UNIVERSIDADE FEDERAL DE SÃO JOÃO DEL-REI

ESTIMAÇÃO SEQUENCIAL BAYESIANA APLICADA A DADOS DE CONTAGEM DE GALHAS DE PSILÍDEOS *Baccharopelma dracunculifoliae*

MARIANA RESENDE

SÃO JOÃO DEL REI - MG

2013

UNIVERSIDADE FEDERAL DE SÃO JOÃO DEL-REI

ESTIMAÇÃO SEQUENCIAL BAYESIANA APLICADA A DADOS DE CONTAGEM DE GALHAS DE PSILÍDEOS *Baccharopelma dracunculifoliae*

Monografia submetida ao corpo docente da coordenação do Curso de Graduação Bacharelado Interdisciplinar em Biossistemas, da Universidade Federal de São João Del-Rei, como parte dos requisitos necessários para a obtenção do título de Bacharel em Biossistemas.

Orientadora: Carla Regina Guimarães Brighenti

MARIANA RESENDE

SÃO JOÃO DEL REI - MG

2013

UNIVERSIDADE FEDERAL DE SÃO JOÃO DEL-REI

ESTIMAÇÃO SEQUENCIAL BAYESIANA APLICADA A DADOS DE CONTAGEM DE GALHAS DE PSILÍDEOS *Baccharopelma dracunculifoliae*

MARIANA RESENDE

Monografia submetida ao corpo docente da coordenação do Curso de Graduação Bacharelado Interdisciplinar em Biossistemas, da Universidade Federal de São João Del-Rei, como parte dos requisitos necessários para a obtenção do título de Bacharel em Biossistemas.

DATA DE DEFESA: 21/06/2013

COMISSÃO EXAMINADORA:

CARLA REGINA GUIMARÃES

Carla Regina Guimarães Brighenti

MARCELO ANGELO CIRILLO

Marcelo Angelo Cirillo

Deodoro Magno Brighenti dos Santos

SÃO JOÃO DEL REI – MG

2013

MARIANA RESENDE

ESTIMAÇÃO SEQUENCIAL BAYESIANA APLICADA A DADOS DE CONTAGEM DE GALHAS DE PSILÍDEOS *Baccharopelma dracunculifoliae*

Defesa Aprovada pela Comissão Examinadora em : 21/06/2013

Comissão Examinadora:

CARLA REGINA GUIMARÃES

Profa. Dra. Carla Regina Guimarães Brighenti
Universidade Federal de São João Del Rei

MARCELO ANGELO CIRILLO

Prof. Dr. Marcelo Ângelo Cirillo
Universidade Federal de Lavras

Prof. Dr. Deodoro Magno Brighenti dos Santos
Universidade Federal de São João Del Rei

DEDICATÓRIA

Dedico este trabalho:

Aos meus pais e irmãos pelo apoio, carinho e incentivo oferecidos em todo momento de minha vida.

Aos meus amigos que me apoiaram e incentivaram para a concretização deste trabalho.

AGRADECIMENTOS

A Deus por ter me oferecido a oportunidade de viver, e me dar forças para evoluir a cada dia.

Aos meus pais, Maria Elisabete de Resende e Abel José de Resende, pelo amor incondicional, pela paciência e incentivo. Por terem feito o possível e o impossível para me oferecer a oportunidade de estudar, acreditar e respeitar minhas decisões e nunca deixando que as dificuldades acabassem com os meus sonhos.

Aos meus irmãos Daniel e João Marcelo, que de uma forma ou de outra, mesmo entre discussões sempre me apoiaram e acreditaram em mim.

A minha orientadora Carla Regina Guimarães Brighenti, pela oportunidade oferecida, pelo empenho, paciência, compreensão, ensinamentos, credibilidade e confiança depositados, obrigada imensamente por tudo. Obrigada pelo compartilhamento do conhecimento e além de tudo pelos ensinamentos para a vida. Fica aqui minha admiração, respeito e carinho pela pessoa e profissional.

As minhas companheiras e “irmãs” de república Ana Luiza, Dayane, Sabrina e Mayara, por terem sentido junto comigo, todas as angústias e felicidades, acompanhando cada passo de perto. Pelo amor, amizade, e apoio depositados, além da companhia por todos esses anos, melhor convívio, não poderia encontrar.

Aos amigos de turma, aos que passaram por minha vida e os que ainda estão presentes, pelas agradáveis lembranças que serão eternamente guardadas no coração.

Agradeço a todas as pessoas do meu convívio que acreditaram e contribuíram, mesmo que indiretamente, para a conclusão deste trabalho.

RESUMO

A ideia geral em análise sequencial Bayesiana é que após cada observação realizada deve-se comparar o denominado 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. Tal técnica pode ser aplicada na estimação da taxa média de infestação de plantas de alecrim do campo (*Baccharis dracunculifolia)* atingidas pelo inseto galhador (*Baccharopelma dracunculifoliae)*, que é importante para produtores de própolis verde. A contagem das galhas desse psilídeo em alecrim pode ser modelada pela distribuição de Poisson (θ) em que o parâmetro θ representa a taxa de infestação de galhas por planta. Dessa maneira, o objetivo desse trabalho foi estimar por meio da metodologia sequencial Bayesiana, o parâmetro θ da Poisson, aplicando este método em dados reais de infestação do psilídeo no alecrim. O critério de parada foi obtido a partir da função perda quadrática e a estimativa do parâmetro θ calculada a partir da média da *posteriori*. Para validação do método, foi elaborada uma rotina no programa R 2.13.1, que permitiu o estudo, via simulação, do grau de informatividade da *priori*. A partir desta rotina e das expressões dos riscos, foi elaborado em Delphi, o aplicativo para estimação do parâmetro. A estimação de θ também foi realizada da maneira frequentista para comparação. Para estimação do parâmetro da Poisson foram estabelecidos 18 casos para cada um dos quatro valores paramétricos fixados, sendo construído o Intervalo de Credibilidade a 95% para cada caso. Destes 72 casos simulados, 13 não continham o verdadeiro valor do parâmetro, sendo estes provenientes de simulações em que o custo foi alto, o que resultou em tamanhos amostrais pequenos ou ainda quando utilizou-se *priori* muito informativas. O aplicativo desenvolvido permitiu maior facilidade na realização da amostragem e análise de dados, sendo que, após realizada a estimação, o usuário tem acesso a um relatório em que consta o resumo da análise. Conclui-se que na área de alecrim do campo avaliada, foi obtida média de infestação de galhas de psilídeos igual a 3,26 galhas/planta, valor próximo ao encontrado pela estimação frequentista.

Palavras-chave: distribuição Gama, risco Bayesiano, alecrim, Delphi.

ABSTRACT

The general idea is Bayesian sequential analysis after each observation should be performed to compare the Bayes risk of subsequently taking an immediate decision with the risk of Bayes a *posteriori* expected to be obtained if more observations are taken. This technique can be applied in the estimation of the average level of infestation of plants rosemary field (*Baccharis dracunculifolia*) affected by gall (*Baccharopelma dracunculifoliae*), which is important for producers of green propolis. The count of galls in rosemary that psyllids can be modeled by the Poisson distribution where the parameter θ is the level of infestation of galls per plant. The aim of this study was to estimate the parameter θ by sequential Bayesian methodology. This method was applying on real data of psyllid infestation in rosemary field. The stopping time was obtained from the quadratic loss function and the estimation of the parameter θ calculated from the average of the posteriori. We created a routine in the program R 2.13.1 to validate the method. The degree of informativeness of the prior was studying via simulation The software in Delphi was developed for parameter estimation. The estimation of θ was also performed for comparison of frequentist method. For the parameter estimation of Poisson 18 cases were established for each of the four parametric values. Was being the credibility interval of 95% for each case. Of these 72 cases simulated 13 does not contain the actual value of the parameter being obtained from these simulations that the cost was high, which resulted in small sample sizes or when it was used prior very informative. The application developed allowed greater ease in performing the sampling and analysis of data, and after the estimation performed, the user has access to a report appearing in the summary of the analysis. In the field of rosemary evaluated was obtained average infestation of gall psyllids equal to 3.26 galls / plant, a value close to that found by estimating frequentist.

Keywords: Gamma distribution, Bayesian risk, rosemary, Delphi

LISTA DE FIGURAS

Figuras	Descrição
Figura 1	Curvas de densidade para as distribuições *a priori* Gama utilizadas para $\theta = 0{,}1$.
Figura 2	Distribuição *a priori* de θ, verossimilhança e distribuição *a posteriori* de $\theta\|x$ utilizadas para $\theta = 0{,}1$ em que θ não pertence ao $ICr_{95\%}$.
Figura 3	Distribuição *a priori* de θ, verossimilhança e distribuição *a posteriori* de $\theta\|x$ utilizadas para $\theta = 0{,}1$ em que θ pertence ao $ICr_{95\%}$.
Figura 4	Curvas de densidade para as distribuições *a priori* Gama utilizadas para $\theta = 1$.
Figura 5	Distribuição *a priori* de θ, verossimilhança e distribuição *a posteriori* de $\theta\|x$ utilizadas para $\theta = 1$ em que θ não pertence ao $ICr_{95\%}$.
Figura 6	Distribuição *a priori* de θ, verossimilhança e distribuição *a posteriori* de $\theta\|x$ utilizadas para $\theta = 1$ em que θ pertence ao $ICr_{95\%}$.
Figura 7	Curvas de densidade para as distribuições *a priori* Gama utilizadas para $\theta = 5$.
Figura 8	Distribuição *a priori* de θ, verossimilhança e distribuição *a posteriori* de $\theta\|x$ utilizadas para $\theta = 5$ em que θ não pertence ao $ICr_{95\%}$.
Figura 9	Distribuição *a priori* de θ, verossimilhança e distribuição *a posteriori* de $\theta\|x$ utilizadas para $\theta = 5$ em que θ pertence ao $ICr_{95\%}$.
Figura 10	Curvas de densidade para as distribuições *a priori* Gama utilizadas para $\theta = 10$.
Figura 11	Distribuição *a priori* de θ, verossimilhança e distribuição *a posteriori* de $\theta\|x$ utilizadas para $\theta = 10$ em que θ não pertence ao $ICr_{95\%}$.

Figura 12	Distribuição *a priori* de θ, verossimilhança e distribuição *a posteriori* de $\theta\|x$ utilizadas para $\theta = 10$ em que θ pertence ao $ICr_{95\%}$.
Figura 13	Galha no alecrim do campo
Figura 14	Psilídeos dentro da galha
Figura 15	Distribuição de Poisson com parâmetro 3,30.
Figura 16	Página inicial para inserção dos dados iniciais e dos dados observados.
Figura 17	Relatório com o resultado da Estimação Sequencial Bayesiana.
Figura 18	Página de armazenamento de dados.
Figura 19	Gráfico dos resultados da amostragem.
Figura 20	Memória de cálculo exportada para o Excel.

LISTA DE TABELAS

Tabelas	Descrição
Tabela 1	Estimativas Bayesianas obtidas por amostragem sequencial via simulação, para o parâmetro da distribuição de Poisson, considerando θ igual a 0,1, em que: n = tamanho da amostra, $IC_{95\%}$ = Intervalo de Credibilidade a 95%.
Tabela 2	Estimativas Bayesianas obtidas por amostragem sequencial via simulação, para o parâmetro da distribuição de Poisson, considerando θ igual a 1, em que: n = tamanho da amostra, $IC_{95\%}$ = Intervalo de Credibilidade a 95%.
Tabela 3	Estimativas Bayesianas obtidas por amostragem sequencial via simulação, para o parâmetro da distribuição de Poisson, considerando θ igual a 5, em que: n = tamanho da amostra, $IC_{95\%}$ = Intervalo de Credibilidade a 95%.
Tabela 4	Estimativas Bayesianas obtidas por amostragem sequencial via simulação, para o parâmetro da distribuição de Poisson, considerando θ igual a 10, em que: n = tamanho da amostra, $IC_{95\%}$ = Intervalo de Credibilidade a 95%.
Tabela 5	Resultado da contagem de galhas de psilídeos em plantas de alecrim e seus respectivos valores de probabilidade e número esperado de plantas de acordo com a distribuição de Poisson.

SUMÁRIO

1. Introdução 12
2. Revisão Bibliográfica 13
   2.1. Amostragem Sequencial 13
   2.2. Inferência Sequencial Bayesiana 14
   2.2.1.Distribuições *a priori* 17
   2.2.2. Inferência Bayesiana para dados de contagem 18
   2.3. Infestação de galhas de psilídeos em alecrim do campo 19
3. Metodologia 21
   3.1. Desenvolvimento da Distribuição *a posteriori* 21
   3.2. Critério de Parada da Amostragem Sequencial 22
   3.3. Simulação da Estimação Sequencial Bayesiana 23
   3.4. Coleta de dados reais 25
4. Resultados e Discussões 26
   4.1. Simulação da Estimação Sequencial Bayesiana 26
   4.2. Estimação Sequencial Bayesiana da contagem de galhas de psilídeos em alecrim do campo 39
   4.2.1. Estimação Frequentista do parâmetro da Poisson da contagem de galhas de psilídeos em alecrim do campo 40
   4.2.2. Software para Estimação Sequencial Bayesiana 43
5. Conclusão 46
6. Referências Bibliográficas 46

1. Introdução

Uma estratégia que tem sido muito utilizada para o monitoramento de pragas em diversas culturas é o Manejo Integrado de Pragas (MIP) que se caracteriza pela preocupação em desenvolver técnicas que interfiram o mínimo possível no meio ambiente (Bastos et al. 2006). Dessa forma, para o desenvolvimento satisfatório do MIP em um agroecossistema se faz necessário a utilização de um tipo de amostragem rápida e eficiente para o controle de pragas tais como o bicho mineiro do cafeeiro *(Perileucoptera coffella)*, a lagarta do algodão *(Albama argillacea)*, traça do tomateiro *(Tuta absoluta)* e a lagarta do milho *(Spodoptera frugiperda)*. A amostragem sequencial, em que o tamanho amostral é uma variável aleatória, é uma das ferramentas mais utilizadas para esse propósito (Bearzoti & Aquino, 1994).

Na amostragem sequencial, extrai-se elementos de uma amostra um por vez, e depois que cada amostra é observada, utiliza-se um critério de parada que será a tomada a decisão (rejeição/aceitação de uma hipótese ou estimação de um parâmetro) de interromper a amostragem ou a continuação da observação de amostras. Nesse tipo de amostragem, a avaliação pode classificar rapidamente a população do organismo de interesse em categorias como baixa, media ou alta. Ou indicar, ainda, se é necessário adotar alguma medida de controle ou não. Assim, o momento de terminar a amostragem e tomar uma decisão depende, em cada etapa, dos resultados obtidos a cada momento. Dessa forma o pesquisador tem a opção de interromper o experimento assim que uma decisão possa ser tomada, não sendo necessário pré-estabelecer um tamanho fixo amostral, buscando minimizar o tempo operacional e diminuir custos em campo (Fernandes et al , 2002).

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* que pode ser influenciada, por exemplo, pela presença de uma praga já observada por um agricultor, pelo índice de chuvas na região, pela flutuação populacional do inseto-praga (Brighenti et al, 2011). A ideia geral em análise sequencial Bayesiana é que após cada observação realizada deve-se comparar o denominado 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).

Portanto, a utilização de métodos inferenciais, especificamente inferência Bayesiana, agrega informações plausíveis de serem atualizadas que auxiliarão na tomada de decisão. Este pode ser o caso, por exemplo, da aplicação da técnica na estimação da taxa média de infestação de plantas de alecrim do campo (*Baccharis dracunculifolia)* atingidas pelo inseto galhador (*Baccharopelma dracunculifoliae)* da família Psilidae, que é importante para produtores de própolis verde. A contagem das galhas desse psilídeo em alecrim pode ser modelada pela distribuição de Poisson (θ) em que o parâmetro θ representa a taxa de infestação de galhas por planta. Assim, é possível ao produtor selecionar áreas de produção de própolis verde de acordo com a contagem estimada (Brighenti et al, 2011).

Dessa maneira, o objetivo desse trabalho é estimar por meio da metodologia sequencial Bayesiana, o parâmetro θ da distribuição de Poisson, aplicando este método de estimação em dados reais de infestação do psilídeo *B. dracunculifoliae* no alecrim do campo, *B. dracunculifolia,* a partir da criação de um software para auxílio na coleta e análise de dados.

2. Revisão Bibliográfica

2.1. Amostragem Sequencial

Basicamente existem dois tipos de amostragem: a amostragem convencional e a amostragem sequencial. O primeiro tipo caracteriza-se, principalmente, por apresentar um número fixo de observações, determinado antes de se iniciar o procedimento de amostragem, e os resultados são avaliados após o seu término, muitas vezes levando a decisões precipitadas com base apenas na experiência que se tem em campo, ou a tomada de decisões tardiamente, induzindo a utilização de inseticidas, o que pode acarretar em vários problemas (Farias et al, 2001).

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; Bianco, 1995; Peng & Brewer, 1995; Farias, 1996). O teste sequencial da razão de probabilidades (TSRP) tem por finalidade testar com o menor número possível de unidades amostrais, a hipótese nula H_0 versus H_1, ou seja, no caso do MIP, a rejeição de H_0 indica a necessidade de controle da praga, e sua aceitação, a não

aplicação de métodos de controle (Fernandes et al , 2002). Já o intervalo de confiança de Iwao consiste em uma metodologia que se apoia no teorema central do limite, que consiste em adaptar um intervalo de confiança em torno da média, gerando duas curvas divergentes, chamadas linhas de decisão do plano para testar a hipótese nula de que a infestação da população está abaixo do nível de controle (Neto et al, 1999).

A amostragem sequencial, desenvolvida por Wald em 1943, está sendo cada vez mais usada em entomologia, visando minimizar os custos, sem que haja uma redução na precisão dos resultados obtidos (Chiaradia et al, 2007). O crescente interesse pela utilização desse tipo de amostragem esta relacionado com o desenvolvimento do programa de MIP, sendo muito aplicada para avaliar a infestação de insetos nas culturas, para a tomada de decisão em relação à escolha de áreas mais ou menos atingidas por algum inseto e sobre o uso, ou não, de inseticidas.

Visto a dificuldade da montagem e acompanhamento de experimentos em campo, planos de amostragem sequencial podem ser considerados instrumentos valiosos quando já se dispõe de conhecimentos básicos sobre as populações das pragas e de seus inimigos naturais, incluindo estudos sobre comportamento e definição dos níveis de controle e de segurança, entre outros requisitos necessários ao desenvolvimento desses planos. Além do que os planos de amostragem sequencial são um instrumento de fácil uso para técnicos e mesmo para agricultores mais esclarecidos, que facilita a tomada de decisão quanto a executar a aplicação de defensivos ou não (Fernandes et al, 2003).

2.2. Inferência Sequencial Bayesiana

Após a realização do experimento, a variável em estudo X assume um valor x e o pesquisador pode então considerar a função $L(\theta,x) = f_x(x|\theta)$, como *função de verossimilhança*, sendo θ o parâmetro da distribuição . Toda a informação relevante que a amostra x pode fornecer sobre o parâmetro θ está contida na função de verossimilhança, possibilitando ao pesquisador estimar θ utilizando apenas a informação amostral. Esta inferência é denominada inferência frequentista (Rossi, 2002).

No entanto, outra possibilidade é utilizar a inferência Bayesiana. A Inferência Estatística Bayesiana é uma metodologia de análise de dados fundamentada em probabilidades condicionais. No modelo Bayesiano um parâmetro θ é considerado um vetor

aleatório desconhecido. A intensidade deste desconhecimento a respeito de θ pode assumir diferentes graus. Do ponto de vista Bayesiano, estes diferentes graus de incerteza são representados através de modelos probabilísticos para θ, formalmente denominados como distribuição *a priori* (Pinho, 2006).

A função de verossimilhança representa a contribuição dos dados para o conhecimento dos parâmetros e, a distribuição *a priori*, reflete a incerteza sobre os possíveis valores dos parâmetros antes da coleta dos dados (Fonseca et al., 2006).

A distribuição *a priori* para a quantidade de interesse θ deve representar, probabilisticamente o conhecimento que se tem sobre θ antes da realização do experimento. Neste caso, a distribuição é representada por uma forma funcional, cujos parâmetros devem ser especificados de acordo com este conhecimento e são denominados de hiperparâmetros para distingui-los do parâmetro de interesse θ (Lindley, 1990).

A combinação de verossimilhança de cada um dos possíveis valores de θ, L(θ,x), e *a priori* p(θ) leva a distribuição a *posteriori* p(θ|x). O teorema de Bayes é a regra de atualização utilizada para quantificar o aumento de informação.

$$p(\theta \mid x) \propto L(\theta; x)\, p(\theta).$$

Assim, a distribuição *a posteriori* é proporcional a verossimilhança × a distribuição *a priori*. A proporcionalidade dada pela constante normalizadora, podendo ser facilmente recuperada, de forma que $p(\theta \mid x) = kL(x \mid \theta)\, p(\theta)$ em que k é chamada distribuição preditiva e dada por $k^{-1} = \int p(x \mid \theta)\, p(\theta)\, d\theta = E_\theta\left[p(X \mid \theta)\right] = p(x)$ (Pinho, 2006).

Desta forma, a inferência Bayesiana fornece uma estrutura mais natural para avaliar as incertezas na estimativa dos parâmetros do que a frequentista. Uma medida direta de incerteza pode ser obtida através da variância da distribuição a *posteriori* de θ (Garthwaite et al, 1995).

O indicador mais apropriado para essa incerteza é denominado intervalo de credibilidade que é o equivalente bayesiano para o intervalo de confiança na teoria frequentista. O intervalo de credibilidade é construído com base na distribuição a *posteriori* de θ. Como o parâmetro θ é um objeto aleatório, o intervalo de credibilidade é aquele no qual θ está inserido com probabilidade (1- α), em que α é o tamanho do erro tipo I pré - estabelecido (Fernandes, 2009).

A partir do conhecimento sobre θ pode-se descrever sua densidade por uma forma funcional particular. Um procedimento de decisão sequencial é denotado por d = (τ,δ). O

primeiro componente, τ, é chamado de "*regra de parada*", e consiste de funções τ_0, $\tau_1(x^1)$, $\tau_2(x^2)$,..., em que $\tau_i(x^i)$ é a probabilidade de parar a amostragem e tomar a decisão de aceitar ou rejeitar a hipótese após x^i ser observado. As observações X_1, X_2,..., X_n são representadas por $X^n = (X_1, X_2,..., X_n)$ que tem densidade $f_n(x^n| \theta)$ e $\pi(\theta)$ denota a densidade *a priori*. O segundo componente, δ, é chamado de "*regra de decisão*" e consiste de uma série de funções de decisão δ_0, $\delta_1(x^1)$, $\delta_2(x^2)$,..., em que $\delta_i(x^i)$ é a ação a ser tomada se a amostragem parar após observar x^i. O tempo de parada é simplesmente o tamanho da amostra até que a regra τ defina parar e tomar uma decisão (Berger, 1985).

Se *n* observações são tomadas sequencialmente então a perda sequencial quando θ é o verdadeiro estado da natureza será denotada por *Perda*(θ,δ,n), ou seja, o valor da perda neste caso é dependente da regra de decisão e do tamanho amostral. A esperança da perda é chamada função risco de um procedimento sequencial e é dada por:

$$R(\theta,d) = P(N=0)\,Perda(\theta,\delta_0,0) + \sum_{n=1}^{\infty} \int_{\{N=n\}} Perda(\theta,\delta_n(x^n),n)dF_n\left(x^n \mid \theta\right) + \sum_{n=1}^{\infty}\left(\sum_{i=1}^{n} c_i\right) P_\theta\left(N=n\right)$$

Em que c_i é o custo associado a i-ésima observação (Berger, 1985).

O risco de Bayes de um procedimento sequencial d é definido por $r(\pi,d) = E^{\pi}\left[R(\theta,d)\right]$, ou seja, a esperança do risco. Uma regra de decisão d* é ótima se tem risco mínimo, isto é, R(d*) < R(d), $\forall$d. Esta regra será denominada regra de Bayes e seu risco, risco de Bayes, então definido por $r(\pi) = \inf_d r(\pi,d)$, sendo assim o menor entre todos os riscos, dada *a priori* π (Garthwaite et al, 1995).

A ideia 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). A Regra de Bayes Sequencial pode ser chamada de aprendizado Bayesiano. A "*posteriori passada*" se comporta como uma "*priori presente"* e é atualizada a partir da observação de dados atuais,ou seja,

$$\pi(\theta \mid y_1) \propto f(y_1 \mid \theta)\pi(\theta) \text{ e}$$
$$\pi(\theta \mid y_1,\ldots,y_m) \propto f(y_m \mid \theta)\pi(\theta \mid y_1,\ldots,y_{m-1}), m = 2,\ldots,n.$$

Dentre as muitas abordagens para a amostragem sequencial Bayesiana, o procedimento look-ahead, considera o risco de Bayes para o tamanho fixo m > 0, com observações X_{n+1}, X_{n+2},...., X_{n+m}, *priori* π^n e *Perda*(θ,δ, n+m), em que θ é o verdadeiro estado

da natureza, δ é a ação escolhida e m é o número adicional de observações tomadas. O risco de Bayes de uma decisão imediata é então $r_0\left(\pi^n, n\right) = \inf r_0\left(\pi^n, \delta, n\right)$. Se X_{n+m} é tomada, então o menor risco de Bayes esperado é denotado por $r^m(\pi^n, n)$. Para m = 1, este será dado por $r^1(\pi^n, n)$ e o procedimento será denominado "*look ahead* a um passo" (Berger, 1985 e Garthwaite et al, 1995, Brighenti et al, 2011).

O estimador de Bayes de θ com respeito à função perda é aquele com menor risco de bayes. No caso da função perda quadrática $L(\theta,\delta, n) = (\theta - \delta)^2$ o estimador de Bayes para o parâmetro θ será a média de sua distribuição atualizada (Berger, 1985).

O resultado clássico em análise Bayesiana mostra que o menor risco de Bayes *a posteriori* é a variância da distribuição *a posteriori*, denotada por var_{post} (n) (Raiffa & Schlaifer, 1961). Daí,

$$r_o(\pi^n, n) = var_{post}(n).$$

A quantidade $r(\pi^n, n)$ representa o menor risco de Bayes que pode ser atingido uma vez observado $\mathbf{x}^n$. Para decidir se deve ou não parar a amostragem, deve-se comparar o risco de uma decisão imediata com o risco de continuar a amostragem. 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 (Pham-Gia, 1998):

$$r^1(\pi^n, n) = E_x[var_{post}(n)].$$

Para um procedimento look-ahead a um passo, o tempo de parada será então o primeiro n para o qual

$$r_o(\pi^n, n) \leq r^1(\pi^n, n).$$

Assim, a regra de parada é dada em função do risco. Deve-se parar no primeiro n tal que o risco esperado *a posteriori* seja maior ou igual a o risco em 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 à *posteriori* com n observações) de tomar uma decisão (Garthwaite et al, 1995).

2.2.1. Distribuições *a priori*

A informação *a priori* que se pretende incorporar na análise é a informação anterior possuída por alguém, que se identifica como "especialista" do problema concreto – seja ele

investigador, estatístico ou demais interessados – e contém elementos subjetivos. Estes elementos são muitas vezes baseados em fontes objetivas tais como dados históricos do problema, problemas análogos ou fatos. No entanto, tais fontes podem ser influenciadas pela formação do especialista de acordo com suas crenças sobre os aspectos desconhecidos ou incertos, necessários na modelagem do problema.

A quantificação probabilística destas crenças *a priori*, especifica a chamada distribuição *a priori*, e exige, por isso, conceitos de probabilidade que são estranhos à Inferência Clássica.

Especificar probabilidades subjetivas que representem os juízos dos especialistas é uma tarefa difícil. Esta dificuldade costuma ser minimizada através da eliciação cuidadosa de resumos do comportamento das distribuições e do seu manuseio por algum método. O método seguramente mais usado consiste na adoção de uma forma distribucional conveniente e coerente com os resumos eliciados, destacando-se neste papel a denominada família das distribuições conjugadas (Paulino et al., 2003).

O método preditivo de eliciação propõe uma maneira de resolver a questão supondo que a distribuição *a priori* pertence a uma família paramétrica de distribuições, resumindo-se então o problema a escolher os hiperparâmetros que resultem numa melhor adaptabilidade. Esta solução é particularmente fácil de implementar para certas famílias como é o caso da chamada família de distribuições conjugadas naturais, devido ao fato de a distribuição preditiva resultante ter em geral uma forma conhecida (Paulino et al., 2003).

2.2.2. Inferência Bayesiana para dados de contagem

O caso mais importante é o das distribuições conjugadas, quando *a priori* e *posteriori* pertencem à mesma classe de distribuições, assim a atualização do conhecimento que se tem de θ envolve apenas uma mudança nos hiperparâmetros (Negrão et al., 2001).

A distribuição de Poisson é uma distribuição de probabilidade discreta que expressa a probabilidade de uma série de eventos ocorrer num certo intervalo se estes eventos ocorrem independentemente de quando ocorreu o último evento. A sua função de distribuição é dada pela expressão:

$$P\left(X = x\right) = \frac{e^{-\theta}\theta^{x}}{x!} \quad \text{x=1,2,...}$$

Em que e é o número de Euler com valor aproximado de 2,718 e θ>0 é o parâmetro da distribuição e pode representar a média ou taxa de ocorrência de um evento em um dado intervalo de tempo, área ou volume (Andrade & Ogliari, 2007).

Após tomadas n observações de uma Poisson, x_1, x_2, x_3, ..., x_n, a função de verossimilhança será dada por (Andrade & Ogliari, 2007):

$$L(x \mid \theta) \propto e^{-n\theta}\theta^{\sum_{i=1}^{n} x_i} \quad \text{x=0,1,2,...}$$

A função de verossimilhança será dada por:

$$L(\theta \mid x) = L(x \mid \theta) = \prod_{i=1}^{n} P(x = x_i) \propto \prod_{i=1}^{n} e^{-\theta}\theta^{x_i} = e^{-n\theta}\theta^{\sum_{i=1}^{n} x_i} = e^{-n\theta}\theta^{t} \text{ em que } t = \sum_{i=1}^{n} x_i$$

assim $L(\theta \mid x) \sim Gama(t+1, n)$.

A distribuição Gama é uma distribuição de probabilidade contínua, com dois parâmetros α (parâmetro de forma) e β (parâmetro de escala), dos quais se exige α > 0 e β > 0; cuja função de densidade para valores x > 0 é:

$$f(x) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$$

Onde Γ é a função gama dada por $\Gamma(x) = \int_0^{\infty} y^{x-1} e^{-y} dy$ (Negrão et al.,2001).

Se α for inteiro então a distribuição Gama é denominada distribuição de Erlang. No caso particular de α igual a 1, a distribuição Gama se reduz a uma exponencial.

Assim, a família de distribuições conjugadas naturais da Poisson é a distribuição Gama. Considerando a família Gama como distribuição *a priori* π(θ) para o parâmetro da Poisson, tem-se a *priori* para θ dada por (Migon, 2006):

$$\text{Se } \theta \sim \text{Gama}(\alpha, \beta), \textit{então } \pi(\theta) \propto \theta^{\alpha-1} \mathrm{e}^{-\beta\theta} .$$

A distribuição *a posteriori* de θ|x segue uma distribuição Gama com parâmetros (α+*t*) e (β+*n*), isto é, $\theta \mid x \sim Gama(\alpha + t, \beta + n)$, pois:

$$\pi(\theta \mid x) \propto L(x \mid \theta)\pi(\theta) = \left(e^{-n\theta}\theta^{t}\right)\left(e^{-\beta\theta}\theta^{\alpha-1}\right) = e^{-(n+\beta)\theta}\theta^{(\alpha+t)-1}$$

2.3. Infestação de galhas de psilídeos em alecrim do campo

O *Baccharis dracunculifolia* é uma planta arbustiva que ocorre no Cerrado brasileiro é popularmente conhecida como alecrim-do-campo ou vassourinha (Fagundes et al., 2005). Em algumas plantas de alecrim são encontradas as chamadas galhas, que 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 dracunculifolia* é um inseto galhador que ataca exclusivamente espécies de *Baccharis*, nas quais induz uma galha foliar elíptica (Burckhardt et al. 2004). 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 colmeia, evitando predadores e oscilações térmicas, mantendo livres de microrganismos e agentes patogênicos (Brighenti et al., 2006).

Para se estimar a infestação de plantas atingidas por galhas de 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 & 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 & Aquino, 2001). 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 taxa de 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 & Ogliari, 2007).

3. Metodologia

3.1. Desenvolvimento da distribuição *a posteriori*

A distribuição *a posteriori* do processo de Poisson inicialmente foi obtida a partir da verossimilhança e da *priori* para θ (Mood et al., 1974).

Sejam X_1, X_2,..., observações tomadas sequencialmente de uma distribuição Poisson com média θ. A família de distribuições conjugadas é a distribuição Gama (α,β) e a *priori* conjugada para θ será:

$$\pi(\theta \mid \alpha, \beta) \propto \theta^{\alpha-1} e^{-\beta\theta}, \quad \theta > 0$$

Após tomadas n observações, $x_n = x_1, x_2, x_3, ..., x_n$, a distribuição a *posteriori* é dada por:

$$\pi^n(\theta \mid x_n) \propto L(x_n \mid \theta).\pi(\theta \mid \alpha, \beta)$$

Então

$$\pi^n(\theta \mid x_n) \propto \left\{ \sum_{i=1}^{n} \theta^{x_i} e^{-\theta} \right\}.\theta^{\alpha-1} e^{-\beta\theta}$$

$$\pi^n(\theta \mid x_n) \propto \theta^{\alpha+\sum x_i - 1} e^{-(\beta+n)\theta}$$

$$\pi^n(\theta \mid x_n) \propto \theta^{\alpha'-1} e^{-(\beta')\theta}$$

em que $\alpha' = \alpha + \sum x_i$ e $\beta' = \beta + n$. Então a distribuição *a posteriori* é também uma distribuição Gama (α',β') (Garthwaite, 1995).

Utilizando a distribuição *a posteriori* após n observações foi calculada a variância *a posteriori* a partir da variância *a priori*,dado que, para a distribuição *a priori* Gama (α, β), a média é dada por $\mu_{priori} = \frac{\alpha}{\beta}$ e sua variância $\text{var}_{priori} = \frac{\alpha}{\beta^2}$. Para a distribuição *a posteriori* de θ dada pela distribuição Gama com parâmetros $\alpha' = \left(\alpha + \sum_{i=1}^{n} x_i\right), \beta' = (\beta + n)$ médios e a variância serão:

$$\mu_{post} = \frac{\alpha'}{\beta'} = \frac{\alpha + \sum_{i=1}^{n} x_i}{(\beta + n)} \qquad \text{e} \qquad \text{var}_{post} = \frac{\alpha'}{\beta'^2} = \frac{\left(\alpha + \sum_{i=1}^{n} x_i\right)}{(\beta + n)^2},$$

dependentes do tamanho da amostra n.

3.2. Critério de Parada da Amostragem Sequencial

Para a obtenção do critério de parada, foi considerada a função perda quadrática $Perda = \left(\theta - \hat{\theta}\right)^2$ +C(n), em que C(n) > 0 é o custo de tomar uma amostra de tamanho n. Para estimar θ, foi calculada a esperança da variância *a posteriori* a qual permite estimar a quantidade $r_0(\pi^n, n)$ que representa o menor risco de Bayes que pode ser atingido uma vez que x^n foi observado.

No caso do processo de Poisson, o risco de Bayes de tomar uma decisão após observar n amostras, ou risco imediato, será dado em função da esperança da perda quadrática da seguinte maneira:

$$r_0\left(\pi^n, n\right) = E\left[\left(\theta - \hat{\theta}\right)^2\right] = \text{var}_{post}(n) = \frac{\alpha'}{\beta'^2} + \sum_{i=1}^{n} c_i \ .$$

Para decidir se deve ou não parar a amostragem, deve-se então comparar o risco de uma decisão imediata r_0 com o risco de continuar a amostragem.

Supondo que seja necessário tomar exatamente m observações adicionais, o risco de Bayes neste procedimento sequencial bayesiano *look-ahead* a m passos, ou risco esperado a *posteriori*, é:

$$r^m\left(\pi^n, n\right) = E\left[var_{post}(n)\right] = E\left[\frac{\alpha' + \sum_{i=n+1}^{n+m} x_i}{(\beta' + m)^2} \mid x_n\right] + \sum_{i=1}^{n+m} c_i$$

$$r^m\left(\pi^n, n\right) = \frac{\alpha' + \sum_{i=n+1}^{n+m} E\left[x_i \mid x_n\right]}{(\beta' + m)^2} + \sum_{i=1}^{n+m} c_i$$

Como $E[x_i|\theta] = \theta$ e $E[\theta|x_n] = \frac{\alpha'}{\beta'}$, após tomadas as n primeiras observações, e $E[x_i|x_n] = \frac{\alpha'}{\beta'}$ para i = n+1, n+2, ...

Então: $$r^m(\pi^n, n) = \frac{\alpha' + m(\alpha'/\beta')}{(\beta' + m)^2} + \sum_{i=1}^{n+m} c_i$$

Sendo o critério de parada para um procedimento look-ahead a um passo o primeiro n para o qual $r_0(\pi^n, n) \leq r^1(\pi^n, n)$ e considerando que o custo de tomar cada observação seja igual a c em todos os casos, a regra é interromper o processo de parada após observar x_n se, e somente se,

$$\frac{\alpha'}{\beta'^2} + nc \leq \frac{\alpha' + (\alpha'/\beta')}{(\beta' + 1)^2} + (n+1)c$$

Fazendo $var_{post}(n) = \frac{\alpha'}{\beta'^2}$, tem-se:

$$var_{post}(n) + nc \leq \frac{(\beta')}{(\beta'+1)} var_{post}(n) + (n+1)c$$

$$var_{post}(n) - \frac{(\beta')}{(\beta'+1)} var_{post}(n) \leq c$$

$$var_{post}(n)\left[1 - \frac{(\beta')}{(\beta'+1)}\right] \leq c$$

Assim, tenha θ *a priori* Gama (α, β) e seja a função perda $L(\theta, a, n) = (\theta - a)^2 + \sum_{i=1}^{n} c_i$, em que $\sum_{i=1}^{n} c_i = C(n) > 0$ é o custo de tomar uma amostra de tamanho n. Se $C(n+1) - C(n) \geq \gamma$ para todo n para uma constante positiva γ, então o critério de parada para a estimação de θ, o parâmetro da Poisson, é dada por $var_{post}(n)\left[1 - \frac{(\beta')}{(\beta'+1)}\right] \leq c$.

3.3.Simulação da Estimação Sequencial Bayesiana

Para estimar o parâmetro da distribuição de Poisson, 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 à distribuição *a posteriori.*

Para obtenção dos resultados e validação do método proposto foi elaborada uma rotina no programa R 2.13.1 (Anexo A) (R Development Core Team, 2011). Foram geradas sequencialmente amostras de uma população Poisson, seguindo os valores paramétricos fixados em θ = 0.1, 1, 5, 10, mantendo-se um truncamento do tamanho amostral no procedimento em $n<1000$. Deve-se ressaltar que no software R, a parametrização da distribuição Gama equivale, neste trabalho, a seguinte associação: α é o parâmetro de forma e $1/\beta$ o parâmetro de escala.

Foram utilizadas seis *prioris* diferentes para cada valor paramétrico simulado. Para definição dos hiperparâmetros da *priori* Gama adotou-se como critério a utilização de valores selecionados de acordo com a media e a variância da *priori* resultante. Assim, utilizou-se seis combinações entre 3 valores de media e 2 valores de variância. Para a media, utilizou-se valores de α e β que fornecessem media igual ao valor paramétrico e outros dois valores sendo um superior e outro inferior ao valor paramétrico. Para a variância da *priori* foram utilizados em cada simulação: (i) valor 10 vezes menor que a média; (ii) valor 2 vezes maior que a média. Em relação ao custo foram adotados três entre os valores 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}, de acordo com o critério de truncamento.

Após o processo de simulação da população e definição da *priori* e custo, foram calculados o risco imediato e o risco esperado de acordo com cada valor obtido sequencialmente. Os valores eram comparados e gerada nova amostra até que $r_o(\pi^n, n) \leq r^1(\pi^n, n)$. A partir da distribuição a *posteriori* o valor estimado do parâmetro da população foi obtido, de acordo com a média da distribuição a *posteriori* obtida até então, assim como a variância da *posteriori* e o intervalo de credibilidade a 95%.

Para estudar via simulação a influência do método e do grau de informatividade da distribuição *a priori* realizou-se a estimação sequencial Bayesiana de acordo com o seguinte algoritmo:

<u>Passo 1</u>

Definir valores do(s):

- hiperparâmetros (α, β) para estabelecimento da distribuição *a priori*,

- custo (C),
- tamanho limite da amostra (N)

Passo 2

Gerar um único valor x_i da distribuição de Poisson com o valor paramétrico θ desejado.

Passo 3

Calcular os parâmetros da distribuição *a posteriori* a partir da soma dos valores obtidos $\sum_{i=1}^{n} x_i$ até a amostragem n.

$$\alpha' = \alpha + \sum_{i=1}^{n} x_i \qquad \beta' = \beta + n$$

Obs.: Os valores de n e $\sum_{i=1}^{n} x_i$ devem ser atualizados a cada amostragem;

Passo 4 → Calcular os riscos imediato e esperado:

$$r_0 = \frac{\alpha'}{\beta'^2} + \sum_{i=1}^{n} c_i \qquad r^1 = \frac{\alpha' + (\alpha'/\beta')}{(\beta'+1)^2} + \sum_{i=1}^{n+1} c_i$$

Passo 5

Regra decisória:

- Se $r^1 < r_0$, continuar a amostragem e atualizar os valores n e $\sum_{i=1}^{n} x_i$ (somar o valor encontrado).
- Se $r^1 > r_0$, interromper a amostragem e realizar o cálculo da média da *posteriori* dada por: $\mu_{post} = \frac{\alpha'}{\beta'}$, que corresponde ao valor da estimativa do parâmetro da Poisson.

Para cada caso de estimação do parâmetro populacional θ = 0,1; 1 ;5 ou 10, foram definidas seis distribuições *a priori* e suas curvas de densidade geradas a partir da rotina em R (Anexo B).

A partir da rotina e das expressões calculadas dos riscos, foi então, elaborado em Delphi, o aplicativo para estimação do parâmetro da distribuição de Poisson. O software foi criado para ser executado na plataforma Microsoft Windows. Além do executável, foi

elaborado para o software um arquivo de configuração de parâmetros iniciais do sistema que envolve as definições de variância e custo, que por estar em formato texto de extensao.ini, possibilita ao usuário abrir o arquivo de configuração e ajustar os parâmetros conforme sua necessidade de ajuste dos valores da priori. O software conta ainda com um arquivo no formato XML, onde o usuário poderá armazenar a memória de cálculo de cada execução para análise futura.

3.4.Coleta de dados reais

Para aplicação da estimação sequencial bayesiana, considerou-se, a variável aleatória X como sendo o número de galhas de psilídeos encontrados em cada planta de alecrim do campo, cujo interesse foi estimar θ, ou seja, a média da quantidade de psilídeos *B. dracunculifoliae* encontrados nas plantas de alecrim do campo, *B. dracunculifolia*, seguindo o modelo de Poisson, determinando-se as áreas mais propicias a instalação de apiários para a produção de própolis verde.

Primeiramente foi realizada a estimação frequentista de θ da distribuição de Poisson para posterior comparação dos resultados com o método bayesiano.

Para verificar o ajustamento de dados a distribuição de Poisson foi realizado um teste Qui quadrado para aderência (Andrade & Ogliari, 2007).

As coletas de dados em campo foram realizadas na cidade de Resende Costa localizada na microrregião de São João Del Rei - MG, com latitude S 21° 08’ 00’’ e longitude W 44°15’40’’ em altitude de 898 m e índice médio pluviométrico anual de 1470 mm.

Foram avaliadas sequencialmente as plantas de alecrim na cidade de Resende Costa. A cada avaliação o número de galhas de psilídeos foi contado e calculados o risco imediato e o esperado. Estes foram comparados e a decisão de continuar a amostragem foi tomada, após interromper a amostragem foi realizada a estimação do parâmetro θ.

4. Resultados e Discussão

4.1. Simulação da Estimação Sequencial Bayesiana

Para estimação do parâmetro da Poisson em que o valor paramétrico foi 0,1 utilizou-se custos iguais a 10^{-3}, 10^{-4} ou 10^{-5}. Os valores dos hiperparâmetros para as seis *prioris* foram calculados pelo sistema de equações obtido supondo a média da *priori* igual ao próprio parâmetro 0,1 ou 0,05 ou ainda 0,5, sendo para a variância atribuído um valor 10 vezes menor que a média da *priori* ou o seu dobro (Figura 1).

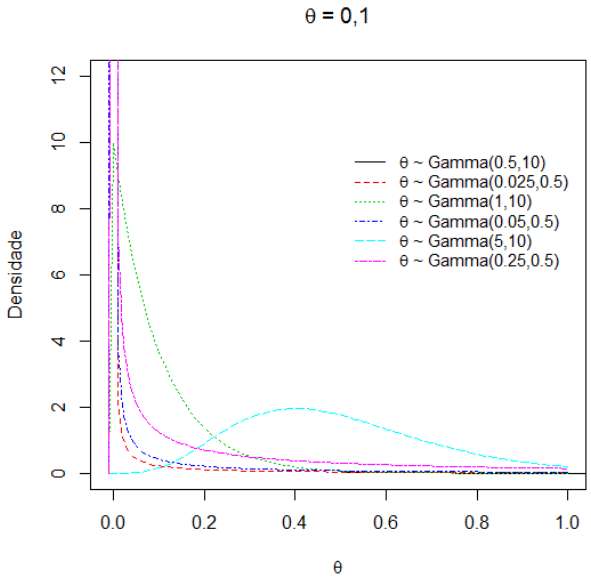


Figura 1 – Densidade para as distribuições *a priori* Gama utilizadas para $\theta = 0{,}1$.

Utilizando as seis *prioris* e os três custos foram então estabelecidos 18 casos para a estimação do $\theta = 0{,}1$ (Tabela 1). Observa-se que em apenas três situações entre as simuladas, o verdadeiro valor do parâmetro não pertence ao intervalo de credibilidade, sendo estes provenientes de simulações em que o custo por observação era igual a 10^{-3}, ou seja, o maior custo utilizado, o que resultou em tamanhos de amostras muito pequenos, em media 5,8 observações, prejudicando a estimação. A exceção para este custo de amostragem se deu quando a média da *priori* coincidia com o valor do parâmetro, evidenciando que, no caso de custo maior, a *priori* teve grande influência.

Quando foi utilizado custo 10^{-4} o tamanho amostral variou entre 7 e 35 observações, sendo a media igual a 22,3. No caso de custo 10^{-5} o tamanho amostral teve maior media, sendo igual a 87,5. Nestes dois casos, pode-se dizer que o estimador teve acurácia, no entanto, no caso de custo menor, obteve-se intervalo de credibilidade com menor amplitude e estimativas para o parâmetro populacional mais precisas.

A partir dos valores de α e β obtidos para a distribuição *a posteriori*, pode-se elaborar os gráficos comparativos entre as densidades da *priori*, verossimilhança e *posteriori*. Nas Figuras 2 e 3 a seguir, encontra-se um modelo de gráficos obtidos, incluindo um caso que o intervalo de credibilidade continha o parâmetro e outro que não continha.

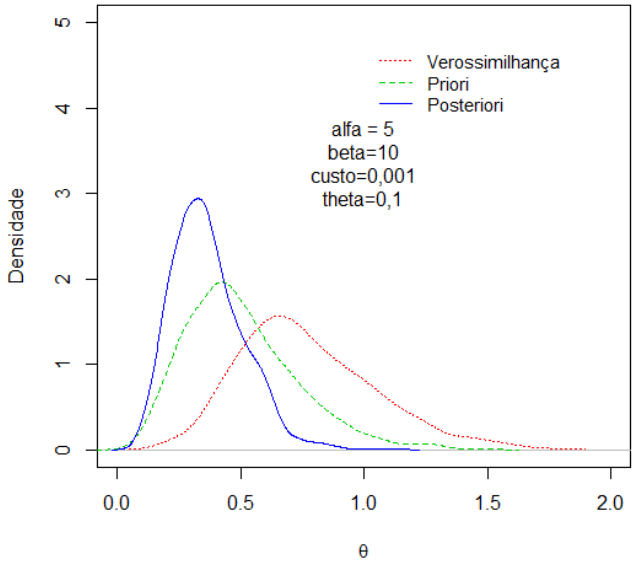


Figura 2 – Distribuição *a priori* de θ, verossimilhança e distribuição *a posteriori* de θ|x utilizadas para θ = 0,1 em que θ não pertence ao $ICr_{95\%}$.

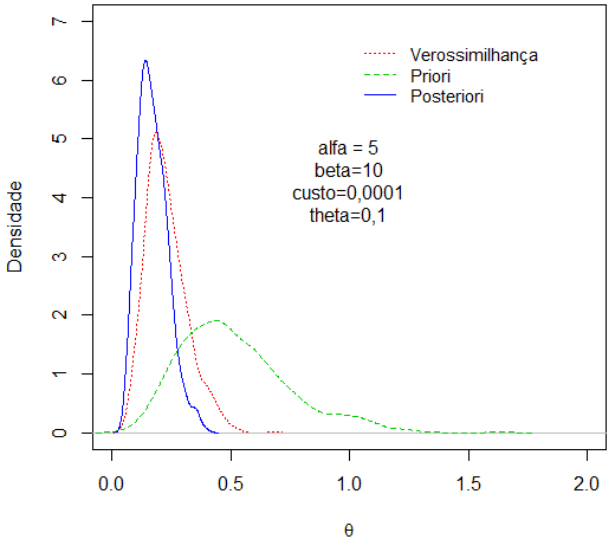


Figura 3 – Distribuição *a priori* de θ, verossimilhança e distribuição *a posteriori* de θ|x utilizadas para θ = 0,1 em que θ pertence ao $ICr_{95\%}$.

Tabela 1 – Estimativas Bayesianas obtidas por amostragem sequencial via simulação, para o parâmetro da distribuição de Poisson, considerando θ igual a 0,1, em que: n = tamanho da amostra, $IC_{95\%}$ = Intervalo de Credibilidade a 95%.

Custo	Parâmetros da *priori* Alfa	Parâmetros da *priori* Beta	Média da *priori*	Variância da *priori*	n	Parâmetros da *posteriori* Alfa'	Parâmetros da *posteriori* Beta'	Média da *posteriori*	Variância da *posteriori*	$ICr_{95\%}$	Amplitude do $ICr_{95\%}$
10^{-3}	0,5	10	0,05	0,005	1	0,05	11	**0,005**	0,0004	**[4,83.10^{-34};0,055]**	0,055
	0,025	0,5		0,1	3	0,025	3,5	**0,007**	0,0020	**[1,35.10^{-65};0,075]**	0,075
	1	10	0,1	0,01	1	1	11	0,091	0,0083	[0,0025; 0,3350]	0,332
	0,05	0,5		0,2	16	4,05	16,5	0,245	0,0149	[0,0670; 0,5390]	0,472
	5	10	0,5	0,05	8	6	18	**0,333**	0,0185	**[0,1220; 0,6480]**	0,526
	0,25	0,5		1	6	0,25	6,5	0,038	0,0059	[4,05 10^{-8}; 0,264]	0,648
10^{-4}	0,5	10	0,05	0,005	7	0,5	17	0,029	0,0017	[2,88.10^{-5}; 0,147]	0,147
	0,025	0,5		0,1	31	3,025	31,5	0,096	0,0030	[0,0199; 0,2306]	0,211
	1	10	0,1	0,01	17	2	27	0,074	0,0027	[0,0089; 0,2063]	0,197
	0,05	0,5		0,2	31	3,05	31,5	0,096	0,0031	[0,0203; 0,2318]	0,212
	5	10	0,5	0,05	35	9	45	0,200	0,0044	[0,0914; 0,3502]	0,259
	0,25	0,5		1	13	0,25	13,5	0,018	0,0014	[1,95.10^{-8}; 0,127]	0,127
10^{-5}	0,5	10	0,05	0,005	101	13,5	111	0,121	0,0011	[0,0656; 0,1945]	0,129
	0,025	0,5		0,1	96	9,025	96,5	0,093	0,0010	[0,0428; 0,1636]	0,121
	1	10	0,1	0,01	64	4	74	0,054	0,0007	[0,0147; 0,1184]	0,104
	0,05	0,5		0,2	74	4,05	74,5	0,054	0,0007	[0,0149; 0,1186]	0,104
	5	10	0,5	0,05	110	17	120	0,141	0,0012	[0,0825; 0,2165]	0,134
	0,25	0,5		1	80	5,25	80,5	0,065	0,0008	[0,0219; 0,1317]	0,109

Para simulação da estimação dos demais parâmetros da Poisson utilizados, considerou-se os custos iguais a 10^{-2}, 10^{-3} ou 10^{-4}. No caso em que se considerou o parâmetro populacional igual a 1, os valores dos hiperparâmetros foram calculados pelo sistema de equações obtido supondo a média da *priori* igual ao próprio parâmetro 1 ou a metade do parâmetro (0,5) ou ainda um valor duas vezes maior (2), sendo para a variância atribuído um valor 10 vezes menor que a média da *priori* ou o seu dobro (Figura 4).

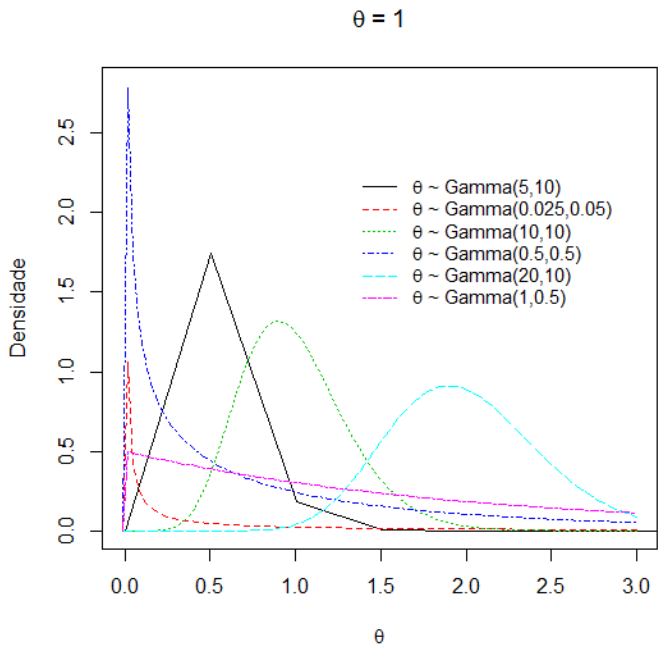


Figura 4 – Curvas de densidade para as distribuições *a priori* Gama utilizadas para $\theta = 1$.

Em apenas dois destes, o intervalo de credibilidade não continha o verdadeiro valor do parâmetro (Tabela 2). No custo 10^{-2}, apesar do pequeno tamanho amostral com media igual a 5 observações, a estimação foi adequada em todos os casos, não sendo verificada nenhuma associação entre o tamanho amostral e a amplitude do intervalo de credibilidade. O tamanho amostral para os custos iguais a 10^{-3} e 10^{-4} apresentaram em media valores de 23,5 e 93,85 observações respectivamente. No caso θ igual a 1 não foi possível identificar influência da *priori* e nem do custo; os casos em que o intervalo de credibilidade não contém o valor real não seguiram um padrão. Nas Figuras 5 e 6, encontra-se um modelo de gráficos obtidos, incluindo um caso que o intervalo de credibilidade continha o parâmetro e outro que não continha.

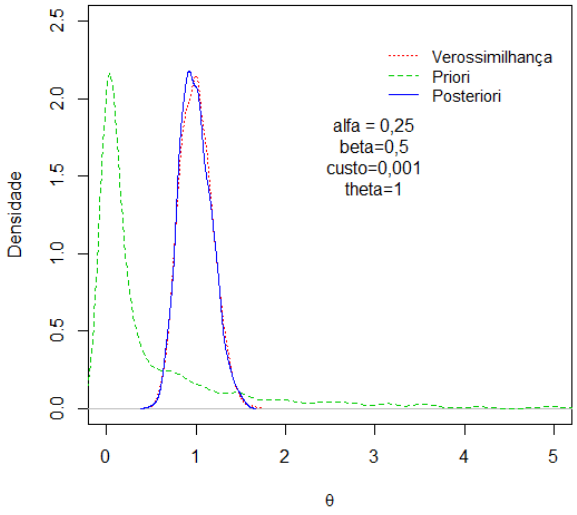


Figura 5 – Distribuição *a priori* de θ, verossimilhança e distribuição *a posteriori* de θ|x utilizadas para θ = 1 em que θ não pertence ao $ICr_{95\%}$.

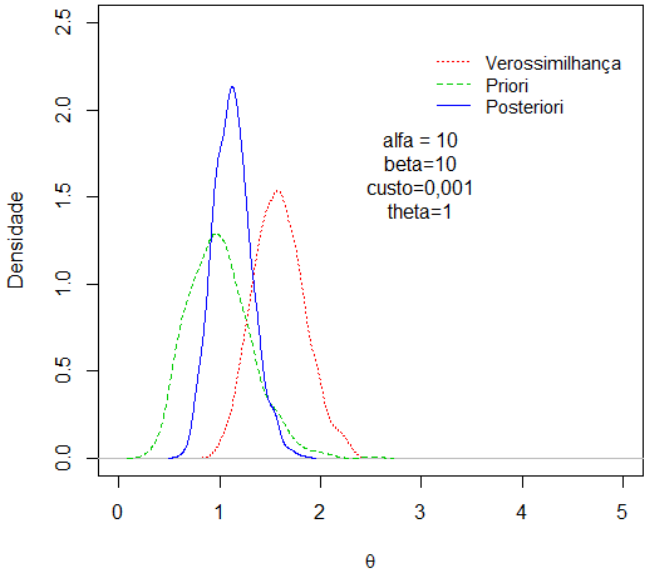


Figura 6 – Distribuição *a priori* de θ, verossimilhança e distribuição *a posteriori* de θ|x utilizadas para θ = 1 em que θ pertence ao $ICr_{95\%}$.

Tabela 2 – Estimativas Bayesianas obtidas por amostragem sequencial via simulação, para o parâmetro da distribuição de Poisson, considerando θ igual a 1, em que: n = tamanho da amostra, $IC_{95\%}$ = Intervalo de Credibilidade a 95%.

Custo	Parâmetros da *priori*		Média da *priori*	Variância da *priori*	n	Parâmetros da *posteriori*		Média da *posteriori*	Variância da *posteriori*	$ICr_{95\%}$	Amplitude do $ICr_{95\%}$
	Alfa	Beta				Alfa'	Beta'				
10^{-2}	5	10	0,5	0,05	1	6	11	0,545	0,0496	[0,200; 1,060]	0,860
	0,25	0,5		1	10	12,25	10,5	1,167	0,1111	[0,607; 1,905]	1,298
	10	10	1	0,1	1	12	11	1,091	0,0992	[0,563; 1,789]	1,226
	0,5	0,5		2	7	4,5	7,5	0,600	0,0800	[0,180; 1,268]	1,088
	20	10	2	0,2	1	22	21	1,048	0,0499	[0,656; 1,528]	0,872
	1	0,5		4	10	12	10,5	1,143	0,1088	[0,590; 1,874]	1,284
10^{-3}	5	10	0,5	0,05	17	19	27	0,704	0,0261	[0,424; 1,053]	0,629
	0,25	0,5		1	14	14	24	**0,583**	0,0243	**[0,318; 0,926]**	**0,608**
	10	10	1	0,1	23	36	33	1,091	0,0331	[0,764; 1,475]	0,711
	0,5	0,5		2	28	22,5	28,5	0,789	0,0277	[0,497; 1,147]	0,650
	20	10	2	0,2	23	36	33	1,091	0,0331	[0,764; 1,475]	0,711
	1	0,5		4	36	47	36,5	1,288	0,0353	[0,946; 1,681]	0,735
10^{-4}	5	10	0,5	0,05	94	113	104	1,087	0,0104	[0,895; 1,295]	0,400
	0,25	0,5		1	94	84,25	94,5	0,892	0,0094	[0,711; 1,091]	0,380
	10	10	1	0,1	86	88	96	0,917	0,0095	[0,735; 1,117]	0,382
	0,5	0,5		2	89	71,5	89,5	**0,799**	0,0089	**[0,624; 0,994]**	**0,370**
	20	10	2	0,2	97	121	107	1,131	0,0106	[0,938;1,341]	0,403
	1	0,5		4	103	111	103,5	1,072	0,0104	[0,882; 1,280]	0,398

Para estimação do parâmetro da Poisson igual a 5, os valores dos hiperparâmetros foram calculados considerando a média da *priori* como os valores 1, 5 ou 25, sendo que para a variância, atribuiu-se um valor 10 vezes menor que a média da *priori* ou o seu dobro (Figura 7).

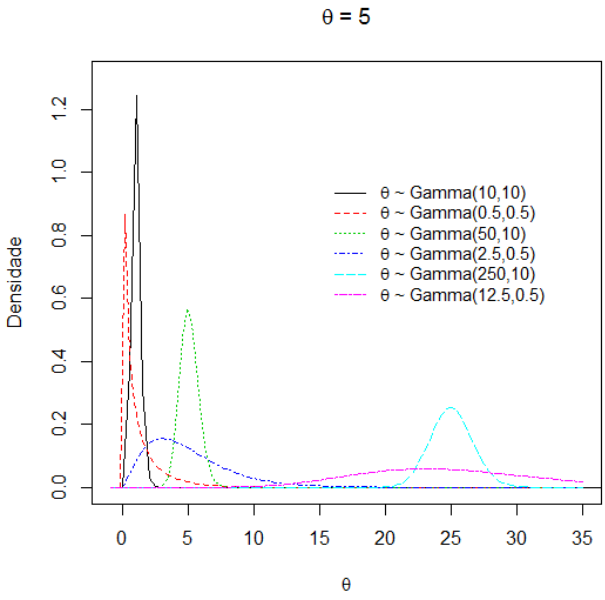


Figura 7 – Densidade para as distribuições *a priori* Gama utilizadas para $\theta = 5$.

Em quatro dos 18 casos simulados, o verdadeiro valor do parâmetro não pertence ao intervalo de credibilidade (Tabela 3), sendo que três deles ocorreram quando utilizou-se a *priori* Gama (250, 10), ou seja, provenientes de uma média da *priori* alta igual a 25 e variância pequena, assim, independente do custo por amostragem, a *priori* teve alto grau de informatividade quando a sua média foi distante do valor real do parâmetro e sua variância pequena. Portanto, a *priori* Gama (250, 10) influenciou negativamente na tomada de decisão, mesmo com o tamanho amostral elevado. O quarto intervalo de credibilidade que não continha o verdadeiro valor do parâmetro foi aquele que a *priori* utilizada foi a distribuição Gama (10,10), sendo o intervalo de credibilidade variando de 0,696 a 2,020, isso ocorreu pois o processo foi interrompido já na primeira amostra. Este fato evidencia que há necessidade de definição de um tamanho amostral mínimo e não somente um limite final para n. Quando o custo foi igual a 10^{-4}, houve em media necessidade de 219,5 observações até que a decisão fosse tomada. No entanto, com o aumento do custo para 10^{-3}, o tamanho

amostral foi reduzido para 67,3 observações em media. As Figuras 8 e 9 apresentam dois casos simulados, um que o intervalo de credibilidade continha o verdadeiro valor do parâmetro e outro não.

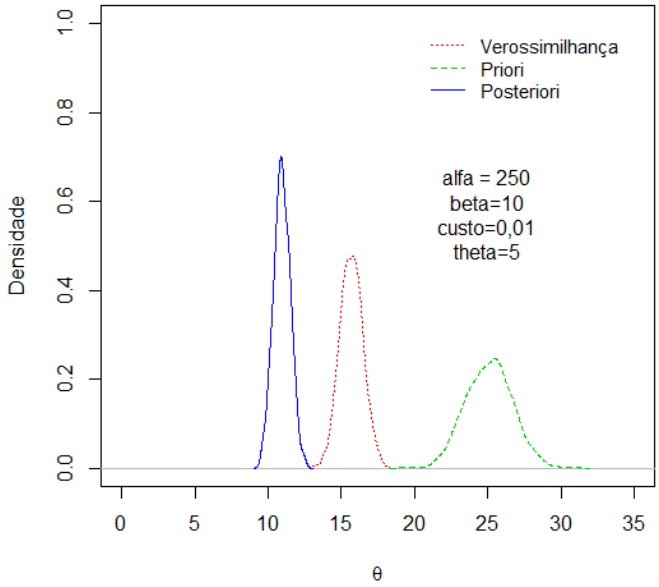


Figura 8 – Distribuição *a priori* de θ, verossimilhança e distribuição *a posteriori* de θ|x utilizadas para θ = 5 em que θ não pertence ao $ICr_{95\%}$.

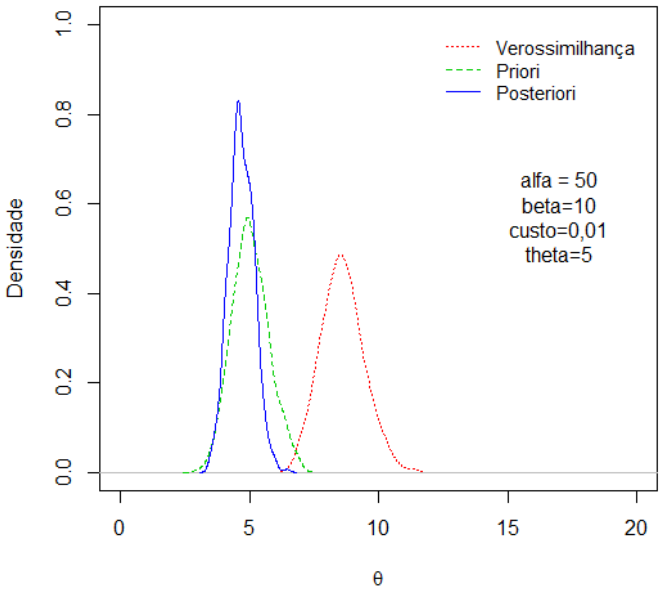


Figura 9 – Distribuição *a priori* de θ, verossimilhança e distribuição *a posteriori* de θ|x utilizadas para θ = 5 em que θ pertence ao $ICr_{95\%}$.

Tabela 3 – Estimativas Bayesianas obtidas por amostragem sequencial via simulação, para o parâmetro da distribuição de Poisson, considerando θ igual a 5, em que: n = tamanho da amostra, $IC_{95\%}$ = Intervalo de Credibilidade a 95%.

Custo	Parâmetros da *priori*		Média da *priori*	Variância da *priori*	n	Parâmetros da *posteriori*		Média da *posteriori*	Variância da *posteriori*	$ICr_{95\%}$	Amplitude do $ICr_{95\%}$
	Alfa	Beta				Alfa'	Beta'				
10^{-2}	10	10	1	0,1	1	14	11	**1,273**	0,1157	**[0,696; 2,020]**	1,324
	0,5	0,5		2	22	113,5	22,5	5,044	0,2242	[4,159; 6,014]	1,855
	50	10	5	0,5	12	110	22	5,000	0,2273	[4,109; 5,977]	1,868
	2,5	0,5		10	22	109,5	22,5	4,867	0,2163	[3,998; 5,819]	1,821
	250	10	25	2,5	23	353	33	**10,697**	0,3242	**[9,610; 11,841]**	2,231
	12,5	0,5		50	22	116,5	22,5	5,178	0,2301	[4,280; 6,159]	1,879
10^{-3}	10	10	1	0,1	56	289	66	**4,379**	0,0663	**[3,888; 4,898]**	1,010
	0,5	0,5		2	69	339,5	69,5	4,885	0,0703	[4,379; 5,148]	0,769
	50	10	5	0,5	59	326	69	4,725	0,0685	[4,226; 5,251]	1,025
	2,5	0,5		10	73	369,5	73,5	5,027	0,0684	[4,876; 5,938]	1,062
	250	10	25	2,5	77	649	87	**7,460**	0,0857	**[6,896; 8,044]**	1,148
	12,5	0,5		50	70	349,5	70,5	4,957	0,0703	[4,451; 5,490]	1,039
10^{-4}	10	10	1	0,1	208	1032	218	4,734	0,0217	[4,449; 5,027]	0,578
	0,5	0,5		2	218	1044,5	218,5	4,780	0,0219	[4,495; 5,074]	0,579
	50	10	5	0,5	214	1121	224	5,004	0,0223	[4,715; 5,301]	0,586
	2,5	0,5		10	222	1105,5	222,5	4,969	0,0223	[4,679; 5,265]	0,586
	250	10	25	2,5	232	1421	242	**5,872**	0,0243	**[5,570; 6,181]**	0,611
	12,5	0,5		50	223	1120,5	223,5	5,013	0,0224	[4,724; 5,311]	0,587

No caso da estimação do parâmetro da Poisson igual a 10, utilizou-se para calculo dos hiperparâmetros, a suposição da media da *priori* igual aos valores 1, 10 ou 100 e para a variância da *posteriori*, sendo um igual ao dobro da media e outro 10 vezes menor (Figura 10).

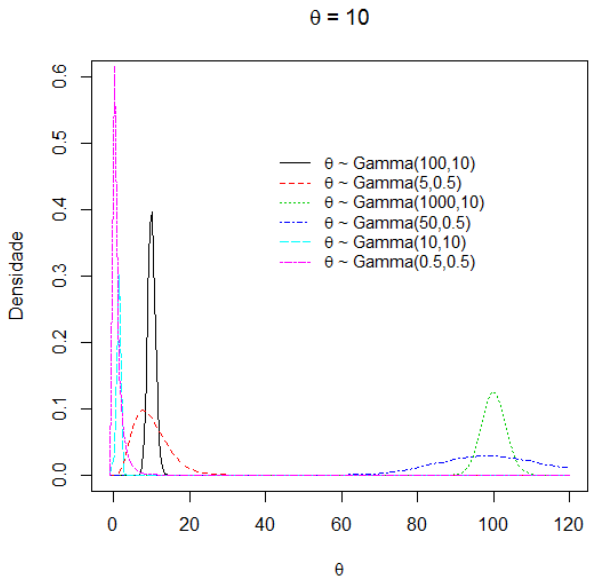


Figura 10 – Densidade para as distribuições *a priori* Gama utilizadas para $\theta = 10$.

Avaliando os intervalos de credibilidade obtidos (Tabela 4), verificou-se que, em 5 deste o parâmetro não pertence ao intervalo, sendo estes intervalos aqueles resultantes das simulações em que a media da *priori* foi considerada muito distante do parâmetro real e com restrição de variância pequena, tal como ocorre no caso da simulação da estimação do parâmetro $\theta = 5$. Deve-se notar ainda que, para tamanhos amostrais altos, media de 29,17 para custo 10^{-2}, de 99,33 para custo 10^{-3} e 317,33 para o custo 10^{-4}, isto não foi o suficiente para melhorar a estimação, quando a *priori* foi muito informativa e distante do valor paramétrico simulado. As Figuras 11 e 12 exemplificam dois casos simulados um que continha o verdadeiro valor paramétrico e outro que não continha.

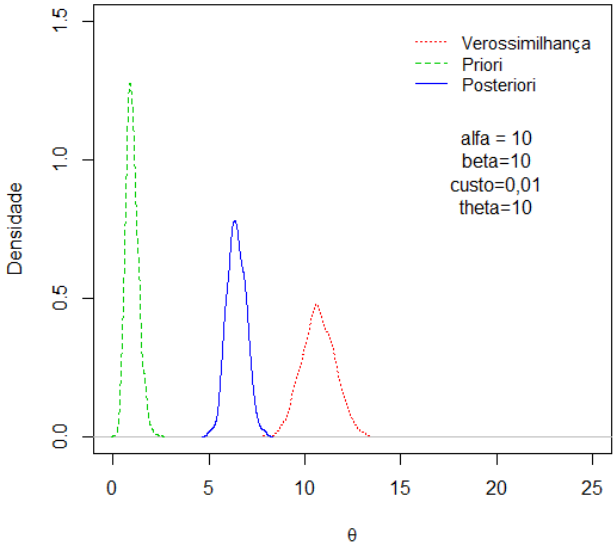


Figura 11 – Distribuição *a priori* de θ, verossimilhança e distribuição *a posteriori* de θ|x utilizadas para θ = 10 em que θ não pertence ao $ICr_{95\%}$.

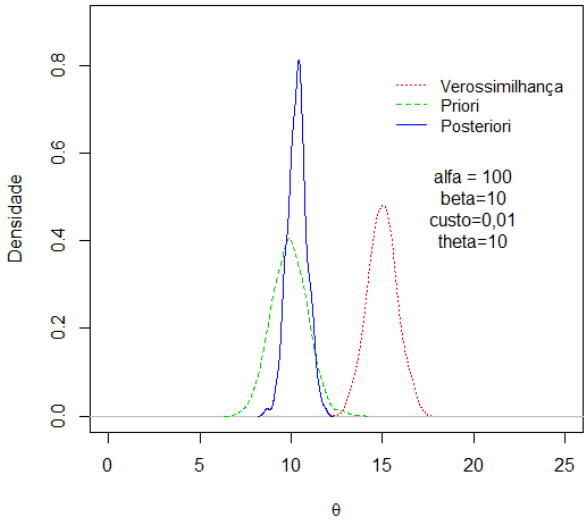


Figura 12 – Distribuição *a priori* de θ, verossimilhança e distribuição *a posteriori* de θ|x utilizadas para θ = 10 em que θ pertence ao $ICr_{95\%}$.

Tabela 4 – Estimativas Bayesianas obtidas por amostragem sequencial via simulação, para o parâmetro da distribuição de Poisson, considerando θ igual a 10, em que: n = tamanho da amostra, $IC_{95\%}$ = Intervalo de Credibilidade a 95%.

Custo	Parâmetros da *priori*		Média da *priori*	Variância da *priori*	n	Parâmetros da *posteriori*		Média da *posteriori*	Variância da *posteriori*	$ICr_{95\%}$	Amplitude do $ICr_{95\%}$
	Alfa	Beta				Alfa'	Beta'				
10^{-2}	10	10	1	0,1	17	191	27	**7,074**	0,2620	**[6,106; 8,112]**	**2,006**
	0,5	0,5		2	30	283,5	30,5	9,295	0,3048	[8,244; 10,408]	2,164
	100	10	10	1	22	320	32	10,000	0,3125	[8,934; 11,125]	2,191
	5	0,5		20	31	313	31,5	9,937	0,3154	[8,866;11,067]	2,795
	1000	10	100	**10**	42	1375	52	**26,442**	0,5085	**[25,063; 27,858]**	**2,269**
	50	0,5		200	33	376	33,5	11,224	0,3350	[10,117; 12,386]	1,209
10^{-3}	10	10	1	0,1	86	876	96	**9,125**	0,0951	**[8,530; 9,739]**	**1,225**
	0,5	0,5		2	98	948,5	98,5	9,629	0,0978	[9,026; 10,251]	1,238
	100	10	10	1	89	978	99	9,879	0,0998	[9,269; 10,507]	1,233
	5	0,5		20	98	959	98,5	9,736	0,0988	[9,129; 10,362]	1,422
	1000	10	100	**10**	122	2292	132	**17,364**	0,1315	**[16,659; 18,081]**	**1,258**
	50	0,5		200	103	1105	103,5	10,676	0,1032	[10,056; 11,314]	0,693
10^{-4}	10	10	1	0,1	302	3042	312	9,750	0,0313	[9,406; 10,099]	0,695
	0,5	0,5		2	315	3133,5	315,5	9,932	0,0315	[9,587; 10,282]	0,699
	100	10	10	1	308	3217	318	10,116	0,0318	[9,769; 10,468]	0,701
	5	0,5		20	320	3289	320,5	10,262	0,0320	[9,914; 10,615]	0,736
	1000	10	100	**10**	342	4367	352	**12,406**	0,0352	**[12,040; 12,776]**	**0,698**
	50	0,5		200	317	3198	317,5	10,072	0,0317	[9,7263; 10,424]	1,209

4.2. Estimação sequencial Bayesiana da contagem de galhas de psilídeo em alecrim do campo

A técnica de estimação sequencial Bayesiana foi aplicada a conjunto de dados reais obtido da contagem de galhas de psilídeos em alecrim na cidade de Resende Costa, microrregião de São João Del Rei.

Foram observadas sequencialmente, plantas do alecrim do campo, verificando-se quantidade de galhas de psilídeo *B. dracunculifoliae* presentes em cada planta (Figura 13 e 14).

Figura 13 – Galha no alecrim do campo

Figura 14 – Psilídeos dentro da galha

Foi considerado o custo por observação igual a 10^{-3}, e o processo como sendo dados de contagem seguindo o modelo Poisson, com hiperparâmetros α=1 e β=1. Os cálculos foram realizados da seguinte maneira:

- 1ª Planta - 0 galhas observadas

Parâmetros da *posteriori*: $\alpha' = 1 \qquad \beta' = 1,5$

Riscos: $r_0 = \frac{1}{(1,5)^2} + 0,001(1) = 0,4454 \qquad r^1 = \frac{1+(1/1,5)}{(1,5+1)^2} + 0,001(1+1) = 0,2687$

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

- 2ª Planta – 2 galhas observadas

Parâmetros da *posteriori*: $\alpha' = 3 \qquad \beta' = 2,5$

Riscos: $r_0 = \frac{3}{(2,5)^2} + 0,001(2) = 0,4820 \qquad r^1 = \frac{3+(3/2,5)}{(2,5+1)^2} + 0,001(1+2) = 0,3458$

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

- 3ª Planta– 0 galhas observadas

Parâmetros da *posteriori*: $\alpha' = 3 \qquad \beta' = 3,5$

Riscos: $r_0 = \frac{3}{(3,5)^2} + 0,001(3) = 0,2479 \qquad r^1 = \frac{3+(3/3,5)}{(3,5+1)^2} + 0,001(1+3) = 0,1945$

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

E assim sucessivamente até a 56ª planta em que;

- 56ª Planta – 6 galhas observadas

Parâmetros da *posteriori*: $\alpha' = 186 \qquad \beta' = 57$

Riscos: $r_0 = \frac{186}{(57)^2} + 0,001(56) = 0,113248$

$$r^1 = \frac{186+(186/57)}{(57+1)^2} + 0,001(1+56) = 0,113261$$

Decisão: Como $r_0 < r_1$ interrompe-se a amostragem

Na 56ª observação realizada com n=56, sendo o parâmetro foi estimado como $\mu_{post} = \frac{\alpha'}{\beta'}$, neste caso, dada por $\mu_{post} = \frac{186}{57} = 3,26$.

Assim, na área de alecrim da cidade de Resende Costa pode-se inferir que a infestação equivale em média a 3,26 galhas de psilídeos por planta.

4.2.1. Estimação Frequentista do Parâmetro da Poisson da contagem de galhas de psilídeo em alecrim do campo

A estimação do parâmetro da Poisson, correspondente a média da quantidade de galhas de *B. dracunculifoliae* encontrados em plantas de alecrim do campo, também foi realizada através do modo frequentista, considerando o valor final de n como o tamanho encontrado no modo sequencial Bayesiano.

O cálculo do valor do parâmetro da distribuição de Poisson (θ), de acordo com os dados observados (Tabela 5) é obtido por:

$$\theta = 0\left(\frac{23}{56}\right) + 1\left(\frac{3}{56}\right) + 2\left(\frac{6}{56}\right) + 3\left(\frac{5}{56}\right) + \ldots + 15\left(\frac{1}{56}\right) = 3,30$$

$$\theta = 3,30$$

A partir do valor de θ foi feito o cálculo da probabilidade de se encontrar 0,1,2,.... galhas de psilídeos por planta. Os valores para cada probabilidade estão na terceira coluna da tabela, calculados conforme o exemplo a seguir.

$$\text{Prob}(Y=0)=\frac{e^{-3,30}(3,30)^{0}}{0!}=0,0368=3,68\%$$

Dentre as plantas observadas, o número esperado de galhas encontrados em cada planta está na quarta coluna da tabela e é dado por:

$$\text{nº esperado (0 galhas)} = 56(0,0368)=2,06$$

Tabela 5 – Resultado da contagem de galhas de psilídeos em plantas de alecrim e seus respectivos valores de probabilidade e número esperado de plantas de acordo com a distribuição de Poisson.

Quantidade de galhas (X)	Nº de plantas observado	Probabilidade de ocorrência de X (%)	Nº de plantas esperado	Desvio (Observado – Esperado)
0	23	3,68	2,06	20,94
1	3	12,14	6,80	-3,80
2	6	20,05	11,23	-5,23
3	5	22,08	12,37	-7,37
4	3	18,24	10,21	-7,21
5	2	12,05	6,75	-4,75
6	2	6,64	3,72	-1,72
7	3	3,13	1,75	1,25
8	1	1,29	0,72	0,28
9	1	0,47	0,27	0,73
10	2	0,16	0,09	1,91
11	1	0,05	0,03	0,97
12	2	0,01	0,01	1,99
13	1	0,00	0,00	1,00
14	0	0,00	0,00	0,00
15	1	0,00	0,00	1,00
Total	56	100	56,00	

Como as frequências esperadas, supondo uma distribuição de Poisson com parâmetro θ = 3,30, para a variável contagem de galhas de psilídeos em plantas de alecrim do campo, são próximas das frequências observadas, de acordo com os pequenos desvios, apresentados na quinta coluna da tabela, conclui que a infestação de

psilídeos *B. dracunculifoliae* em alecrim do campo possivelmente segue uma distribuição de Poisson (Figura 15).

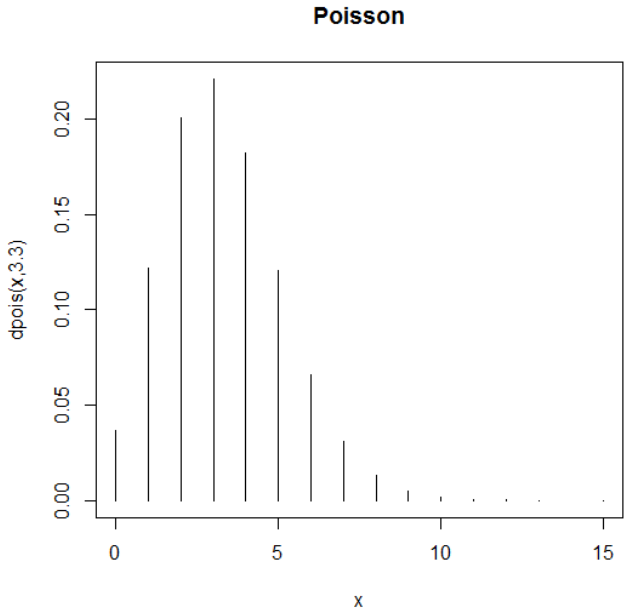


Figura 15 – Distribuição de Poisson com parâmetro 3,30.

Para verificar o ajustamento do conjunto de dados reais a distribuição de Poisson foi utilizando o teste de Qui-quadrado para aderência.

$$\chi^2_{cal} = \frac{\sum(\text{Observado - Esperado})^2}{\text{Esperado}}$$

$$\chi^2_{cal} = \frac{(23\text{-}2)^2}{2} + \frac{(3-7)^2}{7} + \frac{(6-11)^2}{11} + \frac{(5-12)^2}{12} + \frac{(3-10)^2}{10} + \frac{(2-7)^2}{7} + \frac{(2-4)^2}{4} + \frac{(3-2)^2}{2} + \frac{(1-1)^2}{1}$$

$$\chi^2_{cal} = 220,50 + 2,28 + 2,27 + 4,08 + 4,90 + 3,57 + 1,00 + 0,50 + 1$$

$$\chi^2_{cal} = 240,10 \quad \chi^2_{tab(5\%)} = 79,082$$

Como $\chi^2_{cal} > \chi^2_{tab}$, aceita-se a hipótese nula a nível de significância de 5% de que o conjunto de dados da contagem de galhas de psilídeos em alecrim pode ser considerado como proveniente de uma distribuição de Poisson.

4.2.2. Software para Estimação Sequencial Bayesiana

Na primeira sessão do aplicativo encontra-se a entrada de dados da análise e parâmetros para o início da estimação do parâmetro da distribuição de Poisson, tais como as descrições do titulo, data da amostragem, objeto de estudo, contagem esperada, variação e custo, sendo que, cada tópico apresenta uma janela com opção de ajuda para esclarecimento ao usuário sobre a utilização de tais parâmetros.

A contagem esperada e variação incluídas pelo usuário são utilizadas para definição dos parâmetros da distribuição *a priori*, a partir de um método preditivo de eliciação. Como a média é dada por $\mu = \frac{\alpha}{\beta}$ e a variância é $\text{var} = \frac{\alpha}{\beta^2}$, então, considerando a contagem esperada inicial como sendo o valor da média *a priori* e a variação como Var, tem-se que $\alpha = \frac{\mu^2}{Var}$ e $\beta = \frac{\mu}{Var}$. Os parâmetros podem também ser alterados através do arquivo parâmetros.ini, que é instalado juntamente com o programa.

Para cada observação realizada deve ser preenchida a contagem obtida para execução dos cálculos (Figura 16).

Figura 16 – Página inicial para inserção dos dados iniciais e dos dados observados.

A inserção de dados de contagem deve ser continuada até que o programa imprima o relatório final com a estimação realizada em que constam todos os itens do

calculo da media *a posteriori* do parâmetro desejado (Figura 17). Neste mesmo relatório é possível verificar o numero de observações e o somatório das contagens observadas.

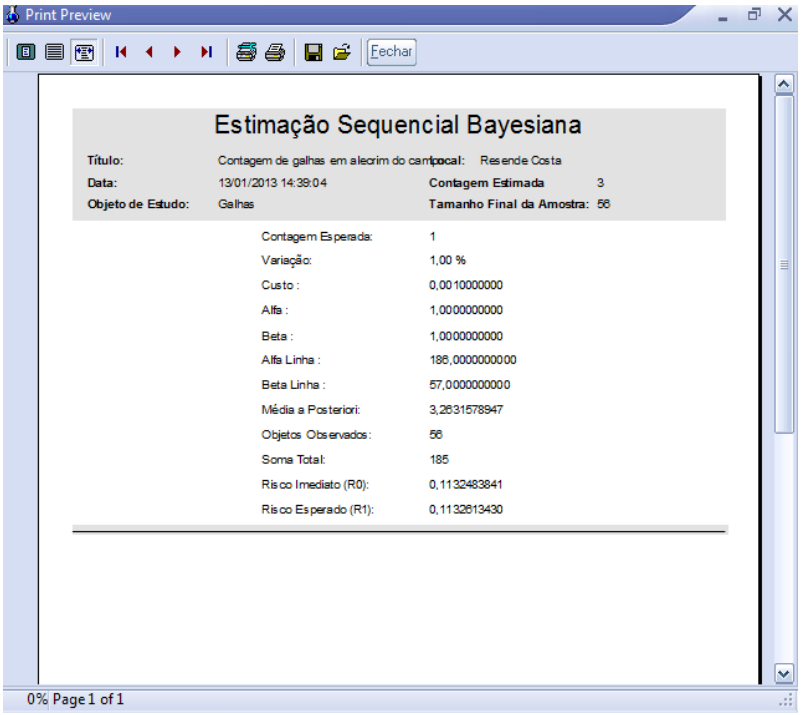


Figura 17 – Relatório com o resultado da Estimação Sequencial Bayesiana.

Após conclusão do processo os dados são armazenados no item "Dados Registrados", onde o usuário poderá ter acesso a todos as analises já feitas (Figura 18).

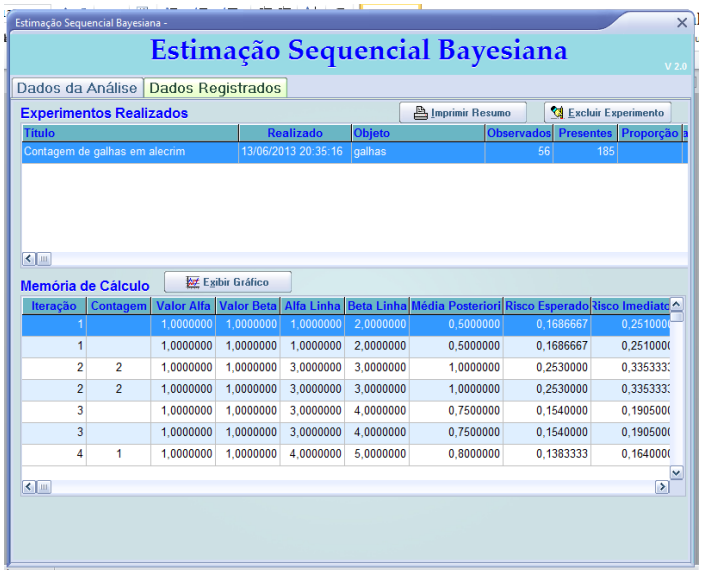


Figura 18 – Página de armazenamento de dados.

Está disponível ao usuário a opção de exibição do gráfico de evolução da analise (Figura 19).

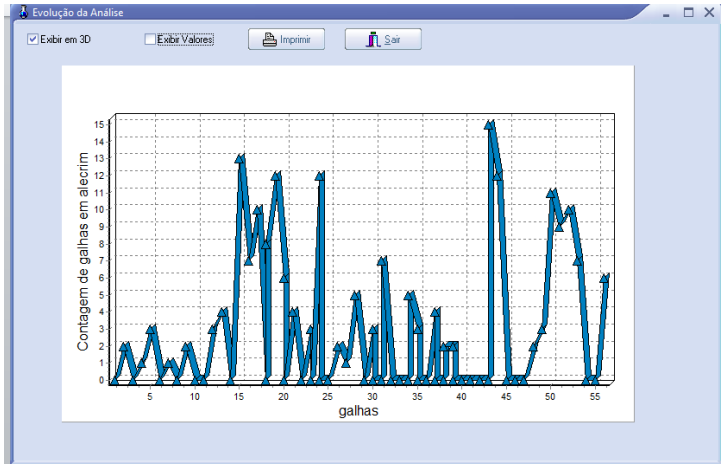
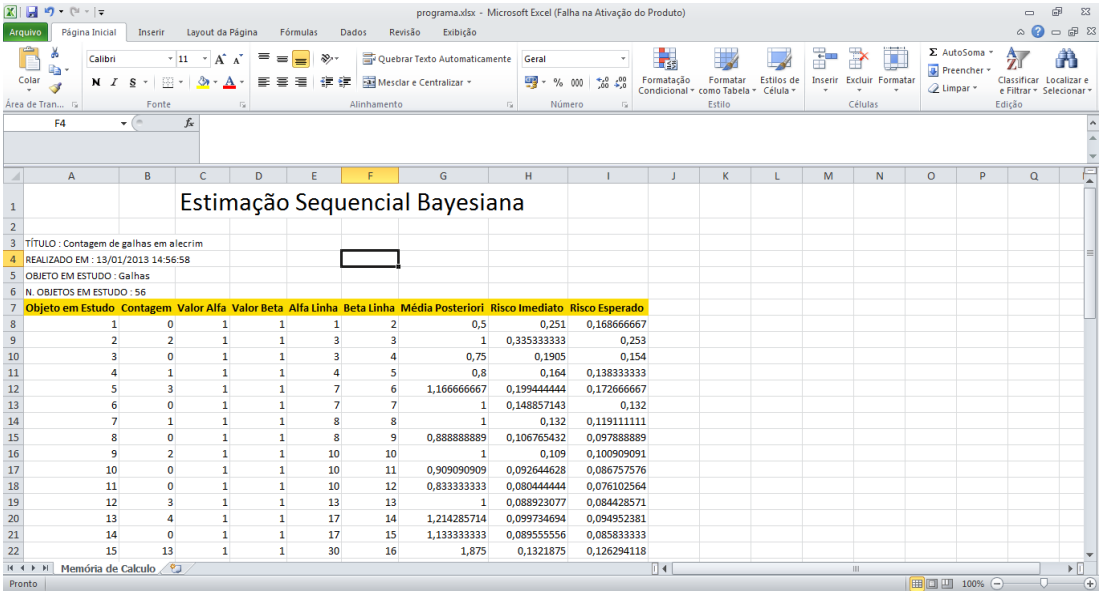


Figura 19 – Gráfico dos resultados da amostragem.

Na página principal do software é possível exportar os dados para uma planilha do Excel e assim disponibilizar uma ferramenta a mais para o usuário (Figura 20).



Estimação Sequencial Bayesiana

TÍTULO : Contagem de galhas em alecrim
REALIZADO EM : 13/01/2013 14:56:58
OBJETO EM ESTUDO : Galhas
N. OBJETOS EM ESTUDO : 56

Objeto em Estudo	Contagem	Valor Alfa	Valor Beta	Alfa Linha	Beta Linha	Média Posteriori	Risco Imediato	Risco Esperado
1	0	1	1	1	2	0,5	0,251	0,168666667
2	2	1	1	3	3	1	0,335333333	0,253
3	0	1	1	3	4	0,75	0,1905	0,154
4	1	1	1	4	5	0,8	0,164	0,138333333
5	3	1	1	7	6	1,166666667	0,199444444	0,172666667
6	0	1	1	7	7	1	0,148857143	0,132
7	1	1	1	8	8	1	0,132	0,119111111
8	0	1	1	8	9	0,888888889	0,106765432	0,097888889
9	2	1	1	10	10	1	0,109	0,100909091
10	0	1	1	10	11	0,909090909	0,092644628	0,086757576
11	0	1	1	10	12	0,833333333	0,080444444	0,076102564
12	3	1	1	13	13	1	0,088923077	0,084428571
13	4	1	1	17	14	1,214285714	0,099734694	0,094952381
14	0	1	1	17	15	1,133333333	0,089555556	0,085833333
15	13	1	1	30	16	1,875	0,1321875	0,126294118

Figura 20 – Memória de calculo exportada para o Excel.

O programa se mostrou adequado para realização da estimação sequencial Bayesiana do parâmetro da Poisson, notou-se apenas a necessidade de haver um valor mínimo para o tamanho amostral, para que a simulação fosse mais eficiente, obtendo-se assim melhor estimativa.

4. Conclusão

A simulação foi uma técnica eficiente para verificar a influencia dos planos de amostragem sequencial por meio de técnicas Bayesianas. Verificou-se que, quando, a media da *priori* considerou-se um valor mais próximo do valor paramétrico real, independente da variância considerada, as estimativas foram adequadas. Quando os valores de média da *priori* são distantes do valor paramétrico real, a variância da *priori* tem influência direta, principalmente quando o custo é alto. No caso de variância pequena, independente do custo por amostragem, o grau de informatividade da *priori* foi alto quando a média *a priori* foi distante do valor paramétrico real.

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.

Foi possível estimar o parâmetro da distribuição de Poisson (θ), a taxa média da infestação de galhas de psilídeo *Baccharopelma dracunculifoliae* no alecrim do campo, *Baccharis dracunculifolia*.

Na área de alecrim avaliada, utilizando-se a técnica de estimação sequencial Bayesiana, foi obtida uma taxa média de infestação de galhas de psilídeos *B. dracunculifoliae* igual a 3,26 galhas/planta durante o verão valores muito próximos também foram encontrados quando utilizada a estimação frequentista.

6. Referências Bibliográficas

ANDRADE, D. F.; OGLIARI, P. J. *Estatística para as Ciências Agrárias e Biológicas: com noções de experimentação.* Florianópolis: UFSC, 2007.

BASTOS, C. S., PICANÇO, M. C., & MARTINS, T. B. Sistemas de amostragem e tomada de decisão no manejo integrado de pragas do algodoeiro. *Rev. bras. ol. fibros., Campina Grande*, v.10, n.3, p.1119-1146, set./dez. 2006.

BEARZOTI, E.; AQUINO, L. H. Plano de amostragem seqüencial para avaliação de infestação de bicho-mineiro (*Lepidóptera:Lyonetiidae)* no sul de Minas Gerais. *Pesquisa Agropecuária Brasileira*, Brasília, v.29, n.5, p.695-705, maio 1994.

BERGER, J. O. *Statistical Decision Therory and Bayesian Analysis*. Second Edition. New York: Springer-Verlag, 1985.

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*. Piracicaba: ESALQ, 1995. 103p. (Tese - Doutorado em Entomologia).

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.

BRIGHENTI, C. R. G., RESENDE, M., & BRIGHENTI, D. M.. Estimação sequencial Bayesiana aplicada à proporção de infestação de psilídeos em alecrim do campo. *Rev. Bras. Biom*, v. 29, n. 2, p. 342-354, 2011.

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). *Journal of Natural History.* 38: 2051 - 2071. 2004.

BUSSAB,W.O.; MORETTIN,P.A. *Estatística Básica.* São Paulo: Editora Saraiva, 2002.

CHIARADIA, L. A., LÚCIO, A. D. C., & NETO, J. V. Seleção de modelos para amostragem seqüencial do ácaro-do-bronzeado Dichopelmus notus Keifer (Acari, Eriophyidae) em erva-mate. *Cerne*, v. 13, n. 3, p. 339-346, 2007.

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*. Piracicaba: ESALQ, 1977. 117p. (Tese - Mestrado em Experimentação e Estatística).

ESPÍRITO-SANTO, M. M.; FERNANDES, G. W. Parasitoid attack and its consequences to the development of the galling psyllid *Baccharopelma dracunculifoliae*. *Basic and Applied Ecology*. 5: 5, 475-484. 2004.

FAGUNDES, M.; NEVES, F. S.; FERNANDES, G. W. Direct and indirect interactions involving ants, insect herbivores, parasitoids, and the host plant *Baccharis dracunculifolia* (Asteraceae). *Ecological Entomology*. 30: 1, 28-35. 2005.

FARIAS, P. R. S.; BARBOSA, J. C.; BUSOLI A. C. Amostragem Seqüencial (Presença-Ausência) para S*podoptera frugiperda* (J.E. Smith) (Lepidoptera: Noctuidae) na Cultura do Milho. **Neotropical Entomology** 30(4): 691-695, 2001.

FERNANDES, M. G.; BUSOLI A. C.; BARBOSA, J. C. Amostragem seqüencial de S*podoptera frugiperda* (J. E. Smith, 1797) (lepidoptera, noctuidae) em algodoeiro. **Revista bras. Agrociência**, v. 8, n. 3, p. 213-218, 2002.

FERNANDES, M. G.; BUSOLI A. C.; BARBOSA, J. C. Amostragem Seqüencial de *Alabama argillacea* (Hübner) (Lepidoptera: Noctuidae) em algodoeiro. **Neotropical Entomology** 32(1):117-122, 2003.

FERNANDES, W.S. Método para a estimação de quantis de enchentes extremas com o emprego conjunto de análise Bayesiana, de informações não sistemáticas e de distribuições limitadas superiormente. *Programa de Pós-graduação em Saneamento, Meio Ambiente e Recursos Hídricos da UFMG*, 2009.

FONSECA, R.; COBUCI, J. A.; PIRES, A.V.; ARAÚJO, C. V.; COSTA, A. R. C. Influência da informação *a priori* na escolha de modelos para avaliação genética de suínos. *Ciênc. agrotec.*, Lavras, v.30, n. 3, p. 538-546, 2006.

GARTHWAITE, P. H.; JOLLIFFE, I; JONES, BYRON. *Statistical Inference.* London: Prentice Hall. p.114-131,1995.

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

IWAO, S. A new method of sequential samplig to classify populations relative to a critical density. *Research on Population Ecology*, Kyoto, v.16, p.281-288, 1975.

LINDLEY, D.V. The present position in Bayesian Statistics. *Statistical Science.* V. 5, n.1, p. 44-89. 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. Ciência e Tecnologia de Alimentos,* Campinas, 28(1): 178-181, jan.-mar. 2008.

MIGON, H. S. Introdução a Inferência Bayesiana**.** *IM and COPPE – UFRJ.* 2006

MOOD, A. M. ; GRAYBILL, F. A.; BOES, D. C. *Introduction to the Theory of Statistics.* Third Edition. Open University Pres, 1974.

NEGRÃO, I. O. ; AQUINO, L. H.; BEARZOTI, E. Avaliação de quatro métodos de estimação dos parâmetros da distribuição Beta-Binomial pela simulação Monte Carlo. *Ciênc. agrotec.*, Lavras, v.25, n.6, p.1370-1381, 2001.

NETO, J. V., DE AQUINO, L. H., BEARZOTI, E., & DE SOUSA, J. C. Otimização da amostragem seqüencial para o monitoramento do bicho-mineiro do cafeeiro Perileucoptera coffeella (Lepidoptera: Lyonetiidae) em lavras, Minas Gerais. *Ciênc. agrotec.*, Lavras, v.23, n.3, p.707-718, jul./set., 1999.

PAULINO, C.D.; TURKMAN, M.A.A.; MURTEIRA, B. Estatística Bayesiana. Fundação Calouste Gulbenkian. Lisboa. 2003.

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

PHAM-GIA, T. Distribution of the stopping time in bayesian sequential sampling. *Australian & New Zealand Journal of Statistics*, 40, p. 221-227, 1998.

PINHO, E. M. Estimação Bayesiana para Medidas de Desempenho de Testes Diagnósticos. *Universidade Federal de São Carlos*, 2006.

R DEVELOPMENT CORE TEAM. R: A language and environment for statistical computing. *R Foundation for Statistical Computing*. Vienna, Austria, 2011. ISBN 3-900051-07-0. Disponível em http://www.r-project.org

RAIFFA, H.; SCHLAIFER, R. *Applied Statistical Decision Theory*. Cambridge, MA: Harvard University Press, 1961.

ROSSI, R. M. *Introdução aos métodos Bayesianos na análise de dados zootécnicos com uso do WinBUGS e R*. p.26-31,2002.

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

ANEXO A

```
##### ROTINA PARA A ESTIMAÇÃO SEQUENCIAL BAYESIANA PARA DADOS DE CONTAGEM #####


#alteração de parâmetros iniciais

alfa=5              #parâmetro de escala da priori Gamma em que alfa>0
beta=0.5            #parâmetro de forma da priori Gamma  em que beta>0
theta=5             #parâmetro da  verossimilhança poisson
custo=0.000001      #custo por observação

#cálculos iniciais

       sn=0                        #soma de todos os valores
       tam=1                       #tamanho da amostra
       mpriori=alfa*(1/beta)       #média da priori
       varpriori=alfa*((1/beta)^2) #Variância da priori
       x = rpois(1,theta)
       sn <- sn + x
       mposteriori=((alfa+sn)*((1/(beta+tam))))         #média da posteriori
       varposteriori=((alfa+sn)*(((1/(beta+tam)))^2))   #Variância da posteriori
       risco=varposteriori + (tam*custo)                #Risco imediato
       riscoposteriori=((alfa+sn)+((alfa+sn)/(beta+tam)))/((beta+tam+1)^2)+((tam+1)*
custo)       #Risco esperado

 while((riscoposteriori <= risco) && (tam<10000))   # limite do tamanho amostral

       {
       x = rpois(1,theta)
       sn <- sn + x
       tam=tam+1
       mposteriori=((alfa+sn)*((1/(beta+tam))))
       varposteriori=((alfa+sn)*(((1/(beta+tam))^2)))
       risco=varposteriori + (tam*custo)
       riscoposteriori=(((alfa+sn)+((alfa+sn)/(beta+tam)))/((beta+tam+1)^2))+((tam+1
)*custo)
       }

#gráficos da verossimilhança, priori e posteriori

priori = rgamma(1000,alfa,(1/beta))
veross = rgamma(1000, (alfa+sn),tam)
posteriori = rgamma(1000,(alfa+sn),(1/(beta+tam)))
```

```
aposteriori = alfa+sn
bposteriori =((beta+tam))

plot(density(veross),xlim=c(0,20),ylim=c(0,3),main="",
xlab=expression(paste(theta)),ylab="Densidade",border=NULL,lty=3)
lines(density(priori),lty=2)
lines(density(posteriori),lty=1)
legend(12,2.5,c("Verossimilhança","Priori","Posteriori"),lty=c(3,2,1),
bty="n",cex=.9)

#texto complementar ao gráfico

text(12,1.7,"
alfa = 10
beta=10
custo=0.0001
theta=5")

#apresentação de resultados

cat("tamanho da amostra = ", tam, "\n")
cat("media a priori = ", mpriori, "\n")
cat("variancia a priori = ", varpriori, "\n")
cat("alfa a posteriori = ", aposteriori, "\n")
cat("beta a posteriori = ", bposteriori, "\n")
cat("media a posteriori = ", mposteriori, "\n")
cat("variancia a posteriori = ", varposteriori, "\n")
cat("risco imediato = ", risco, "\n")
cat("risco a posteriori = ", riscoposteriori, "\n")

#Intervalo de credibilidade

alfaposteriori=alfa+sn
betaposteriori=(beta+tam)
qgamma(0.025,alfaposteriori,betaposteriori)
qgamma(0.975,alfaposteriori,betaposteriori)
```

ANEXO B

```
#### MODELO DA ROTINA PARA PLOTAR CURVAS DAS DENSIDADES A PRIORI GAMA (PARAMETRO IGUAL A 1)####

x <- seq(0, 100, length=200)
plot(x,dgamma(x,5,10),type="l",xlim=c(-0.01,3),ylim=c(0,2.8),
main=expression(paste(theta," = 1")),
xlab=expression(paste(theta)),ylab="Densidade",border=NULL,col=1,lty=1)
curve(dgamma(x,0.025,0.05),type="l",add=T, col=2,lty=2)
curve(dgamma(x,10,10),type="l",add=T, col=3,lty=3)
curve(dgamma(x,0.5,0.5),type="l",add=T, col=4,lty=4)
curve(dgamma(x,20,10),type="l",add=T, col=5,lty=5)
curve(dgamma(x,1,0.5),type="l",add=T, col=6,lty=6)
simbolo <- expression(paste(theta," ~ Gamma(5,10)"), paste(theta," ~ Gamma(0.025,0.05)"), paste(theta," ~ Gamma(10,10)"), paste(theta," ~ Gamma(0.5,0.5)"),paste(theta," ~ Gamma(20,10)"),paste(theta," ~ Gamma(1,0.5)"))
legend(1.3,2.3, simbolo, lty=1:6,col=1:6,bty="n",cex=.9)
```