad (19)$$

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

Application 3. In this application, the sunfish data of Application 1 is reanalyzed. We first decide, based on the approach presented in Section 4.1, which model to use, M_o^* or M_1^*. The normal approximations for the posterior of $Z = \ell og\ c/(1-c)$ and $W = \ell og\ p/(1-p)$, are such that

$\ell_1 = -3.6$; $v_1 = 0.042$; $\ell_2 = -2.6$ and

$v_2 = 0.077$.

In order to see that these approximations are reasonable, a plot of the exact densities of Z and W are given in Figure 1.

Figure 2 below presents a plot of the posterior density of $q = W/Z$, which is given in (16).

The 95% HPD interval for q is (0.59; 0.85) and $q = 1.0$ is not included; neither it is included in the 99% HPD interval. So, the right model to choose is M_1^*. Under this model, a plot of the posterior probability function of N (from (18)) is given in Figure 3 below. The posterior mode is given by 238. The 95% HPD confidence interval for N is (197;546).

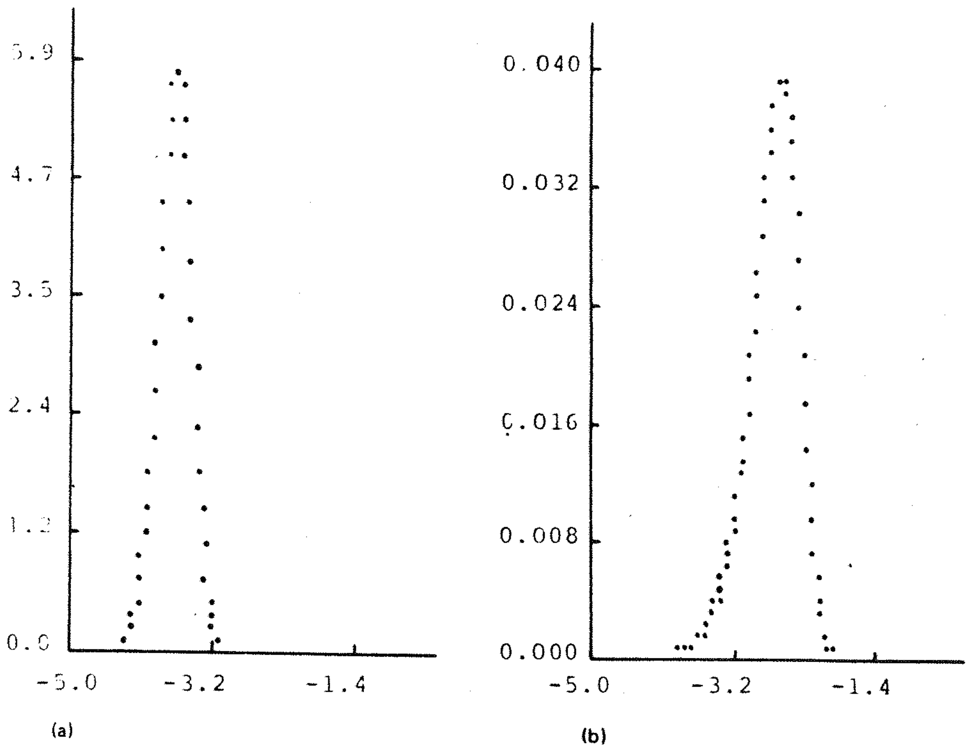

FIG. 1 - Posterior densities. (a) $Z|D$; (b) $W|D$.

Application 4. A second analysis of the data presented in Application 2 is now performed. The normal approximations for Z and W are such that

$$\ell_1 = -0.26; \; v_1 = 0.035; \; \ell_2 = -3.02; \; v_2 = 3.11.$$

The 95% HPD interval for $q = W/Z$ is (0.1;23.5). So, under M_o^*, the posterior mode is $\hat{N} = 37$, and the 95% HPD interval is (36;44).

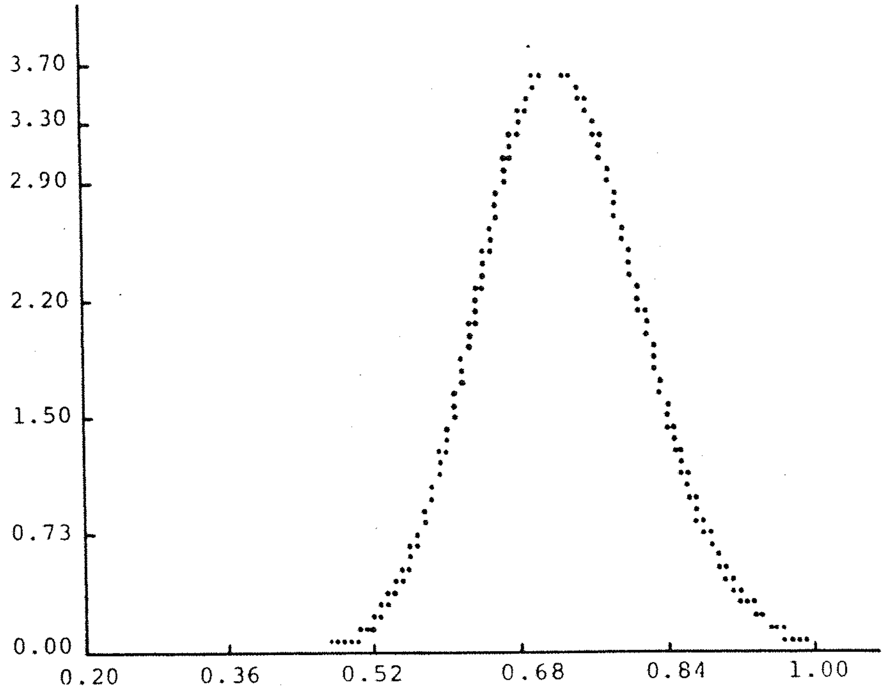

FIG. 2 - Posterior density of q = W/Z.

5. INFORMATIVE PRIORS

In the previous sections, uniform prior distribution was taken for $(\underline{p},\underline{c})$. But in many situations, informative prior knowledge is avaiable about $(\underline{p},\underline{c})$. Thus, it is interesting to investigate the sensitivity of the posterior distribution of N, to changes in the prior structure and the model adopted. In this section, two informative prior structures are investigated.

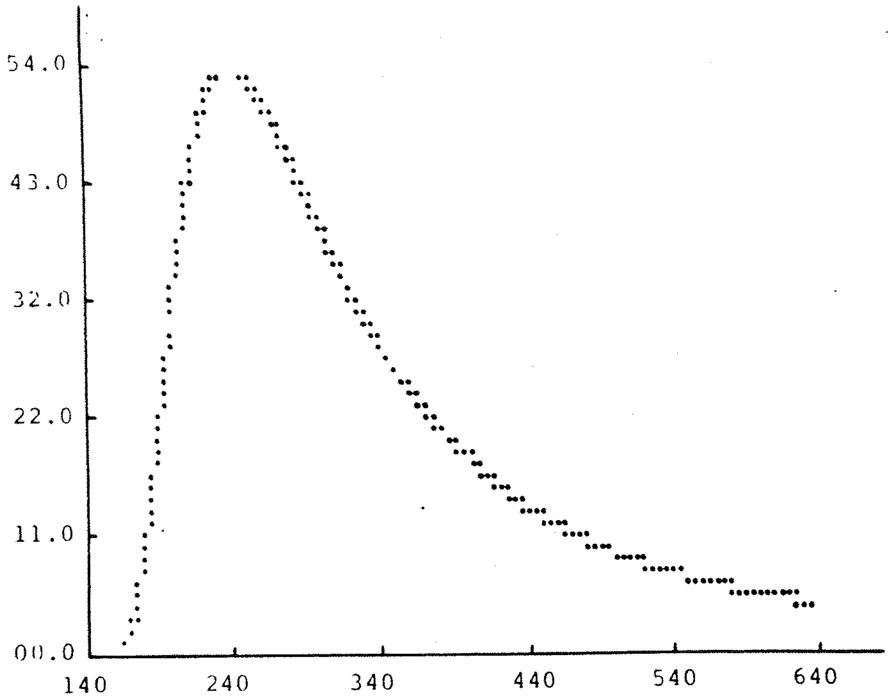

FIG. 3 - Posterior probability function for N under M_1^*.

5.1. Independent Beta priors

If $\underset{\sim}{p}$ and $\underset{\sim}{c}$ are considered to be prior independent, it is natural to consider Beta priors. Therefore, under M_1^*, let

$$\begin{aligned} p_i &= p \sim \text{Beta}(a_1, b_1) \ ; \ i = 1, \ldots, s \ , \\ c_i &= c \sim \text{Beta}(a_2, b_2) \ ; \ i = 2, \ldots, s \ , \end{aligned} \tag{20}$$

with p and c independent. Under M_{1h}^*, let

$$p_i \sim \text{Beta}(a_1,b_1) \ , \ i = 1,\ldots,s$$
$$c_i \sim \text{Beta}(a_2,b_2) \ , \ i = 2,\ldots,s \ , \qquad (21)$$

with p_i and c_i independent, $i = 2,\ldots,s$.

Combining (20) and (21) with (4) and (3), respectively, and using the same approach as in Castledine (1981), we have

$$M_1^*: \Pi(N|D) \alpha \binom{N}{M_{s+1}} \frac{(sN - \sum_{i=1}^{s} M_i - \sum_{i=1}^{s} x_i + b_1 - 1)!}{(sN - \sum_{i=1}^{s} M_i + a_1 + b_1 - 1)!} \Pi(N), \qquad (22)$$

and

$$M_{1h}^*: \Pi(N|D) \alpha \binom{N}{M_{s+1}} \prod_{i=1}^{s} \frac{(N-M_i-x_i+b_1-1)!}{(N-M_i+a_1+b_1-1)!} \Pi(N). \qquad (23)$$

Note that for $a_1=b_1=1$, under model M_{1h}^*, we obtain (13). Also, for $a_1=0$, $b_1=1$, it follows that $\Pi(N|D)=\Pi(N)$. In order to investigate the sensitivity of the posterior distribution of N to changes in the prior structure, as in Castledine (1981), let $\lambda = \Pi(N|D)/\Pi(N)$. Applying Stirling's formula to (22) and (23), we obtain

$$M_1^*: \ln\lambda = \text{const} + (M_{s+1}+a_1)\ln sN + \frac{1}{N}\left\{ \frac{(\sum_{i=1}^{s} M_{i+1}-b_1)(\sum_{i=1}^{s} M_{i+1}-b_1+1)}{s} \right.$$
$$\left. - \frac{(a_1+b_1)(a_1+b_1-1)}{s} + \frac{0.5(M_{s+1}-a_1)}{s} - \frac{\sum_{i=1}^{s} M_i(\sum_{i=1}^{s} M_i + 1)}{s} \right\},$$

and

$$M_{1h}^*: \ln\lambda = \text{const} - (M_{s+1}+sa_1)\ln N - \frac{1}{N}\{2b_1 \sum_{i=1}^{s} M_{i+1} +$$

$$+ s\left[(a_1+b_1)(a_1+b_1-1)-b_1^2 + 0.5(a_1+2b_1)\right]\}.$$

As the number of stages s increases, the dominant terms in M_1^* are M_{s+1} and $\sum_{i=1}^{s} M_{i+1}$. Therefore, the influence of the prior parameters a_1 and b_1 is much greater in model M_{1h}^* than in model M_1^*. Similar result was obtained by Castledine (1981) when the trap response is not considered in the estimation process. As in Castledine (1981), this difference is explained as follows. In model M_1^*, $\bar{p} = \frac{1}{s}\sum_{i=1}^{s} p_i$ can be learned from the experiment an by knowing this, we can form an intuitive estimate,

$$\hat{N} = \frac{1}{\bar{p}s}\sum_{i=1}^{s} x_i + M_{s+1} \quad , \tag{24}$$

whereas in model M_{1h}^*, the value of $\bar{p}$ is known a priori to be $a_1/(a_1+b_1)$, that is, we cannot learn it from the experiment. Note, also, that the effect of changing a_2 and b_2 is null in both models.

Application 5. Table I shows the mode and the 95% HPD intervals of the distribution of N under models M_1^* and M_{1h}^* based on the sunfish population of Application 1. As before, $\Pi(N) \propto 1/N$, $N \geq 1$.

Table I. Posterior Distribution of N for the Sunfish Data.

Models	Prior Parameters a_1	b_1	Mode	95% HPD Interval
M_1^*	1	1	238	(197, 546)
M_1^*	3	100	290	(215, 594)
	3	500	1064	(523, 1124)
	15	500	356	(267, 607)
	30	1000	374	(292, 546)
M_{1h}^*	3	100	385	(305, 535)
	3	500	1625	(1157,1628)
	15	500	393	(333, 474)
	30	1000	394	(338, 465)
Π (Castledine,1981)	3	100	-	(335, 539)
	15	500	-	(347, 489)
	30	1000	-	(350, 478)

5.2. Diriclet Priors

In the previous sections, independent prior distributions were taken for p_i and c_i, i=1,...,s. However, this assumption is not realistic in many practical situations. A prior distribution that takes into consideration possible negative correlation between p_i and c_i is the Diriclet distribution. Therefore, in this section, it is considered that

$$M_{1h}^*: (p_i, c_i) \sim D[\underset{\sim}{a}, a_o] \quad , \quad i=1,\ldots,s \quad , \tag{25}$$

and

$$p_1 \sim B[a_1, a_o+a_2] \quad ,$$

where $\underset{\sim}{a} = (a_1, a_2)$.

Table II. Posterior Distribution of N from the Sunfish Data

Prior Parameters a_o	a_1	a_2	Mode	95% HPD Interval
1	1	1	158	(143 ; 214)
1	1	10	217	(176 ; 348)
10	10	10	139	(135 ; 146)
3	10	100	190	(170 ; 219)

Considering the transformation

$$\Theta_{i1} = p_i \quad \text{and} \quad \Theta_{i2} = \frac{c_i}{1-p_i} ,$$

$i=1,\dots,s$, it is not difficult to see that

(i) Θ_{i1} and Θ_{i2} are independent, $i=2,\dots,s$;

(ii) $\Theta_{i1} \sim B[a_1, a_o+a_2]$ and $\Theta_{i2} \sim B[a_2, a_o+a_1]$.

From (3),(25) and $\Pi(N) \alpha 1/N$, it follows, by integrating out Θ_{i1}, Θ_{i2} that the marginal posterior distribution of N is

$$\Pi(N|D)\alpha \binom{N}{M_{s+1}} \prod_{i=1}^{s} \frac{(N-M_{i+1}+y_i+a_o+a_2-1)!}{(N-M_i+y_i+a_o+a_1+a_2-1)!} \Bigg[1 + \sum_{j=1}^{M_i-y_i} (-1)^j \binom{M_i-y_i}{j} \frac{[N-M_{i+1}+y_i+a_o+a_2+j-1]_j}{[N-M_i+y_i+a_o+a_1+a_2-1]_j} \Bigg] \Pi(N), \quad (26)$$

where $[a]_j = a(a-1)\dots(a-j+1)$. If c_i is small, which is reasonable in many real problems, we can obtain a simplified and approximate version of (26) by considering only the first (or the first two) term in the summation involved.

Application 6. Table II shows the posterior mode and the 95 HPD interval, to illustrate the behavior of the posterior (26) for several values of a_o, a_1 and a_2. The data set employed is the one considered in Application 1.

Acknowledgement: The authors gratefully acknowledge the encouragement of Prof. S. Zacks and the helpful comments of the referee.

BIBLIOGRAPHY

Castledine, B. (1981). "A Bayesian Analysis of the Multiple-Recapture Sampling for a Closed Population", Biometrika, 67, 1, 197-210.

Darroch, J. (1958). The Multiple Recapture Census I: Estimation of a Closed Population. Biometrika, 45, 343-359.

Lehmann, E.L. (1986). "Testing Statistical Hypothesis". Wiley.

Leite, J.G., Oishi, J., Pereira, C.A.B. (1986). "Exact ML estimate of a finite population size: capture-recapture sequential sample data". Prob. in the Engineering and Informational Sciences, 1, 225-236.

Leite, J.G., Bolfarine, H., Rodrigues, J. (1987). "Exact expression for the posterior mode of a finite population size: Capture-recapture sequential sampling", Bras. J. of Prob. and Stat. 1, 91-100.

Pollock, K. (1975). "Building models of capture-recapture experiments". The Statistician, 25, 253-259.

Pollock, K. and Otto, M. (1983). "Robust estimation of population size in closed animal population from capture-recpature experiments". Biometrika 39, 1035-1049.

Smith, A.F.M. and Spiegelhalter, D. (1980). "Bayes factors and choice criteria for linear models". J. R. Statist. Soc. B. (1980), 42, 213-220.

Seber, G.A.F. (1973). "The estimation of animal abuncande". Griffin London.

Received by Editorial Board member March, 1987; Revised February, 1988.

Recommended by Shelemyahu Zacks, State University of New York at Binghamton.

Refereed Anonymously.