



ISSN 2175-6295 Rio de Janeiro- Brasil, 12 e 13 de agosto de 2010

SIMULAÇÃO MONTE CARLO E ALGORITMOS MCMC: ESTUDO COMPARATIVO E SIMULAÇÕES COM FOCO NA MÉDIA A POSTERIORI

Helinton A. L. Barbosa

UFMG – Departamento de Estatística – Av. Antônio Carlos, 6627 – Belo Horizonte/MG
[e-mail: helinton@ufmg.br](mailto:helinton@ufmg.br)

Fernando L. P. de Oliveira

UFMG – Departamento de Estatística – Av. Antônio Carlos, 6627 – Belo Horizonte/MG
[e-mail: fernandoluizoliveira@est.ufmg.br](mailto:fernandoluizoliveira@est.ufmg.br)

Thiago R. dos Santos

UFMG – Departamento de Estatística – Av. Antônio Carlos, 6627 – Belo Horizonte/MG
[e-mail: thiagorezende@yahoo.com.br](mailto:thiagorezende@yahoo.com.br)

Gabriel V. A. Fonseca

UFMG – Departamento de Estatística – Av. Antônio Carlos, 6627 – Belo Horizonte/MG
[e-mail: pyqui@yahoo.com.br](mailto:pyqui@yahoo.com.br)

João V. D. Monteiro

UFMG – Departamento de Estatística – Av. Antônio Carlos, 6627 – Belo Horizonte/MG
[e-mail: piocitos@gmail.com](mailto:piocitos@gmail.com)

Resumo

Neste trabalho apresentamos métodos de simulação Monte Carlo e algoritmos MCMC, definidos como rotinas de simulação Monte Carlo via Cadeias de Markov, muito usados em Estatística Bayesiana, área geralmente com uma necessidade computacional intensa. Também discutimos quais, entre os algoritmos apresentados, são os mais precisos, eficientes e viáveis sob o ponto de vista computacional. O foco principal do estudo é a média *a posteriori*, cujos resultados calculados analiticamente, são comparados com aproximações e simulações obtidas para vários parâmetros e tamanhos amostrais. De forma geral, a simulação Monte Carlo e os algoritmos MCMC são rápidos e eficientes, porém estes últimos possuem construção mais elaborada e requerem algumas particularidades. Ressalvas importantes sobre os métodos foram percebidas e são apresentadas no texto.

Palavras-Chave: Simulação Monte Carlo, Algoritmos MCMC, Estatística Bayesiana.

Abstract

We present methods of Monte Carlo simulation, and MCMC algorithms, or simulation routines Monte Carlo Markov Chain, used here in Bayesian statistics, an area usually computationally intensive. We also discuss which of the algorithms presented are the most accurate, efficient and viable from a computational point of view. The main focus of the study is the average *a posteriori*, the results calculated analytically, are compared with approximations and simulations obtained for various parameters and sample sizes. In general,

the Monte Carlo simulation and MCMC algorithms are fast and efficient, but the latter have a more elaborated and require a special case. Important caveats about the methods were perceived and are presented in the text.

Keywords: Simulation Monte Carlo, MCMC algorithms, Bayesian statistics.

1. INTRODUÇÃO

Em várias situações, alocadas em praticamente todas as áreas do conhecimento, ter definida, ou suposta como conhecida, a distribuição de probabilidade, ou delineamento probabilístico de um conjunto de dados, é de grande importância. Com base em tal conhecimento, calculam-se probabilidades de ocorrência de eventos, como por exemplo, o bloqueio de um servidor em uma rede de filas (Barbosa e Cruz, 2008), estimativas de confiabilidade, estatísticas de interesse e muitos outros assuntos relativos aos dados ou instrumentos modelados. Algumas situações exigem, ou se faz necessária, a atualização dos modelos definidos *a priori* e, esse novo ajuste, é feito pela necessidade de promover uma modelagem correta, ou mais precisa dos dados. Ou ainda, devido a modificações ocorridas durante o processo e que se fazem significativas em todos os cálculos do modelo. Nessas situações, e não raro, precisamos calcular uma integral, intratável analiticamente, ou necessitamos estimar uma quantidade amostral de uma nova distribuição, ainda não conhecida, ou mesmo estimar parâmetros desconhecidos. Para facilitar estas tarefas ou torná-las menos árduas, métodos numéricos, específicos e apropriados a cada situação, podem ser aplicados.

Neste trabalho, apresentaremos métodos de simulação Monte Carlo, bem como algoritmos de simulação Monte Carlo via Cadeias de Markov, descrevendo suas aplicações em Estatística Bayesiana, área que quase sempre necessita de cálculos intensos em suas análises. Discutiremos quais destes métodos são mais coerentes, precisos e viáveis no cálculo da média *a posteriori*. Usaremos nos algoritmos dois tamanhos de amostra, além de diferentes valores de probabilidade de sucesso θ . Os resultados obtidos pelos métodos de simulação serão comparados com os valores analíticos exatos, calculados da distribuição *a posteriori*. Todos os programas foram escritos e implementados no freeware R (Development Core Team, 2005).

Este artigo encontra-se organizado da seguinte forma: Na próxima seção descrevemos rapidamente alguns conceitos introdutórios de Estatística Bayesiana, que são usados no trabalho e, em seguida, na seção III, apresentaremos os métodos e as estruturas das simulações utilizadas no estudo. A seção IV, mostra os resultados computacionais obtidos através da implementação dos algoritmos descritos na seção III. Finalmente, na seção V, apresentamos as principais conclusões, além de alguns comentários sobre o estudo.

2. FUNDAMENTOS DE ESTATÍSTICA BAYESIANA

Os componentes fundamentais da Estatística Bayesiana são: informação *a priori*, resumida através da distribuição *a priori*; a informação trazida pela amostra de dados, que é resumida utilizando a função de verossimilhança; a distribuição *a posteriori*, que é uma atualização da distribuição *a priori* pelos dados, e, em alguns casos, o cálculo da distribuição de futuras observações. De forma objetiva, o problema pode ser colocado da seguinte maneira:

Seja $\theta = (\theta_1, \theta_2, \dots, \theta_k)$ uma quantidade, desconhecida e aleatória, com possíveis valores em um conjunto Θ e admita que se deseje inferir sobre Θ . A informação inicial H , que varia de pessoa a pessoa, sobre θ é resumida probabilisticamente em termos de uma função de densidade:

$$\pi(\theta|H), \theta \in \Theta \quad (1)$$

Que é uma descrição a respeito da incerteza pessoal em relação a θ , com base em toda informação disponível sobre o parâmetro. Esta distribuição (1) é denominada distribuição *a priori* de θ . Com o propósito de atualizar esta informação sobre θ , uma amostra de n valores de um vetor aleatório $X = (X_1, X_2, \dots, X_n)$ relacionado a θ , será observado. Antes de observarmos a amostra, descreveremos a incerteza sobre x , dado θ , através da função:

$$p(x | H, \theta), \quad x \in X, \theta \in \Theta \quad (2)$$

Note que a dependência de θ é fundamental. A informação em $X = (X_1, X_2, \dots, X_n)$ é resumida pela função de verossimilhança (Migon e Gamerman, 1999) com respeito a θ , função esta que associa a cada θ o valor $p(x | H, \theta)$. Para um valor fixado de θ , a função de verossimilhança fornece a probabilidade de ser observado x quando θ é o verdadeiro estado de natureza.

Depois de definidas e atribuídas as funções de probabilidade (1) e (2), é de interesse encontrar uma maneira de atualizar a opinião inicial, com respeito a θ , utilizando a informação contida na amostra. A técnica utilizada para realizar tal operação é o teorema de Bayes, isto é, a distribuição *a posteriori* de θ é dada por:

$$p(\theta | x) = \frac{p(x | \theta)}{p(x)} = \frac{p(x | \theta)\pi(\theta)}{p(x)}, \theta \in \Theta \quad (3)$$

Se utilizarmos a notação $\pi(\theta|x)$ em lugar de $p(\theta|x)$ para enfatizar o fato de que $\pi(\theta|x)$ é uma atualização de $\pi(\theta)$, de (3) temos:

$$\pi(\theta | x) = \frac{p(x | \theta)}{p(x)} = \frac{p(x | \theta)\pi(\theta)}{p(x)}, \theta \in \Theta \quad (4)$$

Note que:

$$p(x) = \int p(x | \theta)\pi(\theta)d\theta \quad (5)$$

Então (5) corresponde a densidade marginal de x , que é denominada distribuição preditiva e é o denominador de (4) (Zuazola, et al., 1996).

2.1. Estimação Pontual

Seja $\alpha(\theta, \delta) = (\theta - \delta)$ a perda esperada, onde θ é o valor real e δ são possíveis estimativas de θ . O estimador de Bayes para θ é o valor δ que torna mínima a perda esperada (6), *a posteriori*, isto é:

$$\hat{\theta}_B = \underset{\delta}{\text{mim}} \quad E(L(\theta, \delta) | x) = \underset{\delta}{\text{mim}} \quad \int L(\theta, \delta)\pi(\theta | x)d\theta \quad (6)$$

Considerando $L(\theta, \delta) = (\theta - \delta)^2$, temos:

$$\begin{aligned} E((\theta - \delta)^2 | x) &= E(\theta^2 - 2\theta\delta + \delta^2 | x) \\ &= E(\theta^2 | x) - 2\delta E(\theta | x) + E(\delta^2 | x) \\ &= E(\theta^2 | x) - 2\delta E(\theta | x) + \delta^2 \end{aligned} \quad (7)$$

Assim, para encontrarmos o estimador de Bayes, δ , com $L(\theta, \delta) = (\theta - \delta)^2$, temos que minimizar (7) com relação a δ . Para isso, devemos fazer $\frac{d(E((\theta - \delta)^2 | x))}{d\delta} = 0$ como mostrado (8).

$$\frac{d(E((\theta - \delta)^2 | x))}{d\delta} = 0 - 2E(\theta | x) + 2\delta = 0 \quad (8)$$

A confirmação de que o estimador escolhido para θ é realmente um ponto de mínimo é realizada através de (9), ou seja:

$$\frac{d^2(E((\theta - \delta)^2 | x))}{d\delta^2} > 0 \quad (9)$$

Verificando (9) que o estimador escolhido para θ é realmente um ponto de mínimo, então $\hat{\theta}_B = E(\theta | x)$ será o estimador de Bayes. Ou seja, se a perda quadrática for utilizada, isto é,

$L(\theta, \delta) = (\theta - \delta)^2$, a média da distribuição *a posteriori* de θ será o estimador de Bayes. Ressaltamos que diferentes funções de perda geram distintos estimadores de Bayes e, frisamos ainda, que a escolha da função de perda é completamente subjetiva. Quando a função de perda escolhida for a 0-1 (leia-se zero um), a moda *a posteriori* é o estimador de Bayes. Porém, ser for trabalhada a perda absoluta, o estimador de Bayes será a mediana *a posteriori* (Migon e Gamerman, 1999).

2.2. Análise Conjugada

Seja $\mathfrak{S} = \{f(x|\theta): \theta \in \Theta\}$ uma família de distribuições amostrais (funções verossimilhança onde Θ é o espaço paramétrico associado à θ). Seja $H = \{h(\theta|a): a \in A\}$ uma família de distribuição de probabilidade, onde A é um conjunto de hiperparâmetros. Dizemos que \mathfrak{S} e H são famílias conjugadas naturais se:

1. $f(x|\theta)$ é proporcional a um membro de H ;
2. H é fechado em relação a produtos, isto é: $h(\theta|X) = \frac{f(x|\theta)h(\theta)}{f(x)}$

2.2.1. Distribuição de Bernoulli

Sejam $X = (X_1, X_2, \dots, X_n)$ variáveis aleatórias que, dado θ , são condicionalmente independentes e identicamente distribuídas (*c.i.i.d.*) com distribuição *Bernoulli*(θ). Então a função de verossimilhança será dada por (10):

$$f(X|\theta) = \prod_{i=1}^n \theta^{x_i} (1-\theta)^{1-x_i} = \theta^{\sum_{i=1}^n x_i} (1-\theta)^{n-\sum_{i=1}^n x_i} \quad (10)$$

A candidata à conjugada natural da família *Bernoulli* é a família de distribuição *Beta*(α, β), com $\alpha, \beta > 0$. Ou seja, de (4), (5) e (10) temos:

$$\begin{aligned} \pi(\theta|X) &= \frac{\theta^{\sum_{i=1}^n x_i} (1-\theta)^{n-\sum_{i=1}^n x_i} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1} (1-\theta)^{\beta-1}}{\int_0^1 \theta^{\sum_{i=1}^n x_i} (1-\theta)^{n-\sum_{i=1}^n x_i} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1} (1-\theta)^{\beta-1} d\theta} \\ &= \frac{\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1+\sum_{i=1}^n x_i} (1-\theta)^{\beta-1+n-\sum_{i=1}^n x_i}}{\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 \theta^{\alpha-1+\sum_{i=1}^n x_i} (1-\theta)^{\beta-1+n-\sum_{i=1}^n x_i} d\theta} = \frac{\theta^{\alpha-1+\sum_{i=1}^n x_i} (1-\theta)^{\beta-1+n-\sum_{i=1}^n x_i}}{\int_0^1 \theta^{\alpha-1+\sum_{i=1}^n x_i} (1-\theta)^{\beta-1+n-\sum_{i=1}^n x_i} d\theta} \\ \pi(\theta|X) &= \frac{\theta^{\alpha-1+\sum_{i=1}^n x_i} (1-\theta)^{\beta-1+n-\sum_{i=1}^n x_i}}{\frac{\Gamma(\alpha+\sum_{i=1}^n x_i)\Gamma(n-\sum_{i=1}^n x_i+\beta)}{\Gamma(\alpha+\beta+n)}} \end{aligned}$$

3. MÉTODOS DE SIMULAÇÃO

3.1. Métodos de Simulação Monte Carlo

Para ilustrar os métodos de simulação Monte Carlo (Gamerman, 1997), seja $h(\theta) = \frac{f(\theta)}{\int f(\theta)d\theta}$,

uma função da qual queremos obter amostras de $h(\theta)$ sem resolver a integral. Suponha que uma amostra é facilmente gerada de uma função $g(\theta)$, chamada de função de referência e de quem desejamos obter uma amostra de $h(\theta)$ - a função $h(\theta)$ deve ser positiva e padronizável. Se tivermos a situação anterior, podemos gerar uma amostra de $h(\theta)$, sabendo apenas a forma funcional de $f(\theta)$ e tendo uma amostra de $g(\theta)$. As técnicas possíveis para tal geração amostral são os métodos da Rejeição e SIR (*Sampling Importance Resampling* ou *Bootstrap* Bayesiano).

3.1.1. Método da Rejeição

Suponha que exista m tal que:

$$\frac{f(\theta)}{g(\theta)} \leq m, \forall \theta \quad (12)$$

O objetivo é encontrar $g(\theta)$, da qual sabemos gerar amostras e, tal que, $g(\theta)m$ seja $f(\theta)$. Para isso, considere o algoritmo a seguir:

1. Gere $\theta_i \sim g(\theta)$ para $i = 1, 2, \dots, T$;
2. Gere $u \sim \text{uniforme}(0,1)$;
3. Se $u \leq \frac{f(\theta)}{mg(\theta)}$, se aceita θ_i , onde m é tal que $\frac{f(\theta)}{g(\theta)} \leq m$;
4. Volte ao passo 1.

O resultado obtido é uma de amostra de $h(\theta), \theta_1, \dots, \theta_k$, $k \leq T$. A distribuição $g(\theta)$ deve ser fácil de ser obtida e conter caudas pesadas, além da amostra gerada ser grande o suficiente para varrer bem todo o espaço paramétrico.

3.1.2. Método SIR (*Sampling Importance Resampling*)

Se m (12) no procedimento anterior não está disponível, uma possibilidade é usar o SIR. Veja o algoritmo a seguir:

1. Escolha uma função de referência;
2. Gere uma amostra $\theta_1, \theta_2, \dots, \theta_n$ de $g(\theta)$;
3. Para cada $i = 1, 2, \dots, n$ calcule:

$$w_i = \frac{f(\theta_i)}{g(\theta_i)} \text{ e } q_i = \frac{w_i}{\sum_{i=1}^n w_i}$$

4. Selecione uma amostra $\theta_1^*, \theta_2^*, \dots, \theta_r^*$ de $g(\theta)$ da amostra original $\theta_1, \theta_2, \dots, \theta_n$ de $g(\theta)$, assumindo $p(\theta = \theta_i) = q_i$;
5. Gere $u \sim \text{uniforme}(0,1)$ e observe se:
 - $u \in (0, q_1)$, escolha θ_1 .
 - $u \in (q_1, q_1 + q_2)$, escolha θ_2 .
 - $u \in (q_1 + q_2, q_1 + q_2 + q_3)$, escolha θ_3 .

As mesmas limitações observadas pelo método da Rejeição são válidas para o SIR.

3.2. Métodos de Simulação Monte Carlo via Cadeias de Markov (MCMC)

3.2.1. Introdução

Idéia central do método é construir uma Cadeia de Markov, da qual seja fácil gerar uma amostra e que haja uma distribuição de equilíbrio, $h(\theta)$, dada pela distribuição de interesse. Para tal, as seguintes condições devem ser observadas:

1. Seja $\theta_1, \theta_2, \dots, \theta_p \sim p(\theta)$ com $p(\theta_1, \theta_2, \dots, \theta_p)$ definida em $\Theta \subset \mathcal{R}^p$;
2. Devemos supor uma Cadeia de Markov homogênea, irreduzível e aperiódica, com espaço de estado Θ , e cuja distribuição de equilíbrio, $p(\theta)$, possa ser construída. Ou seja, deve ser possível construir uma Cadeia de Markov com probabilidade de transição invariante no tempo, onde cada estado possa ser visitado de qualquer outro, com um número finito de interações, e não pode haver estado absorvente. A distribuição estacionária deve ser $p(\theta)$;
3. As amostras das probabilidades de transição devem ser geradas facilmente.

A seguir, apresentamos um dos métodos mais comumente utilizados.

3.2.2. Algoritmo Metropolis-Hastings

Para descrever o algoritmo, suponha que a distribuição de interesse é a distribuição *a posteriori* ($\theta | x$) com $\theta = (\theta_1, \theta_2, \dots, \theta_s)$. Considere também, que todas as condicionais completas *a posteriori* com $(\theta | \theta_{-i}, x)$, $i = 1, \dots, n$, estejam disponíveis, mas que não se saiba gerar amostras diretamente de cada uma. Suponha ainda, que as amostras de um novo valor de θ_i , são geradas a partir de uma distribuição proposta condicional ao valor atual de θ_i , $q(\theta_i^{(p)} | \theta_i^{(a)})$, onde $\theta_i^{(p)}$ é o valor proposto e $\theta_i^{(a)}$ é o valor atual, para $i = 1, \dots, n$. Para facilitar o entendimento, um esquema de amostragem é apresentado a seguir:

1. Inicialize $\theta^{(0)} = (\theta_1^{(0)}, \dots, \theta_s^{(0)})$ e $k = 1$
2. Obtenha um novo valor para $\theta^{(k)}$ a partir de $\theta^{(k-1)}$ através de sucessivas gerações de valores. Para $i = 1$ até S , faça:
 - (i) Gere uma proposta para $\theta_i^{(k)}$ de $\theta_i^{(p)} \sim q(\theta_i | \theta_i^{(k-1)})$
 - (ii) Não rejeite a proposta de probabilidade de aceitação dada por:
$$\alpha = \min \left\{ 1, \frac{p(\theta_i^{(p)} | \theta_i^{(a)}, x) q(\theta_i^{(k-1)} | \theta_i^{(p)})}{p(\theta_i^{(k-1)} | \theta_i^{(a)}, x) q(\theta_i^{(p)} | \theta_i^{(k-1)})} \right\} \quad \text{onde, } \theta_{-i}^{(a)} = (\theta_1^{(k)}, \dots, \theta_{i-1}^{(k)}, \theta_{i+1}^{(k-1)}, \dots, \theta_s^{(k-1)})$$
3. Faça $k = k + 1$, volte para o passo 2 e repita até alcançar a convergência.

O Algoritmo de Metropolis-Hastings é bastante geral e pode, pelo menos em princípio, ser implementado com qualquer distribuição condicional completa *a posteriori* e para qualquer proposta. Entretanto, sob o ponto de vista prático, a escolha da proposta é crucial para o bom desenvolvimento do algoritmo, ou seja, para sua convergência na direção da distribuição *a posteriori*. Algumas propostas mais comuns são apresentadas a seguir.

A- Cadeias Simétricas

Esta situação ocorre quando a distribuição proposta é simétrica em torno da iteração anterior, isto é:

$$q(\theta_i^{(p)} | \theta_i^{(k-1)}) = q(\theta_i^{(k-1)} | \theta_i^{(p)})$$

$$\alpha = \left\{ 1, \frac{p(\theta_i^{(p)} | \theta_i^{(a)}, x)}{p(\theta_i^{(k-1)} | \theta_i^{(a)})} \right\}$$

Dentre as cadeias simétricas, destaca-se o passeio aleatório, $\theta_i^{(p)} = \theta_i^{(k-1)} + \varepsilon$, onde ε tem uma distribuição simétrica em torno zero.

B- Cadeias Independentes

As cadeias são independentes se a proposta não depende do passo anterior, ou seja:

$$q(\theta_i^{(p)} | \theta_i^{(k-1)}) = q(\theta_i^{(p)})$$

e a probabilidade de aceitação é dada por:

$$\alpha = \left\{ 1, \frac{p(\theta_i^{(p)} | \theta_i^{(a)}, x) q(\theta_i^{(k-1)})}{p(\theta_i^{(k-1)} | \theta_i^{(a)}) q(\theta_i^{(p)})} \right\}$$

Um caso particular de cadeias independentes é dado quando a distribuição proposta é a distribuição *a priori* para θ_i , neste caso a probabilidade de aceitação é dada somente pela função de verossimilhança, isto é:

$$\alpha = \left\{ 1, \frac{x | p(\theta_i^{(p)}, \theta_i^{(a)})}{p(x | p(\theta_i^{(k-1)}, \theta_i^{(a)}))} \right\}$$

Outro caso particular de cadeias independentes se dá quando a distribuição proposta é a própria condicional completa *a posteriori*, isto é, $q(\theta_i^{(p)}) = q(\theta_i^{(p)} | \theta_i^{(a)}, x)$. Fazendo isto, a probabilidade de aceitação é igual a 1. Essa característica de gerar da condicional completa e

aceitar sempre em um algoritmo iterativo é a definição do Gibbs Sampler, portanto, um caso particular do algoritmo de Metropolis-Hastings.

3.3.4. Verificação de Convergência

Os métodos de MCMC são ótimas ferramentas para resolução de muitos problemas práticos na análise Bayesiana, bem como em muitas outras áreas do conhecimento. Porém, algumas questões relacionadas à convergência nestes métodos ainda merecem bastante pesquisa. Um ponto que pode surgir é “quantas iterações deve ter o processo de simulação para garantir que a cadeia convergiu para o estado de equilíbrio?” A resposta definitiva para esta questão talvez não possa ser dada, visto que a distribuição estacionária será na prática desconhecida, mas podemos sempre avaliar a convergência das cadeias detectando problemas fora do período de aquecimento. Uma análise de convergência em métodos de simulação pode ser feita preliminarmente analisando os gráficos ou medidas descritivas dos valores simulados da quantidade de interesse θ . As representações gráficas mais frequentes são o gráfico de θ ao longo das iterações e um gráfico da estimativa da distribuição *a posteriori* de θ , além de histograma ou uma densidade Kernel. As estatísticas usuais são a média, o desvio padrão e os quantis (2,5%; 50%; 97,5%). Na detecção do período de *burn-in*, usam-se gráficos como a média ergódica (Rodrigues et al., 2009) e funções de auto-correlação. No gráfico da média ergódica, quando não houver variabilidade significativa, houve convergência.

4. RESULTADOS EXPERIMENTAIS

Nas seções anteriores, nos preocupamos em apresentar ao leitor alguma fonte introdutória sobre o assunto Estatística Bayesiana e Métodos de Simulação Monte Carlo e via Cadeias de Markov. A partir de agora, com o objetivo de avaliar todos os métodos e apresentar conclusões sobre quais são os melhores para cada situação, mostraremos alguns resultados encontrados através da aplicação direta dos algoritmos apresentados anteriormente. Para tal, geramos os dados de uma distribuição Binomial, com parâmetros n e θ ($X \sim bin(n; \theta)$). Assumimos uma distribuição de probabilidade *a priori* pouco informativa para θ , ou seja, uma distribuição *Beta* (1; 1), cuja distribuição *a posteriori*, por conjugação, é uma *Beta*, com parâmetros $\alpha + \sum x_i$ e $n - \sum x_i + \beta$ e média igual $\frac{\alpha + \sum x_i}{n + \alpha + \beta}$. Foram utilizados dois tamanhos

amostrais, ou seja, $n=10$ e $n=100$. No algoritmo Metropolis-Hastings, para a obtenção da distribuição empírica de θ , isto é, uma amostra da distribuição *a posteriori* de θ , consideramos 10.000, 50.000 e 100.000 iterações. Para o método da Rejeição utilizamos 1.000, 10.000 e 100.000 iterações e, finalmente, 1.000 e 2.000 iterações para o método SIR. Foram tomados ainda valores diferentes de probabilidades de sucesso ou $\theta = 0.01, 0.50$ e 0.99 .

4.1. Resultados dos Métodos de Simulação Monte Carlo

4.1.1. Resultados do Método da Rejeição

Nesta seção, utilizamos o método da Rejeição para obter uma distribuição empírica *a posteriori* de θ . Os resultados são apresentados nas Figuras 1, 2 e 3. Na Figura 1, é gerada uma amostra aleatória de uma distribuição Binomial com os parâmetros $n=10$, $n=100$ e $\theta=0.01$. O número de iterações neste caso é 1.000, 10.000 e 100.000. Note que quando aumentamos o tamanho da amostra de 10 para 100, a amostra da distribuição *a posteriori* de θ se aproxima mais do valor real de $\theta=0.01$. Observe também que a distribuição empírica de θ (histograma) é bem similar à distribuição *a posteriori* exata (tracejado). Para valores do número de iterações acima ou iguais a 10.000, os gráficos são bem parecidos. Na Figura 2, com $n=10$, $n=100$ e $\theta=0.50$, como observado na Figura 1, quando aumentamos o tamanho da amostra n , a distribuição empírica *a posteriori* gerada pelo método da Rejeição se concentra

mais em torno do valor de θ assumido. A distribuição empírica *a posteriori* de θ é bem similar à distribuição exata em todos os casos mostrados na Figura 2. Na Figura 3, com $\theta=0.99$, a distribuição *a posteriori* é bem assimétrica, concentrada em torno do valor de θ , assumido principalmente quando utilizamos um tamanho de amostra $n=100$. A distribuição empírica é bem parecida com a distribuição exata, como foi visto também nas figuras anteriores, atestando que o Método da Rejeição é plausível.

A Tabela I apresenta estatísticas descritivas da distribuição empírica *a posteriori* obtidas através do Método da Rejeição. Para $n=10$, a iteração que melhor estimou a média *a posteriori* foi a com 10.000, para qualquer θ , cuja variância foi a menor em relação às outras iterações. Considerando $n=100$, os valores das médias e variâncias ficaram muito próximos um dos outros, mas devemos lembrar que quanto maior a quantidade de iterações maiores são os tempos de simulação. Destaca-se a mediana, que fornece resultados similares aos resultados da média. Vale ressaltar também, que a variância diminui ao passo que o tamanho da amostra aumenta, isto é, a distribuição empírica *a posteriori* de θ se torna mais precisa. Observe ainda que, para $n=10$, a variância está na terceira casa decimal geralmente, enquanto que para $n=100$, a variância está, na maioria dos casos, na quarta casa decimal. Novamente, esta informação, combinada com a diminuição da variabilidade pelo aumento no tamanho amostral, indica uma coerência do algoritmo executado. Em geral, a porcentagem de rejeição do método é grande, sendo necessário um número alto de iterações para obter um número de amostras satisfatório.

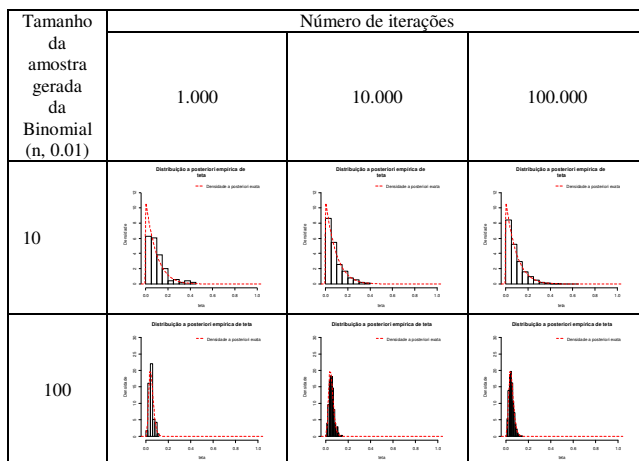


Fig. 1. Método da Rejeição para binomial ($n,0.01$).

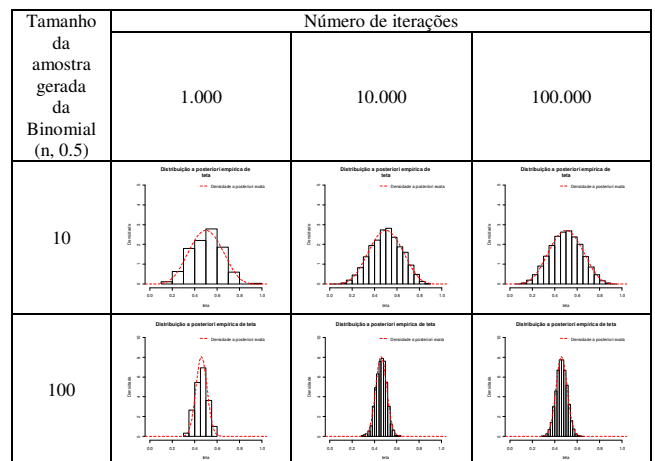


Fig. 2. Método da Rejeição para binomial ($n,0.50$).

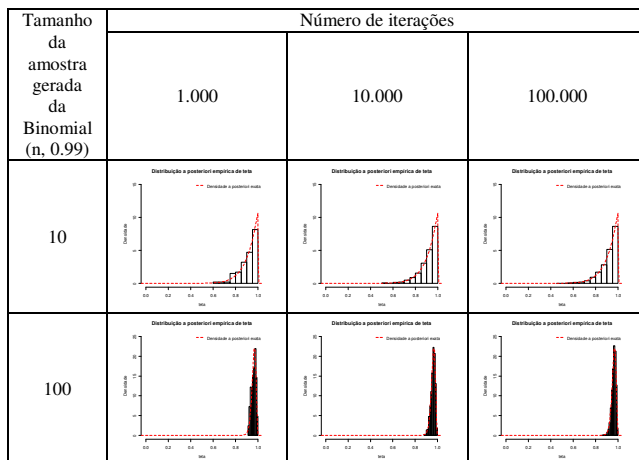


Fig. 3. Método da Rejeição para binomial ($n,0.99$).

TABELA I – ESTATÍSTICAS DO MÉTODO DA REJEIÇÃO

n	Iterações	θ	Estatísticas Descritivas			
			Média <i>a posteriori</i> Exata	Mediana	Média	Variância
10	1.000	0.01	0.08	0.0679	0.0949	0.0070
		0.5	0.5	0.5079	0.5009	0.0189
		0.99	0.92	0.9266	0.9133	0.0059
	10.000	0.01	0.08	0.0623	0.0831	0.0056
		0.5	0.5	0.5043	0.5029	0.0190
		0.99	0.92	0.9404	0.9175	0.0058
	100.000	0.01	0.08	0.0626	0.0845	0.0059
		0.5	0.5	0.5001	0.4993	0.0193
		0.99	0.92	0.9399	0.9399	0.0058
100	1.000	0.01	0.05	0.0471	0.0490	0.0004
		0.5	0.46	0.4664	0.4609	0.0027
		0.99	0.96	0.9670	0.9614	0.0004
	10.000	0.01	0.05	0.0503	0.0524	0.0005
		0.5	0.46	0.4631	0.4613	0.0023
		0.99	0.96	0.9636	0.9604	0.0004
	100.000	0.01	0.05	0.0454	0.0485	0.0004
		0.5	0.46	0.4611	0.4614	0.0024
		0.99	0.96	0.9640	0.9610	0.0004

4.1.2. Resultados do Método SIR

A seguir, apresentamos os resultados do método SIR para obtenção da distribuição empírica a posteriori de θ . Como o método SIR gasta mais tempo, foi determinado um número de iterações de 1.000 e 2.000, que não resultou em perda da qualidade dos resultados das simulações. As Figuras 4, 5 e 6 mostram que o SIR se aproxima bem da distribuição a posteriori para qualquer n e θ , mas com algumas diferenças, sendo o Método da Rejeição mais satisfatório.

Na Tabela II, temos que, com 2.000 iterações, as variâncias são menores. A média a posteriori, tanto para 1.000 quanto para 2.000 iterações, possui valores bem próximos, sendo melhor no segundo caso. Os tempos de simulações são bem maiores em 2.000 iterações (aproximadamente 72,7 segundos) do que com 1.000 iterações (em torno 8,8 segundos), indicado a existência de coerência do algoritmo. Novamente, vale ressaltar o menor número de iterações consideradas no algoritmo do SIR em relação aos outros métodos. Para que o leitor tenha uma idéia de como é demorado o tempo de simulação do SIR em um número maior de iterações, fizemos um experimento, no qual foi considerado $n=100$ e $\theta=0.01$ e 10.000 iterações. O tempo computacional foi de 2 horas e 50 minutos aproximadamente (10.203,8 segundos). Comparando este resultado com o tempo para 2.000 iterações (1 minuto e 13 segundos), percebemos claramente a limitação deste método em relação à quantidade de iterações.

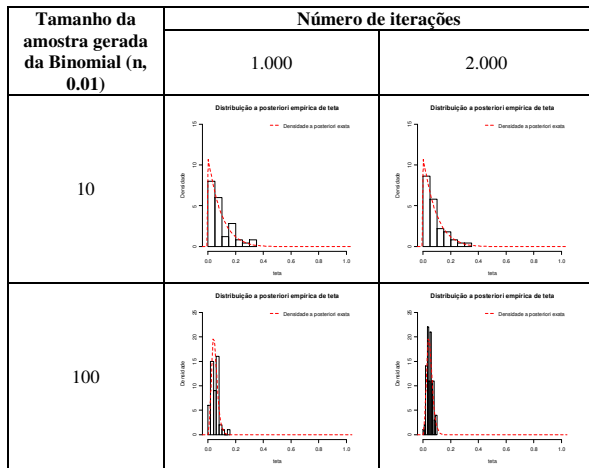


Fig. 4. Método SIR para binomial $(n,0.01)$.

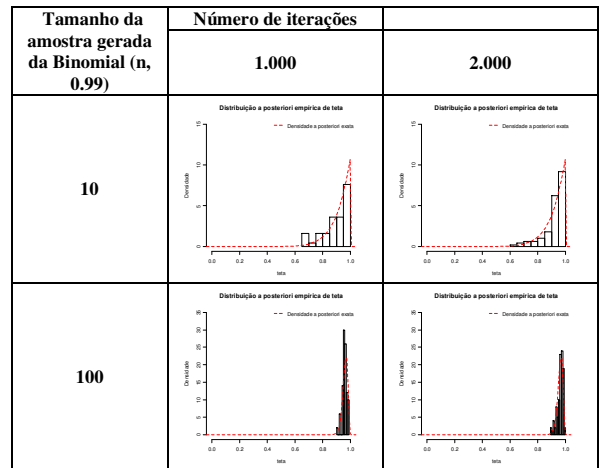


Fig. 6. Método SIR para binomial $(n,0.99)$.

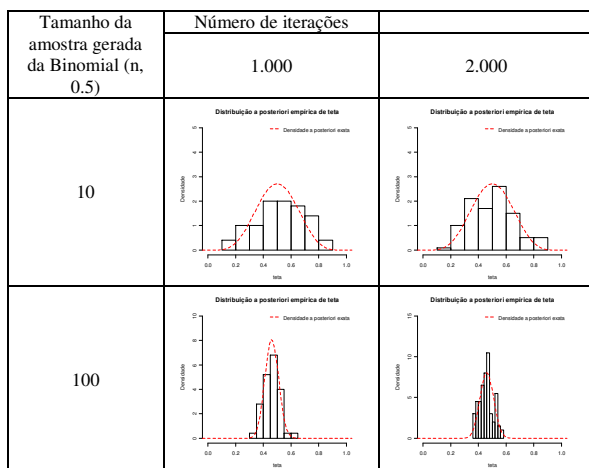


Fig. 5. Método SIR para binomial $(n,0.50)$.

TABELA II – ESTATÍSTICAS DO MÉTODO SIR

n	Número de iterações	θ	Estatísticas Descritivas			
			Média a posteriori Exata	Mediana	Média	Variância
10	1.000	0.01	0.08	0.0544	0.0876	0.0065
		0.5	0.5	0.5522	0.5281	0.0294
	2.000	0.99	0.92	0.9134	0.8892	0.0086
		0.01	0.08	0.0643	0.0785	0.0058
100	1.000	0.5	0.5	0.5022	0.5008	0.0249
		0.99	0.92	0.9463	0.9216	0.0061
	2.000	0.01	0.05	0.0494	0.0487	0.0007
		0.5	0.46	0.4602	0.4575	0.0032
2.000	0.99	0.96	0.9589	0.9595	0.0003	
	0.01	0.05	0.0494	0.0499	0.0004	
2.000	0.5	0.46	0.4579	0.4578	0.0023	
	0.99	0.96	0.9682	0.9628	0.0005	

4.2. Resultados dos Métodos MCMC

4.2.1. Algoritmo Metropolis-Hastings

Na Figura 7, percebemos que quando aumentamos o tamanho da amostra de 10 para 100, a amostra da distribuição *a posteriori* de θ se aproxima mais do valor real de $\theta=0.01$, isto é, a variância diminui. Observamos, também, que a distribuição empírica de θ é bem similar à distribuição *a posteriori* exata. Não há alteração significativa nos gráficos, variando o número de gerações de 10.000 para 50.000 e 100.000. Como observado na Figura 7, na Figura 8, a distribuição empírica *a posteriori* de θ é bem próxima da distribuição exata e há diminuição na variância da distribuição considerando $n=100$. Também não há alterações significativas nos gráficos, variando o número de iterações. Assumindo $\theta=0.5$, a distribuição dos dados é simétrica. Finalmente na Figura 9, a distribuição dos dados é bem assimétrica considerando $\theta=0.99$, conforme se esperava.

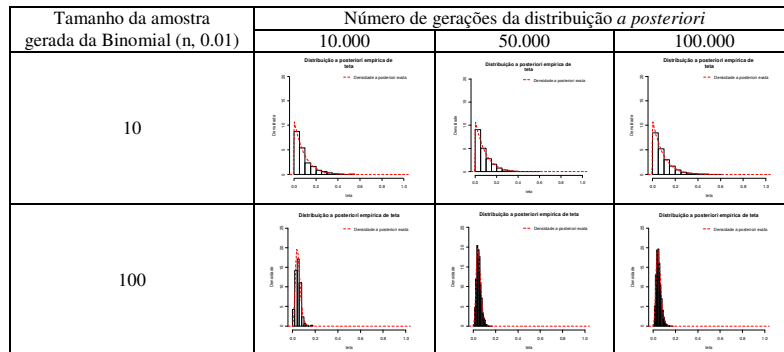


Fig. 7. Método da Metropolis para binomial ($n,0.01$).

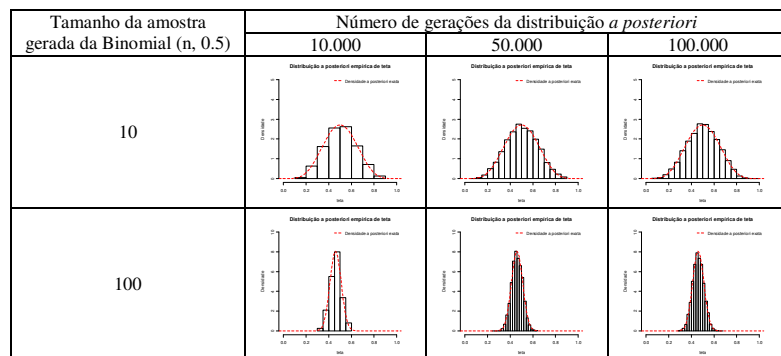


Fig. 8. Método da Metropolis para binomial ($n,0.5$).

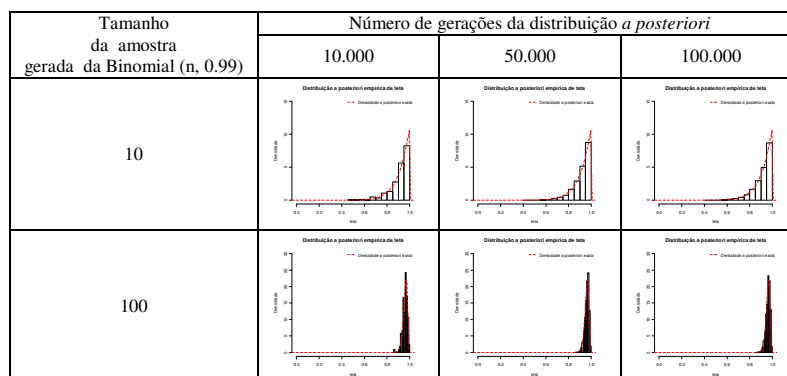


Fig. 9. Método da Metropolis para binomial ($n,0.99$).

Ainda na Figura 9, notamos que a distribuição empírica *a posteriori* de θ é bem similar à distribuição exata e que a variabilidade dos dados diminui aumentando o valor de n . Também não existe alteração na distribuição dos dados, quando variamos o número de iterações. Isso também foi percebido nas Figuras 7 e 8.

A Tabela III mostra que a variância da amostra da distribuição *a posteriori* de θ está na segunda casa decimal com $n=10$, ao passo que, com $n=100$, está na quarta casa decimal, indicando grande coerência do algoritmo. Na Tabela III, notamos ainda que a média da amostra *a posteriori* de θ é bem próxima à média exata e que a mediana é similar à média *a*

posteriori de θ , principalmente nos casos em que $\theta=0.5$, já que a distribuição empírica dos dados com este valor é simétrica, conforme mostra a Figura 8. Avaliando o tamanho da amostra *a posteriori* gerada (TAM) e o número de iterações, percebemos que a perda de amostras está em torno de 90.5% a 95%, dependendo do número de iterações assumido.

Na análise de convergência do algoritmo, pela Figura 10, verificamos que os gráficos da média ergódica, para $\theta=0.01$, começam a se estabilizar depois de 1.000 iterações. Considerando $n=10$ e 100, a média ergódica se estabiliza em torno de 0.08 e de 0.05, respectivamente. Vale salientar que 0.08 e 0.05 são os valores da média *a posteriori* exata, como indicado na Tabela III. A Figura 11 indica que os gráficos da média ergódica começam a se estabilizar após 1.000 iterações e com 3.000 iterações, já há pouca variação. Quando $n=10$ e 100, a média se estabiliza em torno de 0.50 e 0.46, respectivamente. Finalmente, a Figura 12 indica que os gráficos da média, para $\theta=0.99$, começam a se estabilizar após 1.000 iterações.

TABELA III – ESTATÍSTICAS DO MÉTODO METROPOLIS-HASTINGS

n	Iterações	θ	Média <i>a posteriori</i> exata	Estatísticas Descritivas							
				Mínimo	Q1	Mediana	Q3	Máximo	Média	Variância TAM	
10	10.000	0.01	0.08	0.0010	0.0230	0.0593	0.1184	0.5046	0.0840	0.0065	501
		0.5	0.5	0.1697	0.4072	0.5057	0.6007	0.8931	0.5066	0.0190	501
		0.99	0.92	0.4703	0.8825	0.9362	0.9718	0.9996	0.9127	0.0067	501
	50.000	0.01	0.08	0.00002	0.02369	0.05705	0.1138	0.5698	0.07980	0.0056	4501
		0.5	0.5	0.0891	0.4034	0.5006	0.6010	0.8998	0.5019	0.0195	4501
		0.99	0.92	0.4282	0.8835	0.9416	0.9754	1.0000	0.9184	0.0058	4501
	100.000	0.01	0.08	0.00004	0.02577	0.06249	0.1191	0.5.736	0.08367	0.0058	9501
		0.5	0.5	0.0622	0.4031	0.5016	0.5990	0.9502	0.5003	0.0190	9501
		0.99	0.92	0.4249	0.8811	0.9395	0.9732	1.0000	0.9160	0.0060	9501
100	10.000	0.01	0.05	0.0076	0.0315	0.0463	0.0642	0.1710	0.0485	0.0005	501
		0.5	0.46	0.3051	0.4269	0.4617	0.4920	0.5994	0.4601	0.0025	501
		0.99	0.96	0.8695	0.9445	0.9609	0.9737	0.9937	0.9579	0.0004	501
	50.000	0.01	0.05	0.0066	0.0343	0.0466	0.0605	0.1622	0.0488	0.0004	4501
		0.5	0.46	0.2527	0.4274	0.4591	0.4946	0.6261	0.4598	0.0023	4501
		0.99	0.96	0.8538	0.9515	0.9652	0.9748	0.9964	0.9617	0.0004	4501
	100.000	0.01	0.05	0.0043	0.0333	0.0459	0.0608	0.1760	0.0489	0.0005	9501
		0.5	0.46	0.2857	0.4271	0.4595	0.4943	0.6625	0.4604	0.0025	9501
		0.99	0.96	0.8615	0.9499	0.9648	0.9747	0.9978	0.9611	0.0004	9501

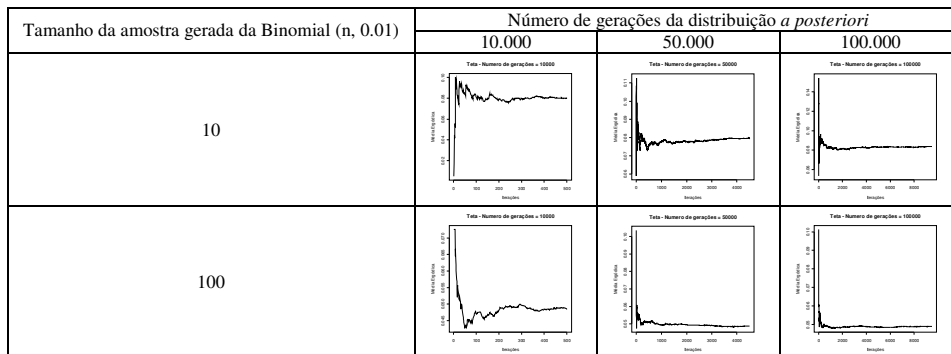


Fig. 10. Média ergódica Metropolis para binomial ($n,0.01$).

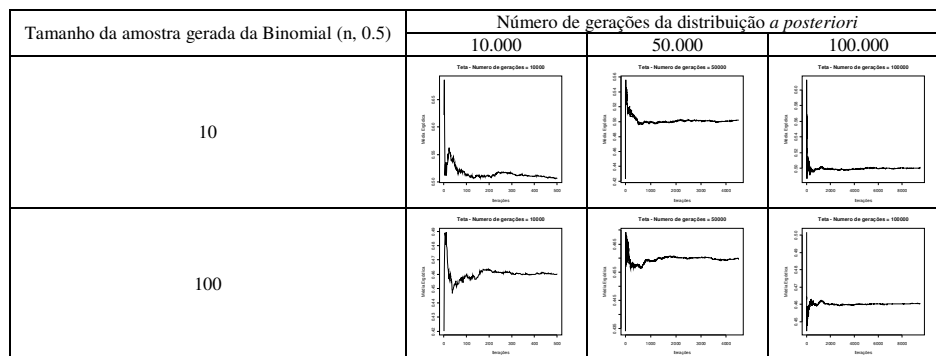


Fig. 11. Média ergódica Metropolis para binomial ($n,0.5$).

Tamanho da amostra gerada da Binomial (n, 0.99)	Número de gerações da distribuição <i>a posteriori</i>		
	10.000	50.000	100.000

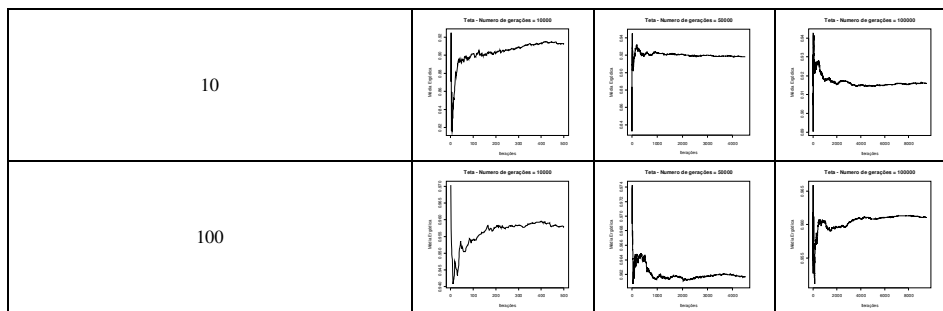


Fig. 12. Média ergódica Metropolis para binomial $(n,0.99)$.

Há pouca variação em todos os gráficos com um número de iterações superior a 3.000. Isso foi observado também para as Figuras 10 e 11. A média ergódica se estabiliza em torno de 0.92 e 0.96 quando $n=10$ e 100, respectivamente. Ressaltamos que estes valores 0.92 e 0.96 são os valores da média *a posteriori* exata, valores estes que podem ser vistos na Tabela III.

5. CONCLUSÕES E OBSERVAÇÕES FINAIS

Neste artigo, descrevemos de maneira resumida, Estatística Bayesiana, métodos de simulação Monte Carlo e algoritmos MCMC, além de compararmos os diferentes métodos em relação à obtenção da média *a posteriori*. Em geral, as implementações de todos os programas foram satisfatórias, não havendo um algoritmo que se destaque muito em relação aos outros. O método da Rejeição, o SIR e o Metropolis-Hastings produziram bons resultados, porém destacamos que o SIR é o método mais lento, motivo pelo qual, optamos por um número de gerações bem menor para ele. Ressaltamos ainda, que o SIR é de grande valia quando m (12) não está disponível, impossibilitando o uso do método da Rejeição. O algoritmo Metropolis-Hastings traz um pouco mais de dificuldade na sua implementação, por se tratar de um método mais elaborado. Porém, o algoritmo é rápido, mesmo com um número grande de iterações (100.000 iterações realizadas em 60 segundos aproximadamente), além de bastante coerente.

Muitas vezes o tempo computacional é um fator decisivo para o pesquisador, entre outros fatores a se considerar, por exemplo, erros de truncamento e aproximação. Considerando o quesito tempo, claramente destacamos novamente o método SIR como o mais lento. Então, cabe ao pesquisador a decisão de qual programa utilizar, pois o método mais apropriado pode variar, dependendo dos fatores e limitações operacionais disponíveis.

A ilustração de alguns métodos, aqui abordados, deve ser apenas uma fonte introdutória de pesquisa, cuja importância cresce à medida que, também aumenta a relevância do trabalho a ser desenvolvido. Leitores interessados em outros algoritmos podem consultar Barbosa et al. (2010), onde são descritos e comparados (entre si e com resultados analíticos) alguns métodos de integração numérica, em relação ao cálculo da média *a posteriori*.

6. REFERÊNCIAS BIBLIOGRÁFICAS

- BARBOSA, H. A. L.; CRUZ, F. R. B. (2008), 'Análise de um Algoritmo de Otimização de Redes de Filas Finitas'. In: XI SPOLM. Rio de Janeiro. Anais [CD-ROM]. Rio de Janeiro : EMA-CASNAV, 2008, p. 1-15.
- BARBOSA, H. A. L.; OLIVEIRA, F. L. P.; SANTOS, T.R.; FONSECA, G.V.A; MONTEIRO, J.V.D. 'Cálculo de Média a Posteriori através de Métodos de Integração Numérica e Simulação Monte Carlo: Estudo Comparativo'. *Revista INGEPRO - Inovação, Gestão e Produção*, v. 2, p. 60-74, 2010.
- BERNADO, J.; BERGER J.; SMITH, A. eds, '*Bayesian Statistics 4*', University Press, Oxford, 763-774.
- COWLES, M. K.; CARLIN, B. 'Markov chain monte carlo convergences diagnostics: a comparative review', *Journal of the American Statistical Association* 91, 883-904, 1996.
- GAMERMAN, D. *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*, Chapman and Hall, London, 1997.
- HASTINGS, W. K. 'Monte carlo sampling methods using markov chains and their applications', *Biometrika* 51(7), 97-109, 1970.

METROPOLIS, N.; ROSENBLUTH, A. W.; ROSENBLUTH, M. N.; TELLER, A. H.; TELLER, E. 'Equations of state calculations by fast computing machines', *Journal of Chemical Physics* 21(6), 1087–1092, 1953.

MIGON, H. S.; GAMERMAN, D. *Statistical Inference: an integrated approach*, Arnold, Rio de Janeiro, 1999.

R DEVELOPMENT CORE TEAM (2005), R: a language and environment for statistical computing, *R Foundation for Statistical Computing*, Vienna, Austria.

RODRIGUES, A.; S., WATANABE, G.K.; CORDER, V.A. 'Convergência de Cadeias'. Departamento de Estatística. Centro de Ciências Exatas e de Tecnologia. Universidade Federal de São Carlos, 2009. http://www.ufscar.br/~polpo/disc/2009_02/grad_estat_bayes/grupo6.pdf

ZUAZOLA, I. P. L.; SILVA, A. J.; VALLE, R. B. A. 'Bayesian inference in errors in variables models under elliptical distribution', 12 SINAPE Anais, 1996. 378–378.