ESTIMAÇÃO SEQUENCIAL BAYESIANA APLICADA À PROPORÇÃO DE INFESTAÇÃO DE PSILÍDEOS EM ALECRIM DO CAMPO

Carla Regina Guimarães BRIGHENTI[1]
Mariana RESENDE[1]
Deodoro Magno BRIGHENTI[2]

- RESUMO: Em análise sequencial Bayesiana as observações são tomadas sequencialmente e deve-se parar a amostragem na observação em que o risco esperado *a posteriori* for maior que o risco imediato. O objetivo deste trabalho foi avaliar a técnica de estimação do parâmetro da proporção, através da análise sequencial Bayesiana e aplicar esse método em dados de infestação de psilídeo no alecrim do campo. A técnica é adequada nesta situação, pois em campo, normalmente o produtor possui uma informação que auxilia na caracterização da área e desta forma, a utilização de estimação com tamanho de amostra fixo pode aumentar os custos. O critério de parada da amostragem sequencial foi obtido a partir da função perda quadrática. Os valores dos parâmetros populacionais da Binomial foram fixados em 0,1; 0,3; 0,7; e 0,9. Estudou-se via simulação a distribuição *a priori* Beta e do custo por observação. Para valores extremos houve maior influência das *priori* adotadas. Já as proporções com valores intermediários as estimativas foram mais próximas do valor real. O grau de informatividade da *priori* foi dependente do custo. Na área de alecrim do campo avaliada, o processo foi interrompido após avaliação de 10 plantas, obtendo-se uma proporção de infestação de psilídeos igual a 28,57%. Concluiu-se que a técnica de estimação sequencial Bayesiana é mais adequada para estimação da proporção quando o valor populacional está mais afastado dos valores extremos.
- PALAVRAS-CHAVE: Proporção; distribuição Beta; função perda; variância *a posteriori*.

1 Introdução

O *Baccharis dracunculifolia* é uma planta arbustiva que ocorre no cerrado brasileiro e popularmente conhecida como alecrim-do-campo ou vassourinha (Fagundes *et al.*, 2005). Em algumas plantas de alecrim são encontradas as chamadas galhas. Elas são respostas morfogenéticas de plantas a induções, mecânicas ou químicas, por organismos indutores denominados galhadores como insetos, vírus, bactérias ou fungos. Esta associação entre a planta e o galhador não causa benefícios para o vegetal. Mas o galhador obtém abrigo e nutrição, além de garantir sua dispersão (Espírito-Santo & Fernandes, 2004).

O psilídeo *Baccharopelma* é um inseto galhador que ataca exclusivamente espécies de *Baccharis*, nas quais induz uma galha foliar elíptica (Burckhardt *et al.*, 2004).

[1] Universidade Federal de São João Del Rei – UFSJ, Departamento de Zootecnia, CEP: 36301-160, São João Del Rei, MG, Brasil. E-mail: *carlabrighenti@ufsj.edu.br* / *naninha_mr@yahoo.com.br*
[2] Apiário Brighenti Ltda. E-mail: *deodorobrighenti@gmail.com*

Sabe-se que é através da coleta de resina pelas abelhas africanizadas nesta planta é que ocorre a produção da chamada própolis verde de alto valor comercial (Maróstica Junior *et al*., 2008). A própolis é utilizada pelas abelhas para se protegerem contra ação de intempéries do meio ambiente, em fechar frestas e aberturas na colméia, evitando predadores e oscilações térmicas, mantendo livres de microrganismos e agentes patogênicos (Brighenti *et al*., 2006).

Tem-se observado recentemente, em condições de campo, que há plantas do alecrim que são mais atraídas pelas abelhas africanizadas (São-Thiago *et al*., 2008). Esse fato foi notado através das injúrias observadas em algumas plantas e relatado por apicultores em eventos recentes da área (Brighenti *et al*., 2010). A maior atratividade pode ocorrer pela liberação de compostos voláteis do *Baccharis* como mecanismo de defesa a insetos herbívoros, merecendo especial atenção neste caso, o psilídeo *Baccharopelma dracunculifoliae* (Hemiptera: Psilidae) (Lara e Fernandes, 1994; Espírito-Santo e Fernandes, 2004; São-Thiago *et al*., 2008).

Para se estimar a proporção de plantas atingidas por psilídeos é necessário construir um planejamento amostral que permita representar de maneira adequada a população de interesse. A avaliação dos níveis de infestação de insetos-pragas é essencial em um programa de manejo integrado de pragas (Hollingsworth e Gatsonis, 1990). Tal avaliação é realizada através de práticas de amostragem sequencial, cujos esforços podem ser otimizados. Este método sequencial se constitui em um método simples e apresenta vantagem por ter um número de amostras variável, requerendo um número diferente dependente do grau de infestação, ao contrário do procedimento convencional, onde o número de coletas é sempre fixo (Bearzoti e Aquino, 1994). No entanto, a mesma técnica pode ser utilizada, além do controle de pragas, para a tomada de decisão em relação à escolha de áreas mais ou menos atingidas por algum inseto, como a infestação de *B. drancunculifoliae* em área de alecrim. Ao avaliar sequencialmente o experimento, o pesquisador tem a opção de interromper o experimento assim que uma decisão possa ser tomada. Dessa forma busca-se minimizar o tempo operacional e diminuir custos (Andrade e Ogliari, 2007).

Para desenvolver planos de amostragem sequencial, duas metodologias têm sido bastante utilizadas, o teste sequencial da razão de probabilidades de Wald (1947) e o intervalo de confiança de Iwao (1975) (Estefanel, 1977; Fuxa *et al*., 1989; Lynch *et al*., 1990; Walker e Allsopp, 1993; Hall *et al.*, 1994; Bianco, 1995; Peng e Brewer, 1995; Farias, 1996). Contudo, a validação do plano de amostragem sequencial depende de aspectos regionais e sazonais. Uma maneira de incorporar as informações referentes a esses aspectos é atualizar o plano de amostragem por meio de técnicas Bayesianas, incorporando uma informação a *priori* (Garthwaite *et al.*, 1995).

A idéia geral em análise sequencial Bayesiana é que após cada observação realizada deve-se comparar o risco de Bayes *a posteriori* de tomar uma decisão imediata com o risco de Bayes *a posteriori* esperado que será obtido se mais observações são tomadas (Berger, 1985). O risco de Bayes de um procedimento sequencial d é definido por $r(\pi, d) = E^{\pi}[R(\theta, d)]$, ou seja, a esperança do risco dada *a priori* p(π).

O estimador de Bayes de π com respeito à função perda é aquele com menor risco de Bayes. No caso da função perda quadrática o estimador de Bayes para o parâmetro π será a média de sua distribuição atualizada.

O resultado clássico em análise Bayesiana tem mostrado que o menor risco de Bayes *a posteriori* é a variância da distribuição *a posteriori*, denotada por var_{post} (n) (Raiffa e Schlaifer, 1961). Daí, $r_o(\pi^n, n) = var_{post}(n)$ e o menor risco de Bayes *a posteriori* esperado, se outra observação é tomada, isto é, $r^1(\pi^n, n)$, é a esperança desta variância, ou seja, $r^1(\pi^n, n) = E_x[var_{post}(n)]$ (Pham-Gia, 1998). O tempo de parada dado em função do risco é o primeiro n para o qual $r_o(\pi^n, n) \leq r^1(\pi^n, n)$. Pode-se dizer ainda que a amostragem deva continuar sempre que a esperança do risco *a posteriori* (ou a esperança da variância se mais uma observação for realizada) for menor que o risco "atual" (variância *a posteriori* com n observações) de tomar uma decisão (Garthwaite *et al*., 1995).

Nota-se que a estimação da proporção, por exemplo, de plantas atingidas por psilídeos, em um contexto estatístico torna-se um objeto de pesquisa, no qual, a utilização de métodos inferenciais, especificamente, inferência bayesiana agrega informações precisas e plausíveis de serem atualizadas que auxiliarão na tomada de decisão.

Dessa maneira, o objetivo desse trabalho é estimar através da metodologia sequencial bayesiana, o parâmetro π (probabilidade de sucesso) da distribuição da proporção, aplicando esse método de estimação em dados reais de infestação do psilídeo *Baccharopelma dracunculifoliae* no alecrim do campo.

2 Metodologia

As coletas de dados em campo foram realizadas na microrregião de São João Del Rei, sudeste de Minas Gerais. Suas coordenadas geográficas com latitude S 21°08'00'' e longitude W 44°15'40'' com altitude de 898 metros. A vegetação da região é típica de cerrado com áreas muito degradadas. E entre as plantas nativas da região encontra-se o alecrim do campo.

Foram investigadas cinco áreas situadas em locais onde os apicultores da Apis Del Rei – Associação de Apicultores de São João Del-Rei mantêm seus apiários com colônias de produção de própolis verde que indica a presença de alecrim do campo.

A partir dos dados coletados foi realizada a estimação da proporção de infestação de psilídeos *B. dracunculifoliae* em alecrim do campo por inferência bayesiana.

De acordo com o objetivo proposto, a metodologia realizada encontra-se descrita nas seções 2.1- Critério de parada da amostragem sequencial, na qual se determina qual a regra de parada será utilizada; e em 2.2- Simulação da estimação sequencial Bayesiana, na qual se elaborada uma rotina no programa R 2.12.0, para obtenção dos resultados e validação do método proposto.

2.1 Critério de parada da amostragem sequencial

Em procedimento sequencial, extraem-se elementos de uma amostra um por vez, e depois que cada amostra é observada, utiliza-se um critério de parada e é tomada a decisão (rejeição/aceitação de uma hipótese ou estimação de um parâmetro) ou a observação de amostras é continuada.

Para a obtenção do critério de parada, foi considerada a função perda quadrática $Perda = (\pi - \hat{\pi})^2 + C(n)$, em que C(n) > 0 é o custo de tomar uma amostra de tamanho n. Para estimar π, foi calculada a esperança da variância *a posteriori* a qual permite estimar a quantidade $r(\pi^n, n)$ que representa o menor risco de Bayes que pode ser atingido uma vez que $\mathbf{x}^n$ foi observado.

Para decidir se deve ou não parar a amostragem, deve-se então comparar o risco de uma decisão imediata com o risco de continuar a amostragem.

Utilizou-se como procedimento de inspeção sequencial o procedimento Look-ahead, em que a inspeção termina no primeiro n (n = 0, 1,...) para o qual $r_o(\pi^n, n) = r^m(\pi^n, n)$, em que r_0 é o risco de Bayes imediato e r^m é o risco após observadas m valores de x, sendo chamado procedimento look-ahead *m*-passos. Para um procedimento "olhar à frente" a um passo, o tempo de parada é o primeiro n para o qual $r_o(\pi^n, n) \le r^1(\pi^n, n)$.

2.2 Simulação da estimação sequencial Bayesiana

Para estimar o parâmetro de proporção considerando o tamanho da amostra como variável aleatória definida por uma regra de decisão sequencial baseada na avaliação da função perda, inicialmente estudou-se via simulação o grau de informatividade da distribuição *a priori* em relação a distribuição *a posteriori*.

Para obtenção dos resultados e validação do método proposto foi elaborada uma rotina no programa R 2.12.0 (R Development Core Team, 2009). Foram geradas sequencialmente amostras de ensaios de Bernoulli, seguindo os valores paramétricos fixados em π = 0,1; 0,3; 0,7; e 0,9, mantendo-se um truncamento no procedimento com n=1000. Após o processo de simulação da população, foram calculados o risco imediato e o risco esperado de acordo com cada valor obtido sequencialmente, adotando-se como hiperparâmetros da *priori* Beta os valores 1, 3 e 5, tanto para o parâmetro α quanto para o parâmetro β. Quando $\alpha = \beta = 1$, a densidade de Beta se reduz à uniforme no intervalo (0,1), sendo uma distribuição não informativa e quando α ou $\beta = 3$ e α ou $\beta = 5$, tem-se respectivamente, as *prioris* de menor e maior grau de informatividade. Em relação ao custo foi adotado o valor 10^{-4}, conforme sugerido em Pham-Gia (1998). Foram também simulados as mesmas situações paramétricas, adotando-se outros valores para o custo, 10^{-3} e 10^{-5}. Os valores eram comparados e gerada nova amostra até que $r_o(\pi^n, n) \le r^1(\pi^n,n)$, quando a estimativa do parâmetro da proporção era então calculada a partir da média da *posteriori* obtida até então.

A partir dos parâmetros calculados e hiperparâmetros fixados foram plotados os gráficos da verossimilhança, *priori* e *posteriori* obtida.

3 Resultados e discussão

Os resultados apresentados encontram-se descritos nas seções 3.1- Critério de parada de amostragem sequencial; 3.1.1- Esperança da variância *a posteriori* da distribuição Beta, 3.2- Simulação da estimação sequencial Bayesiana; e 3.3- Aplicação a partir de dados reais.

3.1 Critério de parada de amostragem sequencial

Seja π o parâmetro de um experimento de Bernoulli. Então $S_n = \sum_{i=1}^{n} X_i$ é uma estatística suficiente, em que x é o valor observado no ensaio de Bernoulli, e

$S_n \sim Bin(n,\pi)$. A verossimilhança para a soma de ensaios de Bernoulli s_n, correspondente a uma Binomial e pode ser escrita como:

$$f(s_n \mid \pi) = f_{Sn}(s_n) = \binom{n}{s_n} \pi^{s_n} (1-\pi)^{n-s_n} . \tag{1}$$

A distribuição *a priori* conjugada para $p(\pi)$ será uma Beta dada por:

$$p(\pi \mid \alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \pi^{\alpha-1} (1-\pi)^{\beta-1}, \text{ em que } \alpha > 0 \text{ e } \beta > 0 . \tag{2}$$

Então *a posteriori* π^n para o parâmetro π será dada por:

$$\pi^n (\pi \mid s_n) \propto f(s_n \mid \pi) \, x \, p(\pi \mid \alpha, \beta) . \tag{3}$$

Incluindo o termo constante para completar a função densidade de probabilidade Beta:

$$p(\pi \mid s_n) = \frac{\pi^{\alpha+s_n-1} (1-\pi)^{\beta+n-s_n-1}}{\dfrac{\Gamma(\alpha+s_n)\Gamma(\beta+n-s_n)}{\Gamma(\alpha+\beta+n)} \underbrace{\int_0^1 \frac{\Gamma(\alpha+\beta+n)}{\Gamma(\alpha+s_n)\Gamma(\beta+n-s_n)} \pi^{\alpha+s_n-1} (1-\pi)^{\beta+n-s_n-1} d\pi}_{\text{Função densidade de probabilidade de uma Beta } (\alpha+s_n, \beta+n-s_n)}} .$$

Esta expressão mostra que a família conjugada natural é a família Beta de distribuições com os parâmetros $\alpha + s_n, \beta + n - s_n$.

Para a distribuição *a priori* Beta (α, β), a média é dada por $\dfrac{\alpha}{(\alpha+\beta)}$ e sua variância $\text{var}_{prior} = \dfrac{\alpha\beta}{\left[(\alpha+\beta)^2 (\alpha+\beta+1)\right]}$. Para a distribuição *a posteriori* de π dada pela distribuição beta com parâmetros $\alpha + s_n, \beta + n - s_n$ média e a variância serão:

$$\mu_{post} = \frac{\alpha + s_n}{(\alpha + s_n + \beta + n - s_n)}, \tag{4}$$

$$\text{var}_{post} = \frac{(\alpha+s_n)(\beta+n-s_n)}{(\alpha+\beta+n)^2 (\alpha+\beta+n+1)}, \tag{5}$$

dependentes do tamanho da amostra n.

Utilizando a distribuição a *posteriori* conjugada após n observações $p^n(\pi \mid s_n)$, e sua variância *a posteriori*, obteve-se o seguinte critério de parada, baseado no procedimento look-ahead *1*-passos, em que:

$$r_o(\pi^n, n) \leq r^1(\pi^n, n)\,. \tag{6}$$

Assim, para determinar o critério de parada, foi necessário calcular o risco imediato que é dado pela variância *a posteriori* acrescida do custo de n observações e o risco esperado dado pela esperança da variância *a posteriori* e acréscimo no custo de mais uma observação, assim:

$$\mathrm{r_o}(\pi^{\mathrm{n}}, \mathrm{n}) = \frac{(\alpha+s_n)(\beta+n-s_n)}{(\alpha+\beta+n)^2(\alpha+\beta+n+1)} + C(n). \tag{7}$$

$$\mathrm{r}^1(\pi^{\mathrm{n}}, \mathrm{n}) = \mathrm{E}\left[\mathrm{var}_{post}\right] + C(n+1). \tag{8}$$

Esperança da variância *a posteriori* da distribuição Beta

Sendo W(n) a variância *a posteriori d*e Beta, em que:

$$W(n) = \frac{(\alpha+s_n)(\beta+n-s_n)}{(\alpha+\beta+n)^2(\alpha+\beta+n+1)}, \tag{9}$$

e $s_n = \sum_{i=1}^{n} X_i$, tem-se que a esperança da variância de tomar uma observação x_{n+1} será:

$$E\left[W(n)\right] = W(n+1)P[sucesso] + W(n+1)P[fracasso].$$

$$E\left[W(n)\right] = \frac{(\alpha+s_n+1)(\beta+n+1-s_n-1)}{(\alpha+\beta+n+1)^2(\alpha+\beta+n+2)}\pi + \frac{(\alpha+s_n+0)(\beta+n+1-s_n-0)}{(\alpha+\beta+n+1)^2(\alpha+\beta+n+2)}(1-\pi).$$

Como Beta é *a priori* do parâmetro p da Bernoulli, então:

Probabilidade de sucesso π = E(π) = média da Beta *a posteriori* com *n* observações, ou, seja, $\alpha' = (\alpha+s_n)$ e $\beta' = (\beta+n-s_n)$,

$$\pi = \frac{(\alpha+s_n)}{(\alpha+s_n)+(\beta+n-s_n)} = \frac{\alpha'}{\alpha'+\beta'} = \frac{(\alpha+s_n)}{(\alpha+\beta+n)}.$$

Então: $(1-\pi) = \dfrac{(\beta+n-s_n)}{(\alpha+\beta+n)}$.

$$E\left[W(n)\right] = \frac{(\alpha+s_n+1)(\beta+n-s_n)(\alpha+s_n)}{(\alpha+\beta+n+1)^2(\alpha+\beta+n+2)(\alpha+\beta+n)} + \frac{(\alpha+s_n)(\beta+n-s_n+1)(\beta+n-s_n)}{(\alpha+\beta+n+1)^2(\alpha+\beta+n+2)(\alpha+\beta+n)}.$$

$$E\left[W(n)\right] = \frac{(\alpha+s_n)(\beta+n-s_n)}{(\alpha+\beta+n)(\alpha+\beta+n+1)}\frac{\left[\alpha+\beta+n+2\right]}{(\alpha+\beta+n+1)(\alpha+\beta+n+2)}\frac{(\alpha+\beta+n)}{(\alpha+\beta+n)}.$$

Substituindo o valor correspondente por W(n), tem-se:

$$E\left[W(n)\right]=W(n)\frac{(\alpha+\beta+n+2)}{(\alpha+\beta+n+1)}\frac{(\alpha+\beta+n)}{(\alpha+\beta+n+2)}.$$

$$E\left[W(n)\right]=W(n)\frac{(\alpha+\beta+n)}{(\alpha+\beta+n+1)}. \tag{10}$$

Então, o risco esperado será dado por:

$$r^1\left(\pi^n,n\right)=E\left[W(n)\right]+C(n+1)=\frac{\alpha+\beta+n}{\alpha+\beta+n+1}W(n)+C(n+1). \tag{11}$$

Após interromper a amostragem utiliza-se como estimador Bayesiano da proporção, considerando uma função perda quadrática, a média da distribuição beta com parâmetros $\alpha+s_n, \beta+n-s_n$ (Winkler, 1972) dada por:

$$\frac{\alpha+s_n}{(\alpha+s_n+\beta+n-s_n)}. \tag{12}$$

3.2 Simulação da estimação sequencial Bayesiana

Para estudar via simulação o grau de informatividade da distribuição *a priori* em relação a distribuição *a posteriori,* considerou-se alguns parâmetros da Beta, que poderiam modelar diferentes distribuições *a priori* no intervalo [0,1].

A partir dos valores estabelecidos para *a priori* e o custo obtiveram-se as seguintes estimativas para o parâmetro da proporção (Tabela 1).

A partir dos resultados obtidos via simulação com o custo 10^{-4}, observa-se que, os valores encontrados para proporção igual a 0,1 foram aqueles em que se obteve o pior desempenho da estimação sequencial bayesiana. Considerando-se o intervalo de credibilidade a 95%, pode-se afirmar que as melhores estimativas foram obtidas quando da utilização das *priori* dadas por Beta (1,5), Beta (3,1) e Beta (3,3). Para a proporção 0,9, as piores estimativas foram obtidas quando se utilizou as *priori* Beta (1,5), Beta (3,3) e Beta (5,5). Assim, as *priori* utilizadas tiveram influência considerável quando da estimação da proporção em casos mais extremos. Em proporção menor ocorreu uma superestimação e para proporção maior ocorreu uma subestimação.

Os resultados obtidos na simulação com proporção 0,3 tiveram estimativas muito próximas do parâmetro real. Todos os valores encontrados podem ser considerados adequados. Com destaque para algumas *priori* tais como Beta (1,3), Beta (3,5) e Beta (5,5), em que as estimativas encontradas foram **0,29** e **0,30**. Alguns valores obtidos se encontraram mais distante do real como *a priori* não informativa Beta(1,1), mesmo assim pode ser considerada adequada, evidenciando a não influencia da distribuição *a priori* quando a proporção populacional assumiu o valor 0,3.

Considerando a proporção populacional igual a 0,7, os valores obtidos também foram adequados. As *priori* iguais a Beta (1,3) e Beta (3,3) foram aquelas que forneceram as melhores estimativas, sendo iguais a **0,71** e **0,70** respectivamente. *A priori* que forneceu a estimativa mais distante do parâmetro populacional real foi a Beta (1,5), com parâmetro estimado em 0,53.

Tabela 1 - Estimativas Bayesianas obtidas por amostragem sequencial via simulação, para o parâmetro da proporção (π), considerando custo por observação igual a 10^{-4} e diferentes parâmetros populacionais e hiperparâmetros

π	Parâmetros da *priori* α	β	Média da *priori*	Tamanho da amostra	Parâmetros da *posteriori* α'	β'	Média da *posteriori*	$ICr_{95\%}(\pi)$
0,1	1	1	0,50	40	10	32	0,24	[0,124;0,376]
		3	0,25	37	9	32	0,22	[0,108;0,356]
		5	0,17	34	8	32	0,20	[0,093;0,335]
	3	1	0,75	36	8	32	0,20	[0,093;0,335]
		3	0,50	32	7	31	0,18	[0,079;0,320]
		5	0,38	33	9	32	0,22	[0,108;0,356]
	5	1	0,83	35	9	32	0,22	[0,108;0,356]
		3	0,63	33	9	32	0,22	[0,108;0,356]
		5	0,50	32	10	32	0,24	[0,124;0,376]
0,3	1	1	0,50	46	19	29	0,40	[0,264;0,536]
		3	0,25	42	14	32	**0,30**	[0,182;0,443]
		5	0,17	42	17	31	0,35	[0,227;0,493]
	3	1	0,75	37	9	32	0,22	[0,108;0,356]
		3	0,50	41	16	31	0,34	[0,214;0,480]
		5	0,38	37	13	32	**0,29**	[0,168;0,429]
	5	1	0,83	37	11	32	0,26	[0,139;0,395]
		3	0,63	40	17	31	0,35	[0,227;0,493]
		5	0,50	36	14	32	**0,30**	[0,182;0,443]
0,7	1	1	0,50	45	31	16	0,66	[0,519;0,786]
		3	0,25	41	32	13	**0,71**	[0,572;0,832]
		5	0,17	43	26	23	0,53	[0,392;0,667]
	3	1	0,75	43	31	16	0,66	[0,519;0,786]
		3	0,50	40	32	14	**0,70**	[0,557;0,818]
		5	0,38	40	30	18	0,63	[0,485;0,755]
	5	1	0,83	42	30	18	0,63	[0,485;0,755]
		3	0,63	35	32	11	0,74	[0,605;0,861]
		5	0,50	38	29	19	0,60	[0,464;0,736]
0,9	1	1	0,50	38	32	8	0,80	[0,665;0,907]
		3	0,25	36	32	8	0,80	[0,665;0,907]
		5	0,17	37	32	11	0,74	[0,605;0,861]
	3	1	0,75	36	32	8	0,80	[0,665;0,907]
		3	0,50	35	32	9	0,78	[0,644;0,891]
		5	0,38	30	31	7	0,82	[0,679;0,920]
	5	1	0,83	32	31	7	0,82	[0,679;0,920]
		3	0,63	30	31	7	0,82	[0,679;0,920]
		5	0,50	31	32	9	0,78	[0,644;0,891]

A priori Beta (1,5) foi aquela que mais se destacou em relação as demais, sendo considerada a mais informativa.

Obteve-se a correlação entre a estimativa e o tamanho amostral para cada valor paramétrico estimado, sendo esta igual a 0,357; 0,807; -0,408 e -0,527 para as proporções 0,1; 0,3; 0,7 e 0,9 respectivamente, nos quais os tamanhos médios das amostras e seus erros padrões foram: 35(0,88), 40(1,10), 41(1,00) e 34(1,05).

As mesmas conclusões foram obtidas para a simulação com custos iguais a 10^{-3} e 10^{-5}. Proporções maiores ou menores, próximas a 0 ou 1, tiveram influência das *priori* adotadas e resultaram em estimativas distantes da proporção real. As proporções com valores intermediários tiveram mais sucesso em todos os casos avaliados. Vale ressaltar que, para custo maior o tamanho de amostra foi reduzido em todos os casos, enquanto que, para o custo menor este foi aumentado como era de se esperar.

Quanto ao grau de informatividade das *priori*, pode se observar que para custos menores houve menor influencia da *priori*, assim, no caso da estimação sequencial bayesiana, para custos elevados, deve-se verificar a adequação da *priori*, pois ela é de alta informatividade, já que, deve ocorrer estimação do parâmetro populacional com redução do tamanho amostral para diminuição de custos.

3.3 Aplicação a partir de dados reais

A técnica de estimação sequencial bayesiana de uma proporção foi aplicada a conjunto de dados reais obtido de experimentos de campo realizados na microrregião de São João Del Rei. Para facilitar tal aplicação, estabeleceu-se o seguinte algoritmo:

Algoritmo

Passo 1 → Entrar com os seguintes valores:

hiperparâmetros (α, β) para estabelecimento da distribuição *a priori*,
tamanho da amostra (n),
soma dos ensaios de Bernoulli (s_n) e
custo (C);

- Os valores de n e s_n devem ser atualizados a cada amostragem;

Passo 2 → Cálculo dos parâmetros da distribuição *a posteriori*:

$$\alpha' = \alpha + s_n , \quad \beta' = \beta + n - s_n .$$

Passo 3 → Cálculo dos riscos imediato e esperado:

$$r_0 = \frac{(\alpha + s_n)(\beta + n - s_n)}{(\alpha + \beta + n)^2 (\alpha + \beta + n + 1)} + C(n) \ ,$$

$$r_1 = \frac{(\alpha + \beta + n)}{(\alpha + \beta + n + 1)} r_0 + C(n+1) .$$

Passo 4 → Regra decisória:

- Continue – enquanto $r^1 < r_0$, atualizar n (acrescentar uma unidade) e s_n (somar 1 quando obter sucesso ou 0 quando fracasso).

- Pare – quando $r^1 > r_0$, fazer o cálculo da média da *posteriori* dada por:

$$\mu_{post} = \frac{\alpha'}{(\alpha' + \beta')} \ .$$

APLICAÇÃO:

Considerando o custo por observação igual a 10^{-3}, hiperparâmetros $\alpha = 1$ e $\beta = 3$, foram observadas sequencialmente, plantas do alecrim do campo, verificando-se a presença (1) ou ausência (0) do psilídeo *B. dracunculifoliae*, obtendo-se o seguinte resultado:

- Planta 1= psilídeo presente

  $s_1 = 1$, $\alpha' = 2$, $\beta' = 3$.

$$r_0 = \frac{(1+1)(3+1-1)}{(1+3+1)^2(1+3+1+1)} + 0,001(1) = 0,041 \ ,$$

$$r_1 = \frac{(1+3+1)}{(1+3+1+1)} 0,041 + 0,001(2) = 0,0353 \ .$$

Decisão: Como $r_0 > r_1$ continua-se a amostragem

- Planta 2= psilídeo ausente

  $s_2 = 1$, $\alpha' = 2$, $\beta' = 4$

$$r_0 = \frac{(1+1)(3+2-1)}{(1+3+2)^2(1+3+2+1)} + 0,001(2) = 0,0337 \ ,$$

$$r_1 = \frac{(1+3+2)}{(1+3+2+1)} 0,0317 + 0,001(3) = 0,0302 \ .$$

Decisão: Como $r_0 > r_1$ continua-se a amostragem.

E assim sucessivamente, até que o $r_0 < r_1$ e a amostragem foi interrompida com $n = 10$ e a proporção estimada, a partir da média *a posteriori* em 0,2857.

Conclusões

A simulação foi uma técnica eficiente para estudar os planos de amostragem sequencial por meio de técnicas bayesianas. E possibilitou estimar a parâmetro π (probabilidade de sucesso) da proporção da infestação do psilídeo *Baccharopelma dracunculifoliae* no alecrim do campo.

Verificou-se via simulação que a distribuição *a priori* é mais informativa ao valor estimado para o parâmetro de proporção, nos casos em que o parâmetro populacional está próximo aos extremos 0 e 1.

A Estimação Sequencial Bayesiana é mais adequada quando o parâmetro é estimado com valores mais centrais no espaço paramétrico [0,1].

O tamanho amostral médio para a estimação sequencial Bayesiana foi igual a 37,5, quando estabelecido um custo de 10^{-4} para a obtenção de uma nova amostra. Quando o custo foi considerado maior houve redução do tamanho amostral.

A partir da técnica de estimação estudada foi possível estabelecer um algoritmo para ser utilizado na estimação em experimentos de campo, realizando-se a amostragem sequencial concomitante ao cálculo dos riscos para a tomada de decisão.

Na área de alecrim do campo avaliada, utilizando-se a técnica de estimação sequencial Bayesiana, foi obtida uma proporção de infestação de psilídeos *B. dracunculifoliae* igual a 28,57%.

Agradecimentos

Os autores agradecem a Fundação de Amparo a Pesquisa do Estado de Minas Gerais (FAPEMIG) pelo apoio financeiro ao Projeto APQ-01655-10.

BRIGHENTI, C. R. G; RESENDE, M.; BRIGHENTI, D. M. Sequential Bayesian estimation applied to the proportion of psyllids infestation in rosemary field. *Rev. Bras. Biom.*, São Paulo, v.29, n.2, p.342-354, 2011.

- *ABSTRACT: In sequential Bayesian analysis the observations are sequentially taken and must stop the sampling on the observation that the expected posteriori risk is higher than the immediate risk. The objective of this study was to evaluate the technique of estimating the parameter of the proportion, through the sequential Bayesian analysis and apply this method on data from psyllid infestation in the rosemary field. The technique is appropriate in this situation because in the field, the producer usually has some information that helps in the characterization of the area and thus the use of sample size estimation with fixed costs may increase. The criterion of stop of the sampling sequence was obtained starting from the quadratic loss function. The values of the population parameters Binomial were fixed in 0,1; 0,3; 0,7; and 0,9.It was Studied via simulation the degree of informativeness of the Beta a prior distribution and the cost per observation. For extreme values were more influence by priori adopted. Since the proportions with intermediate values the estimates were closer of the actual value. The degree of informativeness of the prior was dependent on the cost. In the area of rosemary, which was evaluated, was obtained a proportion of infestation psyllids equal to 28.57% thereby obtaining the sample size same to 10 plant evaluated. It was concluded that the Bayesian sequential estimation technique is best suited for estimation of the proportion when the population value is farthest from the extreme values.*
- *KEYWORDS: Proportion; Beta distribution; loss function; posteriori variance.*

Referências

ANDRADE, D. F.; OGLIARI, P. J. *Estatística para as ciências agrárias e biológicas: com noções de experimentação*. Florianópolis: UFSC, 2007. 438p.

BEARZOTI, B.; AQUINO, L. H. Plano de amostragem seqüencial para avaliação de infestação de bicho-mineiro (*Lepidoptera: Lyonetiidae)* no sul de Minas Gerais. *Pesq. Agrop. Bras.*, Brasília, v.29, n.5, p.695-705, 1994.

BERGER, J. O. *Statistical decision therory and Bayesian analysis*. 2.ed. New York: Springer-Verlag, 1985. 624p.

BIANCO, R. *Construção e validação de planos de amostragem para o manejo da lagarta docartucho - Spodoptera frugiperda (J. E. Smith, 1797) (Lepidoptera, Noctuidae) na cultura do milho.* 1995. 103f. Tese (Doutorado em Entomologia) – Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 1995.

BRIGHENTI, C. R. G.; BARROS, D. L.; BRIGHENTI, D. M.; OLIVEIRA, T. G. S. Amostragem sequencial para seleção de área para produção de própolis verde. In: CONGRESSO BRASILEIRO DE APICULTURA, 2010, Cuiabá. *Anais...*

BRIGHENTI, M. D; BRIGHENTI, C. R. G; SANTOS, F. C. Método para intensificar a produção de própolis. *Mensagem Doce*, São Paulo, n.85, p.2-6, 2006.

BURCKHARDT, D.; ESPÍRITO-SANTO, M. M.; FERNANDES, G. W.; MALENOVSKÝ, I. Gall-inducing jumping plant-lice of the Neotropical genus *Baccharopelma* (Hemiptera: Psylloidea) associated with *Baccharis* (Asteraceae). *J. Nat. Hist.*, Chichester, v.38, p.2051-2071, 2004.

ESPÍRITO-SANTO, M. M.; FERNANDES, G. W. Parasitoid attack and its consequences to the development of the galling psyllid *Baccharopelma dracunculifoliae*. *Basic Appl. Ecol.*, Jena, v.5: n.5, p.475-484, 2004.

ESTEFANEL, V. *A amostragem sequencial baseada no teste sequencial da razão de probabilidades e seu uso no controle de lagartas da soja no Estado do Rio Grande do Sul*. 1977. 117f. Dissertação (Mestrado em Experimentação e Estatística) – Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 1977.

FAGUNDES, M.; NEVES, F. S.; FERNANDES, G. W. Direct and indirect interactions involving ants, insect herbivores, parasitoids, and the host plant *Baccharis dracunculifolia* (Asteraceae). *Ecol. Entomol.*, Oxford, v.30, n.1, p.28-35, 2005.

FARIAS, P. R. S. *Distribuição espacial e amostragem sequencial de Spodoptera frugiperda (J. E. Smith, 1797) na cultura do milho.* 1996. 129f. Dissertação (Mestrado em Entomologia Agrícola) – Faculdade de Ciências Agrárias e Veterinárias, Universidade Estadual Paulista, Jaboticabal, 1996.

FUXA, J. R.; MITCHELL, F. L. Sampling to determine when to control fall armyworms. *Louisiana Agric.*, Baton Rouge, v.33, n.1, p.3-5, 1989.

GARTHWAITE, P. H.; JOLLIFFE, I.; BYRON, J. *Statistical inference*. 2.ed. London: Prentice Hall, 1995. p.114-131.

HALL, D. G.; CHILDERS, C. C.; EGER, J. E. Spatial dispersion and sampling of citrus rust mite (Acari: Eriophyidae) on fruit in "Hamlin" and "Valencia" orange groves in Florida. *J. Econ. Entomol.,* Lanham, v.87, n.3, p.687-698, 1994.

HOLLINGSWORTH, C. S.; GATSONIS, C. A. Sequential sampling plans for green peach aphid (Homoptera: Aphididae) on potato. *J. Econ. Entomol.,* Lanham, v.83, n.4, p.1365-369, 1990.

IWAO, S. A new method of sequential samplig to classify populations relative to a critical density. *Res. Popul. Ecol.*, Kyoto, v.16, p.281-288, 1975.

LARA, A. C. F.; FERNANDES, G. W. Distribuição de galhas de *Neopelma baccharidis* (*Homoptera: Psyllidae*) em *Baccharis dracunculifolia (Asteraceae). Rev. Bras. Biol.*, São Paulo, v.54, n.4, p.661–668, 1994.

LYNCH, A. M.; FOWLER, S. W.; SIMMONS, G. A. Sequential sampling plans for spruce budworm (Lepidoptera: Tortricidae) egg mass density using Monte Carlo simulation. *J. Econ. Entomol.,* Lanham, v.83, n.4, p.1479-1484, 1990.

MARÓSTICA JUNIOR, M. R.; DAUGSCH, A.; MORAES, C. S.; QUEIROGA, C. L.; PASTORE, G. M., PARK, Y. K. Comparison of volatile and polyphenolic compounds in Brazilian green propolis and its botanical origin *Baccharis dracunculifolia. Ciênc. Tecnol. Alim.*, Campinas, 28, n.1, p.178-181, 2008.

PENG, C.; BREWER, G. J. Sampling plans for estimating a chene damage by the red sunflower seed weevil (Coleoptera: Curculionidae). *Can. Entomol.*, Ottawa, v.127, p.7-14, 1995.

PHAM-GIA, T. Distribution of the stopping time in bayesian sequential sampling. *Austral. & New Zeal. J. Stat.*, Oxford, v.40, n.2, p.221-227, 1998.

R DEVELOPMENT CORE TEAM. R:. *A language and environment for statistical computing. R Foundation for Statistical Computing*. Vienna, Austria, 2009.

RAIFFA, H.; SCHLAIFER, R. *Applied statistical decision theory*. Cambridge, MA: Harvard University Press, 1961. 356p.

SÃO-THIAGO, P.; SANTANA, R. T. M.; COSTA, A. G. F. C.; BASTOS, E. M. A. F. Visitação de abelhas *Apis mellifera* em populações de *Baccharis dracunculifolia* para coleta de resinas. In: ENCONTRO SOBRE ABELHAS, 8., 2008. Ribeirão Preto. *Anais…* CD Rom.

WALD, A. *Sequential analysis*. New York: John Wiley, 1947. 212p.

WALKER, P. W.; ALLSOPP, P. G. Sampling distributions and sequential sampling plans for *Eumargarodes laingi* and *Promargarodes* spp. (Hemiptera: Margarodidae) in Australian sugarcane. *Environ. Entomol.*, College Park, v.22, n.1, p.10-15, 1993.

WINKLER, R. L. *Introduction to Bayesian inference and decision*. 2.ed. New York: Holt, Renehart & Winston, 1972. 464p.

Recebido em 05.04.2011

Aprovado após revisão em 02.08.2011