UNIVERSIDADE DE SÃO PAULO

UTILIZAÇÃO DA INFERÊNCIA BAYESIANA EM EXPERIMENTOS DE CAPTURA - RECAPTURA

SILVANA APARECIDA CEREGATO
JOSEMAR RODRIGUES

Nº 31

NOTAS

Instituto de Ciências Matemáticas de São Carlos

Instituto de Ciências Matemáticas de São Carlos

ISSN - 0103-2577

UTILIZAÇÃO DA INFERÊNCIA BAYESIANA EM EXPERIMENTOS DE CAPTURA - RECAPTURA

SILVANA APARECIDA CEREGATO
JOSEMAR RODRIGUES

Nº 31

N O T A S D O I C M S C
Série Estatística

São Carlos
Jun./1996

UTILIZAÇÃO DA INFERÊNCIA BAYESIANA EM EXPERIMENTOS DE CAPTURA-RECAPTURA

Silvana Aparecida CEREGATO*

Josemar RODRIGUES*

- **RESUMO:** O objetivo deste trabalho é propor um ajuste para o estimador de Petersen (1896) via maximização da densidade preditiva, similar ao ajuste de Bartlett (1937). Um conjunto de dados simulados é utilizado para estudar a eficiência deste ajuste.

- **UNITERMOS:** Inferência Bayesiana, Captura-Recaptura, Densidade Preditiva.

1. Introdução

Um problema de grande interesse em Ciências Biológicas é o de estimação do tamanho de uma população animal.

Tal problema pode ser resolvido empregando-se o método de captura-recaptura, descrito da seguinte maneira: duas ou mais amostras são selecionadas da população em estudo. Todos os animais capturados na primeira amostra são marcados. Em seguida, todos os marcados (ou parte deles) são devolvidos à população. Para cada uma das amostras subseqüentes são registradas as capturas dos animais marcados e não marcados. Todos os animais não marcados recebem uma marca e, em alguns casos, os animais anteriormente marcados também recebem uma nova marca, indicando a época de sua captura. Em seguida todos os animais (ou parte deles) são devolvidos à população.

Neste trabalho apresentamos uma forma de estimar o tamanho de uma população animal utilizando o método de captura-recaptura via Inferência Bayesiana Preditiva (ver, por exemplo, O'Hagan, 1994; Press, 1989).Comparamos o estimador obtido com o estimador de Petersen (1896) utilizando um conjunto de dados simulados.

*Instituto de Ciências Matemáticas de São Carlos - USP - CP 668 - 13.560-970 - São Carlos - SP - Brasil

2. Etapas do Experimento de Captura-Recaptura

Suponhamos que exista uma população animal cujo tamanho precisa ser estimado. O processo de captura-recaptura é formulado através das seguintes etapas:

1ª etapa: Supondo que o tamanho da população é N (desconhecido), selecionamos sem reposição, através de um mecanismo aleatório, n_0 animais que são marcados e devolvidos à população;

2ª etapa: Considerando a população animal anterior selecionamos com reposição n_1 animais e observamos, dentre eles, m animais já marcados na primeira etapa. A probabilidade de capturar um animal marcado nesta segunda etapa é dada por $c = \frac{n_0}{\mathrm{N}}$.

Assim, temos que $M|n_1,\, c \sim \text{Binomial}(n_1,\, c)$, portanto,

$$P[M = m|n_1,\, c] = \binom{n_1}{m} c^m (1-c)^{n_1 - m}. \tag{1}$$

A média e a variância de m são, respectivamente:

$$E(M) = n_1 c \qquad\qquad V(M) = n_1 c(1-c).$$

Considerando apenas essas duas etapas iniciais do experimento de captura-recaptura, podemos obter o estimador de Petersen (1896), muito conhecido e utilizado em problemas de estimação do tamanho de uma população animal.

Esse estimador é obtido igualando-se a proporção de marcados na população, representada por $\frac{n_0}{\mathrm{N}}$, à proporção de marcados na amostra, representada por $\frac{m}{n_1}$, ou seja,

$$\hat{\mathrm{N}}_{\mathrm{P}} = \frac{n_0 n_1}{m}. \tag{2}$$

3. Previsão Bayesiana no Experimento de Captura-Recaptura

A fim de empregarmos a previsão Bayesiana (ver, por exemplo, O'Hagan, 1994; Press, 1989) no experimento de captura-recaptura devemos considerar a hipótese de que, ao realizarmos uma terceira etapa de captura na população estudada, a probabilidade de captura de cada animal marcado é dada por $\pi = \frac{n_0}{\mathrm{N}}$, ou seja, $\pi = c$.

Assim, a terceira etapa a ser realizada futuramente pode ser descrita da seguinte forma: supondo que a probabilidade de captura de um animal marcado é π devemos capturar com reposição, dentro da população de tamanho N, um certo número de animais, denotado por n.

Podemos dizer, então, que nesta terceira etapa realizaremos N ensaios de Bernoulli, cuja probabilidade de sucesso é π. Logo, $n|\mathrm{N}, \pi \sim \mathrm{Binomial}(\mathrm{N}, \pi)$, ou seja,

$$P[n = n_0|\mathrm{N}, \pi] = \binom{\mathrm{N}}{n_0} \pi^{n_0} (1-\pi)^{\mathrm{N}-n_0}. \tag{3}$$

A terceira etapa é exatamente a segunda etapa onde substituímos n_1 por N.

Com o objetivo de aplicar o Método Bayesiano para solucionar nosso problema de captura-recaptura, podemos supor que o parâmetro π, que representa a probabilidade de captura de um animal marcado, dentro da população de tamanho N, tem uma distribuição a priori conjugada (Press, 1989), por exemplo uma distribuição $\mathrm{Beta}(a,b)$, onde $a > 0$ e $b > 0$. Assim,

$$p(\pi) = \frac{1}{B(a,b)} \pi^{a-1} (1-\pi)^{b-1}. \tag{4}$$

Os dados incorporados no nosso estudo são provenientes da segunda etapa do experimento de captura-recaptura, ou seja, a função de verossimilhança, $l(m|n_1, \pi)$, é dada pela expressão (1). Logo, combinando as expressões (1) e (4), obtemos a distribuição a posteriori para π, dada por:

$$p(\pi|n_1, m) = \frac{p(\pi) l(m|n_1, \pi)}{\int_0^1 p(\pi) l(m|n_1, \pi) d\pi} = \frac{1}{B(\alpha, \beta)} \pi^{\alpha-1} (1-\pi)^{\beta-1}. \tag{5}$$

onde: $\alpha = a + m$ e $\beta = b + n_1 - m$.

Portanto, os parâmetros da distribuição a posteriori representam os parâmetros da distribuição a priori de π atualizados pelos dados observados na segunda etapa do experimento de captura-recaptura.

Podemos determinar a densidade preditiva (ver, por exemplo, Press, 1989) para a variável aleatória n integrando o produto das expressões (3) e (5) com respeito ao parâmetro π, da seguinte forma:

$$P[n = n_0|\mathrm{N}] = \int_0^1 P[n = n_0|\mathrm{N}, \pi] p(\pi|n_1, m) d\pi.$$

Logo, a densidade preditiva para n corresponde a uma distribuição $\mathrm{Beta-Binomial}(\mathrm{N}, \alpha, \beta)$ (Bernardo & Smith, 1994):

$$P[n = n_0|\mathrm{N}] = \binom{\mathrm{N}}{n_0} \frac{\Gamma(\alpha+\beta)\Gamma(\mathrm{N}-n_0+\beta)\Gamma(n_0+\alpha)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\alpha+\beta+\mathrm{N})}. \tag{6}$$

Nosso objetivo nesta terceira etapa do experimento de captura-recaptura é estimar o tamanho da população animal, N, de modo que a probabilidade de recapturar todos os animais marcados na primeira etapa seja a máxima possível, ou seja, desejamos maximizar $P[n = n_0|\mathrm{N}]$ em relação a N.

Considerando a expressão (6) como uma função de N apresentamos a seguir os valores de N que maximizam $P[n = n_0|\mathrm{N}]$.

Tabela 1: Valores de $\hat{\mathrm{N}}$ que maximizam $P[n = n_0|\mathrm{N}]$

Possíveis valores de β			$\hat{\mathrm{N}}$
$\beta = 1$			n_0
$\beta > 1$	$n_0 \leq \dfrac{\alpha}{\beta - 1}$		n_0
$\beta > 1$	$n_0 > \dfrac{\alpha}{\beta - 1}$	$n_0 - 1 + n_0 \dfrac{\beta - 1}{\alpha} \in Z$	$\left\Vert n_0 - 1 + n_0 \dfrac{\beta - 1}{\alpha} \right\Vert + 1$
$\beta > 1$	$n_0 > \dfrac{\alpha}{\beta - 1}$	$n_0 - 1 + n_0 \dfrac{\beta - 1}{\alpha} \notin Z$	$n_0 + n_0 \dfrac{\beta - 1}{\alpha} = \dfrac{n_0}{\dfrac{\alpha}{\alpha + \beta - 1}}$ (7)

Observação: $\|x\|$ corresponde ao menor inteiro não superior a x e Z é o conjunto dos números inteiros relativos.

Notemos que o estimador $\hat{\mathrm{N}}$ dado em (7) pode ser escrito da seguinte forma:

$$\hat{\mathrm{N}}_{\mathrm{B}} = \mathrm{A}\hat{\mathrm{N}}_{\mathrm{P}} \tag{8}$$

onde $\mathrm{A} = \dfrac{m}{n_1}\left(\dfrac{\alpha + \beta - 1}{\alpha}\right)$.

Concluímos então que o estimador Bayesiano, denotado por $\hat{\mathrm{N}}_{\mathrm{B}}$, que maximiza a densidade preditiva da variável aleatória n corresponde ao estimador de Petersen multiplicado por um ajuste.

Se trabalharmos com uma priori não informativa para π, ou seja, $a = b = 0$, temos que os parâmetros da posteriori de π são $\alpha = m$ e $\beta = n_1 - m$. Neste caso,

$$\hat{\mathrm{N}}_{\mathrm{B}} = \hat{\mathrm{N}}_{\mathrm{P}}\left(1 - \frac{1}{n_1}\right). \tag{9}$$

Se o número de animais capturados na segunda etapa do experimento de captura-recaptura é razoavelmente grande o estimador $\hat{\mathrm{N}}_{\mathrm{B}}$ é aproximadamente o estimador de Petersen.

Além disso, em (9) o ajuste para o estimador Bayesiano é similar às chamadas correções de Bartlett (1937), cujo objetivo é tornar os estimadores menos viciados.

No exemplo a seguir verificamos a influência de tal ajuste nos estimadores obtidos e comparamos o estimador (8) com o estimador de Petersen.

4. Exemplo

Podemos ilustrar esse problema de captura-recaptura com um exemplo numérico, através de uma simulação utilizando o *software* MINITAB.

Considerando conhecidos os valores de: N: tamanho da população animal, n_0: número de animais capturados na 1ª etapa do experimento e n_1: número de animais capturados na 2ª etapa do experimento, geramos 200 dados a partir de uma distribuição Binomial com parâmetros n_1 e $\frac{n_0}{N}$. Esses dados gerados correspondem aos valores de m: número de animais marcados observados na 2ª etapa do experimento. Substituindo esses valores nos estimadores (2) e (8), podemos compará-los com o verdadeiro valor de N e verificar qual deles é o mais eficiente.

Seja N = 600 e $n_0 = 180$. Assim, $\pi = \frac{n_0}{N} = 0.3$.

Consideremos algumas distribuições a priori para c com diferentes parâmetros a e b, apresentadas a seguir:

Tabela 2: Parâmetros a e b para a priori $\text{Beta}(a,b)$ de π

a	b	$E(\pi)$	$Var(\pi)$
1	2	0.3333	0.055556
6	14	0.3000	0.010000
9	21	0.3000	0.006774
16	37	0.3019	0.003903
32	73	0.3048	0.001999

Notemos que, à medida que os parâmetros a e b aumentam, mais informativa se torna a distribuição a priori para π. Além disso, consideremos diferentes valores para n_1 (30, 50, 70 e 90) para verificar qual seu impacto nos estimadores de Petersen e Bayesiano.

Utilizando o *software* MINITAB, geramos 200 valores para m e verificamos o comportamento da eficiência relativa dos estimadores de Petersen e Bayesiano para diferentes distribuições a priori através da seguinte expressão:

$$M_{\hat{N}_i} = \frac{\sum_{i=1}^{200}\left(\frac{\hat{N}_i - N}{N}\right)^2}{200}. \qquad (10)$$

Também apresentamos uma comparação da eficiência relativa do estimador Bayesiano em relação ao estimador de Petersen, utilizando o seguinte critério:

$$Ef\left(\hat{N}_B\right) = \sum_{i=1}^{200} \frac{\left(\frac{\hat{N}_{P_i} - N}{\hat{N}_{B_i} - N}\right)^2}{200}.$$

Assim, nas Tabelas 3 e 4, temos os seguintes resultados:

Tabela 3: Eficiência relativa de $\hat{N}_P$ e $\hat{N}_{B(a,b)}$

parâmetros da priori (a **e** b)	$M_{\hat{N}}$	$n_1 = 30$	$n_1 = 50$	$n_1 = 70$	$n_1 = 90$
—	$\hat{N}_P$	0.1499	0.0588	0.0425	0.034
(0, 0)	$\hat{N}_{B(\text{não-inf})}$	0.13286	0.05533	0.04046	0.03286
(1, 2)	$\hat{N}_{B(1,2)}$	0.09126	0.04608	0.03555	0.02962
(6, 14)	$\hat{N}_{B(6,14)}$	0.031871	0.024589	0.022249	0.019979
(9, 21)	$\hat{N}_{B(9,21)}$	0.0208	0.0184	0.01760	0.01630
(16, 37)	$\hat{N}_{B(16,37)}$	0.00993	0.010692	0.01112	0.010923
(32, 73)	$\hat{N}_{B(32,73)}$	0.003555	0.004757	0.005307	0.005597

Tabela 4: Comparação da eficiência relativa de $\hat{N}_B$ em relação a $\hat{N}_P$

		$Ef\left(\hat{N}_B\right)$			
parâmetros da priori (a **e** b)	**Estimador**	$n_1 = 30$	$n_1 = 50$	$n_1 = 70$	$n_1 = 90$
(0, 0)	$\hat{N}_{B(\text{não-inf})}$	1.0524688	0.9862195	1.0084661	1.0210054
(1, 2)	$\hat{N}_{B(1,2)}$	1.5322839	1.2544771	1.2064427	1.2106201
(6, 14)	$\hat{N}_{B(6,14)}$	3.4654519	2.0316789	1.7189549	1.5527626
(9, 21)	$\hat{N}_{B(9,21)}$	5.2200724	2.6993872	2.1467539	1.8612969
(16, 37)	$\hat{N}_{B(16,37)}$	12.19422	5.1588392	3.62635	2.9654916
(32, 73)	$\hat{N}_{B(32,73)}$	134.82521	46.495151	22.592091	20.653773

Pela Tabela 3 observamos que o estimador Bayesiano é mais consistente que o estimador de Petersen, pois produz medidas menores, principalmente nos casos em que foram utilizadas prioris informativas.

Através da Tabela 4, verificamos a eficiência do estimador Bayesiano em relação ao estimador de Petersen, que se torna significativa à medida que incorporamos informações a priori e trabalhamos com amostras pequenas. No caso de amostras grandes tal eficiência não é tão significativa pois as informações a priori são absorvidas pelos dados contidos na amostra.

5. Conclusões

A utilização da Inferência Bayesiana em problemas de captura-recaptura se torna viável quando o pesquisador dispõe de informações a priori a respeito do comportamento da população animal em estudo.

Neste trabalho consideramos problemas de captura-recaptura com homogeneidade, ou seja, cuja probabilidade de captura é a mesma para todos os animais da população.

O fato do estimador Bayesiano ser proporcional ao estimador de Petersen é muito importante pois, numa situação de total desinformação (na qual utilizamos uma priori não-informativa) o fator de ajuste do estimador Bayesiano se torna similar às chamadas correções de Bartlett (1937), cujo objetivo é diminuir o vício dos estimadores.

Para pequenas amostras, notamos que o estimador Bayesiano é muito mais eficiente que o estimador de Petersen, principalmente quando incorporamos informações confiáveis (prioris informativas) no nosso estudo. Esse resultado pode ser empregado, por exemplo, no estudo de uma espécie em extinção, na qual N é razoavelmente pequeno e, conseqüentemente, também n_1 o será. Qualquer informação a priori incorporada ao estudo será de grande auxílio para a obtenção do estimador Bayesiano para N de modo que este seja mais eficiente que o estimador de Petersen.

Uma proposta de estudo é analisar situações mais complexas como problemas de captura-recaptura com heterogeneidade, ou então situações nas quais a probabilidade de recaptura de um animal é diferente da probabilidade de captura.

Além disso, deixamos como proposta um estudo para obtenção de um critério visando a escolha adequada de n_1 ou o desenvolvimento de uma Inferência Bayesiana considerando uma distribuição a priori para n_1, por exemplo uma distribuição Poisson.

6. Agradecimentos

Agradecemos ao Prof. Dr. José Galvão Leite - IME - USP por nos fornecer os cálculos matemáticos necessários para a confecção da Tabela 1.

- **ABSTRACT:** The purpose of this paper is to suggest an adjustment for the Petersen's estimator (1896), via maximization of the predictive density, which is similar to Bartlett's correction (1937). A genereted data set is used to analyse the efficiency of this adjustment.

- **KEYWORDS:** Bayesian Inference, Capture-Recapture, Predictive Density.

7. Referências

Bartlett, M.S. (1937). Properties of Sufficiency and Statistical Tests, Proc. Roy. Soc. Ser. A 160, 268-282.

Bernardo, J. M. e Smith, A. F. M. (1994). Bayesian Theory. Wiley, Chichester.

O'Hagan, A. (1994). Kendall's Advanced Theory of Statistics - Vol. 2B - Bayesian Inference. Edward Arnold.

Petersen, G. G. J. (1896). The Yearly Immigration of Young Plaice Into the Limfjord from the German Sea. Rept. Danish Biol. Sta. 6, 1-48.

Press, S. J. (1989). Bayesian Statistics: Principles, Models, and Applications. John Wiley and Sons.

NOTAS DO ICMSC

SÉRIE ESTATÍSTICA

030/96 ACHCAR, J.A. - Bayesian inference for software reliability models considering interfailure time data.

029/96 ACHCAR, J.A. - Bayesian inference for software reliability models using homogeneous poisson process.

028/96 RODRIGUES, J.; LEITE, J.G. - Inference for the software reliability using imperfect recapture debbugging model.

027/96 ACHCAR, J.A.; DEY, D.K.; NIVERTHI, M. - A bayesian approach using nonhomogeneous Poisson process for software reliability models.

026/96 ANDRADE, M.G.; VAL, J.B.R. do - Um método numérico baseado na solução do valor médio para a equação de Helmholtz parte II: Rede triangular.

025/96 ANDRADE, M.G.; VAL, J.B.R. do - Um método numérico baseado na solução do valor médio para a equação de Helmholtz parte I: malha quadrada.

024/96 MAZUCHELLI, J.; ACHCAR, J.A. - Análise bayesiana para modelos de crescimento.

023/96 ANDRADE, M.G.; SOARES, S.; CRUZ Jr., G.; VINHAL, C.D.N - Uma abordagem estocástica para o planejamento a longo prazo da operação de sistema hidrotérmicos.

022/95 ACHCAR, J. - A generalized Moranda software reliability model: a bayesian approach.

021/95 RODRIGUES, J. - Inference for the software reliability using asymmetric loss functions: A hierarchical Bayes approach.