

Cálculo de média a posteriori através de métodos de integração numérica e simulação monte carlo: estudo comparativo

Helinton Andre Lopes Barbosa (UFMG) helinton@ufmg.br
Fernando Luiz Pereira Oliveira (UFMG) fernandoest@ig.com.br
Thiago Rezende Santos (UFMG) thiagorezende@yahoo.com.br
Gabriel Vinicius Araújo Fonseca (UFMG) pyqui@yahoo.com.br
João Vitor Dias Monteiro (UFMG) piocitos@gmail.com

Resumo: Neste trabalho são apresentados métodos de aproximação numérica e simulação Monte Carlo, usados em Estatística Bayesiana, além de se discutir quais, entre os algoritmos apresentados, são os mais eficientes, precisos e rápidos computacionalmente. O foco principal do estudo é a média a posteriori, cujos resultados calculados analiticamente, são comparados com aproximações e simulações obtidas para vários parâmetros. De forma geral, os métodos de quadraturas são rápidos e eficientes, assim como a simulação Monte Carlo. Algumas ressalvas sobre estes algoritmos foram percebidas e são apresentadas no estudo.

Palavras-chaves: Integração numérica; Simulação monte carlo; Estatística bayesiana.

1. INTRODUÇÃO

Em diversas situações, desenvolvidas em praticamente todas as áreas do conhecimento, ter definida, ou suposta como conhecida, a distribuição de probabilidade de um conjunto de dados é de grande importância. A partir de tal conhecimento, calculam-se probabilidades de ocorrência de eventos, confiabilidade, estatísticas de interesse e vários assuntos relativos aos dados ou instrumentos modelados.

Existem casos em que é necessária a atualização dos modelos definidos *a priori* e, esse novo ajuste, se dá pela necessidade de promover uma modelagem correta dos dados, ou ainda devido a modificações ocorridas durante o processo e que se fazem significativas em todos os cálculos do modelo. Nessas situações, muitas vezes 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, métodos numéricos, específicos e apropriados a cada situação, podem ser aplicados.

Neste trabalho, apresentaremos algumas aproximações numéricas e descreveremos suas aplicações em Estatística Bayesiana, área que quase sempre necessita de cálculos intensos em suas análises. Discutiremos quais destes métodos são mais eficazes, precisos e rápidos no cálculo da média *a posteriori*. Usaremos nas aproximações vários tamanhos de amostra, além de diferentes valores de probabilidade de sucesso θ . Os resultados obtidos pelos algoritmos de aproximação 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) e encontram-se disponíveis no Anexo ao final deste texto.

Este artigo encontra-se organizado da seguinte forma: No próximo capítulo descrevemos rapidamente alguns conceitos elementares de Estatística Bayesiana e, em seguida, no capítulo 3, apresentaremos os métodos numéricos e as simulações utilizadas no estudo. O capítulo 4, mostra os resultados computacionais obtidos através da implementação

dos algoritmos descritos no capítulo 3. Finalmente, no capítulo 5, apresentamos as principais conclusões do estudo, encerrando, no Anexo, com os códigos dos programas.

2. Estatística Bayesiana

2.1 Introdução

Os componentes básicos 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 maneira simples, o problema pode ser colocado da seguinte forma:

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:

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

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 & 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 atribuídas as funções de probabilidade em (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). Para uma explicação mais detalhada ver Zuazola et al. (1996).

2.2 Estimação Pontual

A distribuição *a posteriori* é toda a informação disponível sobre θ após observarmos os dados. Apesar disto, muitas vezes precisamos resumi-la em um único número. Este número é o estimador de Bayes.

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 torna mínima a perda esperada (6), a posteriori, isto é:

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

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 , e para isso, devemos fazer $\frac{d(E((\theta - \delta)^2 | x))}{d\delta} = 0$ como mostrado em (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:

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

Confirmando (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, quando 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.

Devemos ressaltar 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 & GAMERMAN, 1999).

2.3 Análise Conjugada

Seja $\mathfrak{T} = \{f(x|\theta) : \theta \in \Theta\}$ 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 hiper-parâmetros. Dizemos que \mathfrak{T} 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:

$$h(\theta | X) = \frac{f(x|\theta)h(\theta)}{f(x)}$$

Exemplo

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 $Poisson(\theta)$. Então a função de verossimilhança será:

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

A candidata à família conjugada natural da família Poisson é a família de distribuição $Gama(\alpha, \beta)$, com $\alpha, \beta > 0$, ou seja, por (4), (5) e (10) temos:

$$\begin{aligned} \pi(\theta | X) &= \frac{\frac{e^{-n\theta} \theta^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!} \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\beta\theta}}{\int_0^\infty \frac{e^{-n\theta} \theta^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!} \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\beta\theta} d\theta} = \frac{e^{-(n+\beta)\theta} \theta^{\alpha + \sum_{i=1}^n x_i - 1}}{\int_0^\infty e^{-(n+\beta)\theta} \theta^{\alpha + \sum_{i=1}^n x_i - 1} d\theta} = \frac{e^{-(n+\beta)\theta} \theta^{\alpha + \sum_{i=1}^n x_i - 1}}{\frac{\Gamma(\alpha + \sum_{i=1}^n x_i)}{(n + \beta)^{\alpha + \sum_{i=1}^n x_i}}} \\ \Rightarrow (\theta | X) &\sim Gama(\alpha + \sum_{i=1}^n x_i; n + \beta) \end{aligned}$$

3. Métodos Numéricos

3.1 Quadraturas

Suponha que desejamos calcular uma integral definida (Anton, 2000), $I = \int_a^b f(\theta) d\theta$, no intervalo (a, b) . Basicamente, os métodos de quadratura usuais consistem em:

1. Escolher “pontos”, $\theta_1, \theta_2, \dots, \theta_n$ no intervalo (a, b) .
2. Determinar pesos w_{is} , tais que, $I = \sum_{i=1}^n w_i f(\theta_i)$.

Desta forma, consegue-se calcular o valor aproximado da integral em estudo. Vejamos, a seguir, alguns métodos de quadraturas.

3.1.1 Método Newton-Cotes

O método de Newton-Cotes (Filho, 2000) é também denominado como aproximação pelo ponto médio, já que em cada subintervalo de amplitude h , o valor da função é tomado no ponto médio do subintervalo. Aproximações pelo extremo esquerdo e direito dos subintervalos não são muito utilizadas.

Um esquema desta abordagem pode ser apresentado como:

1. Divida o intervalo (a, b) em n partes iguais;
2. Avalie a função $f(\theta)$ no ponto médio de cada intervalo;
3. A integral I , é aproximada por (11):

$$\hat{I}_{NC} = h \sum_{i=1}^n f\left(a + \frac{(2i-1)h}{2}\right), \text{ para } h = \frac{b-a}{n} \quad (11)$$

3.1.2 Regra Trapezoidal

As aproximações pelos extremos esquerdos e direito são raramente usadas em aplicações, porém, se é tomada a média entre elas, obtemos um resultado chamado de Aproximação Trapezoidal ou Regra do Trapézio (Filho, 2000), a qual é usada comumente. O nome aproximação trapezoidal pode ser explicado considerando-se o resultado no qual $f(x)$ é maior ou igual a zero em $[a,b]$, de tal forma que, $\int_a^b f(x)$, representa a área abaixo de $f(x)$ e acima de $[a,b]$. Geometricamente, a aproximação trapezoidal resulta em se aproximar esta área pela soma das áreas dos trapézios. Neste caso, a integral I é aproximada por (12):

$$\hat{I}_T = h \left[\frac{f(a)}{2} + \sum_{i=1}^n f\left(a + \frac{(2i-1)}{2}h\right) + \frac{f(b)}{2} \right], \text{ para } h = \frac{b-a}{n} \quad (12)$$

3.1.3 Regra de Simpson

Para melhorar as aproximações pelos pontos médio e trapezoidal, substituem-se os contornos superiores lineares por contornos superiores curvos, escolhidos para se ajustarem mais estritamente à forma da curva $y = f(x)$. Esta é a idéia subjacente da Regra de Simpson (Filho, 2000), a qual usa curvas parabólicas da forma $y = ax^2 + bx + c$.

A Regra de Simpson é obtida dividindo-se o intervalo $[a,b]$ por um número par de subintervalos, de igual comprimento h , aproximando a área abaixo de $y = f(x)$ em pares sucessivos de subintervalos. A soma destas aproximações serve, então, como uma estimativa de $\int_a^b f(x)$. A aproximação da integral I por este método é dada por (13):

$$\hat{I}_S = \frac{h}{3} \left[f(a) + 4 \sum_{i=1}^{n/2} f\left(a + \frac{4i-1}{2}h\right) + 2 \sum_{i=1}^{n/2} f\left(a + \frac{4i+3}{2}h\right) + f(b) \right] \quad (13)$$

para $h = \frac{b-a}{n}$.

4. 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).

4.1 Método da Rejeição

Suponha que exista m tal que,

$$\frac{f(\theta)}{g(\theta)} \leq m, \quad \forall \theta \quad (14)$$

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.

A amostra obtida 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.

4.2 Método SIR (Sampling Importance Resampling)

Se m (14) no procedimento anterior não está disponível, uma possibilidade é usar o SIR. Veja o algoritmo a seguir:

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

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

4. Selecionar uma amostra $\theta_1^*, \theta_2^*, \dots, \theta_n^*$ 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.

5. Resultados Experimentais

Com o objetivo de avaliar todos os métodos e apresentar conclusões sobre quais são os melhores para cada situação, mostraremos, a seguir, alguns resultados encontrados através da aplicação dos algoritmos apresentados anteriormente. Geramos os dados de uma distribuição Binomial, com parâmetros n e θ ($X \sim \text{bin}(n; \theta)$). Assumimos uma distribuição *a priori* para θ sendo uma distribuição *Beta* (1; 1), cuja distribuição *a posteriori*, por conjugação, é uma *Beta*, com parâmetros $\alpha + \sum x_i$ e $n - \sum x_i + \beta$, com média igual $\frac{\alpha + \sum x_i}{n + \alpha + \beta}$.

Foram utilizados vários tamanhos amostrais, ou seja, $n=10$ e $n=100$. Quanto ao número de iterações, temos 10, 100 e 1.000 para os métodos de quadraturas; 1.000, 10.000 e 100.000 para o método da Rejeição e 1.000 e 2.000 para o SIR, bem como valores diferentes de probabilidades de sucesso, isto é, $\theta=0.01$; 0.50 e 0.99.

5.1 Resultados de Quadraturas

Nesta seção, para valores de $\theta=0.01$, $\theta=0.5$, $\theta=0.99$, $n=10$, $n=100$ e sendo o número de iterações iguais a 10, 100 e 1000, apresentamos os resultados obtidos nas Tabelas 1, 2 e 3. Para um maior poder de comparação, foram calculados valores exatos das médias *a posteriori* para diversos θ e n .

Nas Tabelas 1, 2 e 3, notamos que a partir da milésima iteração, a aproximação da média *a posteriori* converge para o valor exato. Porém, observamos ainda que o tempo gasto é maior quando temos mais iterações e que, com somente dez iterações, não observamos um bom valor para a média, independente do valor de θ ou de n . Salientamos também que o aumento de n só acarretará na perda da informação da distribuição *a priori*, pois quando n é muito grande, a média de X (amostra) é mais informativa que a média dos tetras (*a priori*). Os tempos de simulação são bem pequenos, sendo pouco importantes se comparados uns com os outros.

TABELA 1 – Resultados Newton-Cotes

n	θ	Iterações	Média <i>a posteriori</i>	Média Exata	Tempo(seg.)
10	0.01	10	0.08756	0.08333	0.00
		100	0.08337		0.00
		1000	0.08333		0.60
	0.50	10	0.50001	0.50000	0.00
		100	0.50000		0.00
		1000	0.50000		0.05
	0.99	10	0.86755	0.91667	0.00
		100	0.91616		0.00
		1000	0.91666		0.04
100	0.01	10	0.09046	0.049019	0.00
		100	0.04902		0.02
		1000	0.04902		0.05
	0.50	10	0.46505	0.46078	0.00
		100	0.46078		0.00
		1000	0.46078		0.04
	0.99	10	1.33989	0.96078	0.00
		100	0.95987		0.00
		1000	0.96078		0.04

TABELA 2– Resultados Trapezoidal

n	θ	Iterações	Média <i>a posteriori</i>	Média Exata	Tempo(seg.)
10	0.01	10	0.08756	0.08333	0.00
		100	0.08338		0.00
		1000	0.08333		0.05
	0.50	10	0.50001	0.50000	0.00
		100	0.50000		0.00
		1000	0.50000		0.05
	0.99	10	1.41755	0.91667	0.00
		100	0.97116		0.02
		1000	0.92216		0.03
100	0.01	10	0.09046	0.04902	0.00
		100	0.04902		0.00
		1000	0.04902		0.05
	0.50	10	0.46505	0.46078	0.00
		100	0.46078		0.00
		1000	0.46078		0.05
	0.99	10	1.33989	0.96078	0.00
		100	0.95987		0.00
		1000	0.96078		0.05

TABELA 3 – Resultados Simpson

n	θ	Iterações	Média <i>a posteriori</i>	Média Exata	Tempo(seg.)
10	0.01	10	0.08333	0.08333	0.00
		100	0.08174		0.02
		1000	0.08332		0.05
	0.50	10	0.49883	0.50000	0.00
		100	0.50000		0.01
		1000	0.50000		0.04
	0.99	10	1.07428	0.91667	0.00
		100	0.93452		0.00

		1000	0.91849		0.05
100	0.01	10	0.00523	0.04902	0.00
		100	0.04856		0.01
		1000	0.04901		0.04
0.50	10	10	0.54732	0.46078	0.00
		100	0.46078		0.00
		1000	0.46078		0.04
0.99	10	10	0.89371	0.96078	0.00
		100	0.96509		0.00
		1000	0.96078		0.05

5.2. Resultados dos Métodos de Simulação Monte Carlo

5.2.1 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 $\theta=0.01$. 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. Na Tabela 4, são apresentadas as estatísticas descritivas da distribuição empírica *a posteriori* de obtida 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 quarta casa decimal. Em geral, a porcentagem de rejeição do método de Rejeição é grande, sendo necessário um número alto de iterações para obter um número de amostras satisfatório.

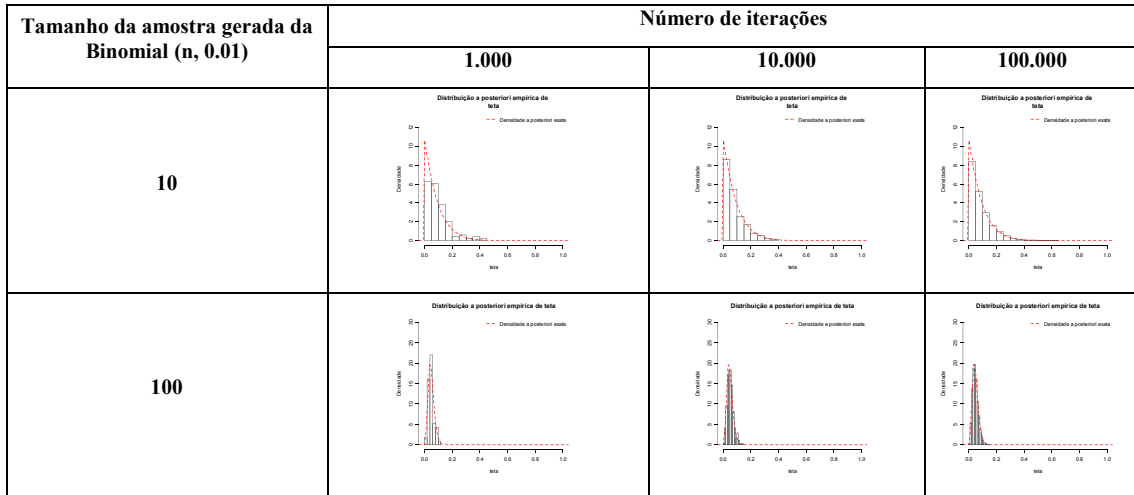


FIGURA 1 - Método da Rejeição para binomial (n,0.01).

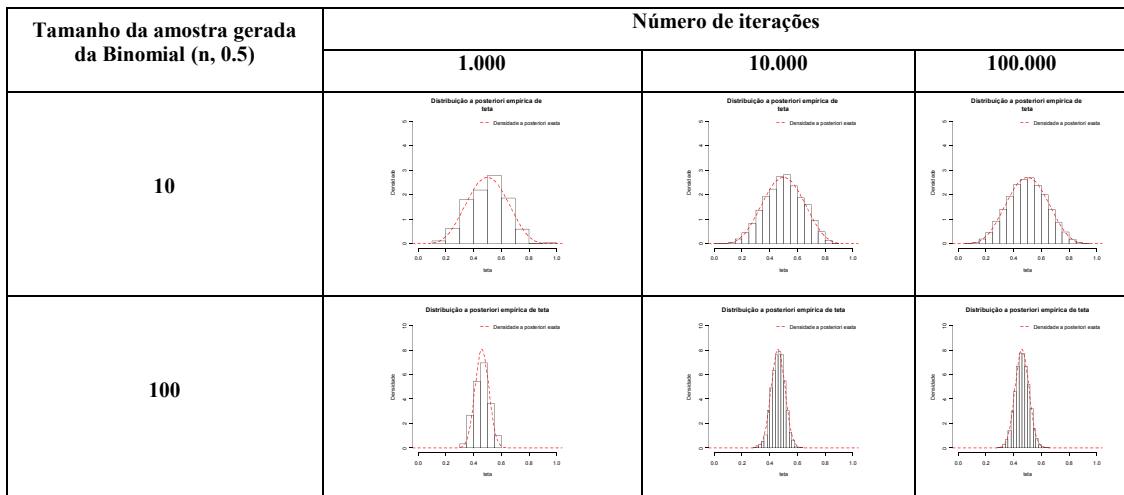


FIGURA 2 - Método da Rejeição para binomial (n,0.50).

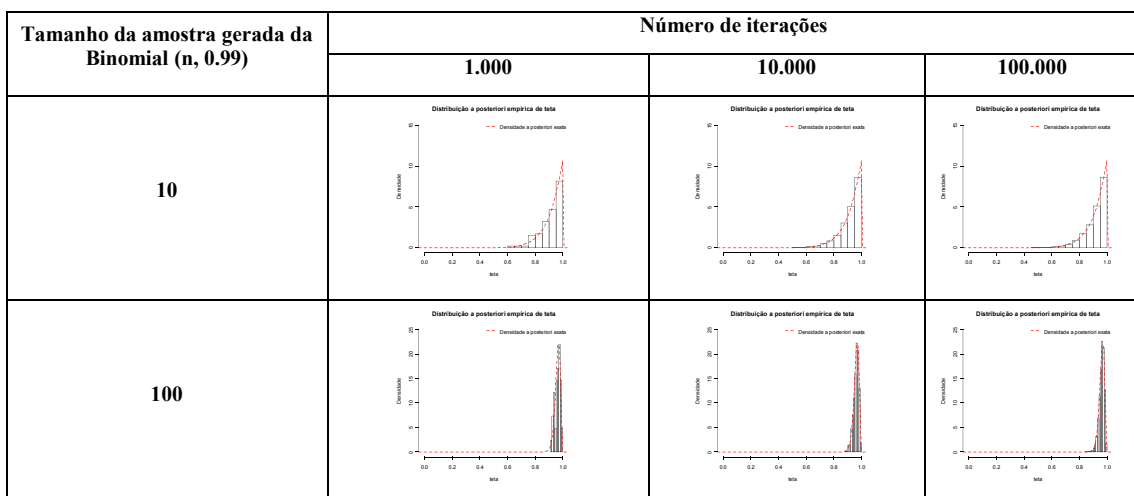


FIGURA 3 - Método da Rejeição para binomial (n,0.99).

TABELA 4 – Estatísticas do método da rejeição

<i>n</i>	Iterações	θ	Média	Estatísticas Descritivas		
			<i>a posteriori</i>	Mediana	Média	Variância
			Exata			
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

5.2.2 Resultados do Método SIR

A seguir, apresentamos os resultados do método SIR para obtenção da distribuição empírica *a posteriori* de θ , porém, como o Método SIR gasta mais tempo, escolhemos um número de iterações de 1.000 e 2.000.

As Figuras 4, 5 e 6 mostram que o método se aproxima bem da distribuição *a posteriori* (tracejado) para qualquer n e θ , mas com algumas diferenças, sendo o Método da Rejeição mais satisfatório.

Na Tabela 5, temos que, com 2.000 iterações, as variâncias são menores, porém, os tempos de simulações são bem maiores do que com 1.000 iterações. Em relação à média *a posteriori*, tanto para 1.000 quanto para 2.000 iterações, os valores estão bem próximos, sendo melhor com 2.000. Vale ressaltar que o número de iterações consideradas do algoritmo do Método SIR é bem menor se comparado com os outros métodos, pelo fato deste método ser mais demorado.

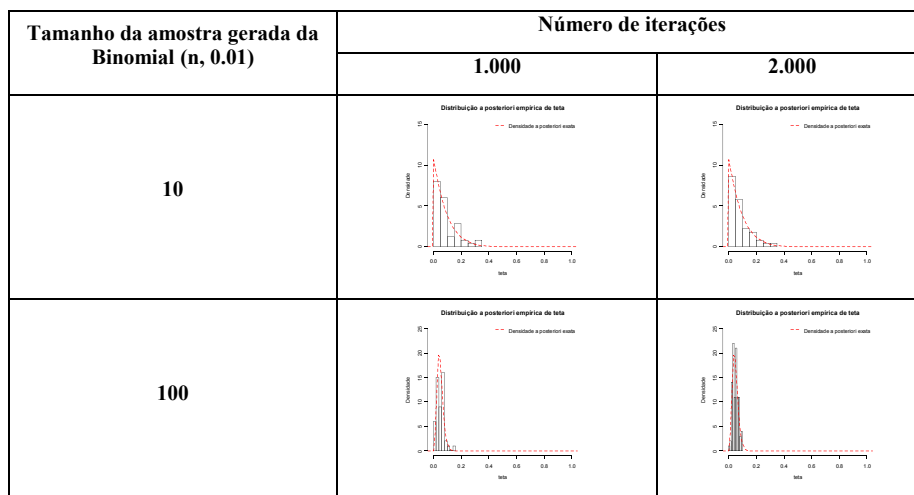


FIGURA 4 - Método SIR para binomial ($n, 0.01$).

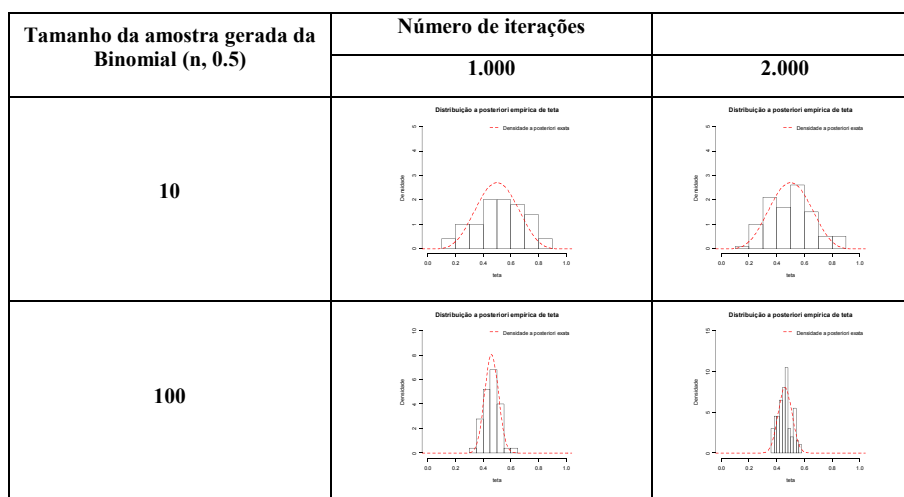


FIGURA 5 - Método SIR para binomial ($n, 0.50$).

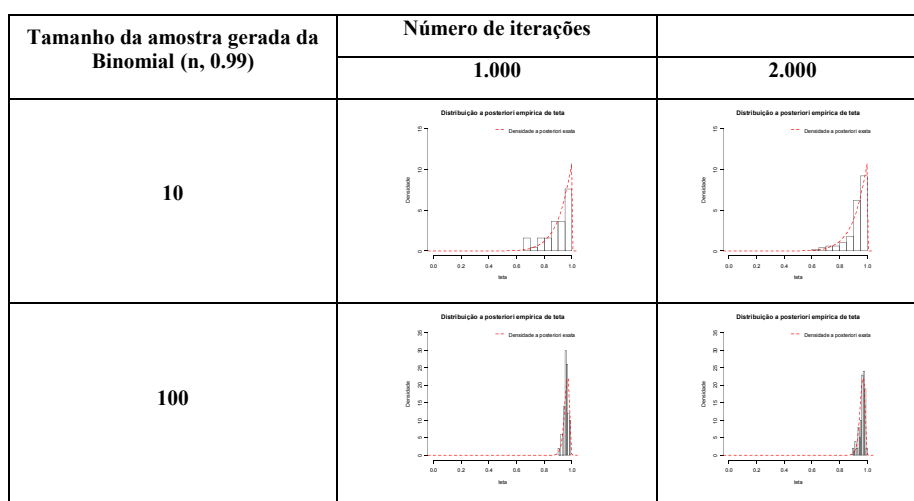


FIGURA 6 - Método SIR para binomial ($n, 0.99$).

TABELA 5 – Estatística do método SIR

n	Iterações	θ	Média <i>a posteriori</i> Exata	Estatísticas Descritivas		
				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
		0.99	0.92	0.9134	0.8892	0.0086
	2.000	0.01	0,08	0.0643	0.0785	0.0058
		0.5	0.5	0.5022	0.5008	0.0249
		0.99	0.92	0.9463	0.9216	0.0061
100	1.000	0.01	0,05	0.0494	0.0487	0.0007
		0.5	0.46	0.4602	0.4575	0.0032
		0.99	0.96	0.9589	0.9595	0.0003
	2.000	0.01	0,05	0.0494	0.0499	0.0004
		0.5	0.46	0.4579	0.4578	0.0023
		0.99	0.96	0.9682	0.9628	0.0005

6. Considerações Finais

Neste artigo, descrevemos de forma resumida, Estatística Bayesiana, algoritmos de aproximação numérica e simulação Monte Carlo, além de compararmos diferentes métodos em relação à obtenção da média *a posteriori*. As implementações de todos os programas foram satisfatórias. Todos os três tipos de quadraturas (Newton-Cotes, Trapezoidal e Simpson) foram muito rápidas e eficientes, com um número de iterações razoável e não havendo grandes diferenças entre elas. Salientamos que a dificuldade na implementação das quadraturas cresce à medida que o número de parâmetros aumenta, principalmente, para um número superior a 10. O Método da Rejeição e o SIR produziram bons resultados, porém destacamos que este segundo é o método mais lento, motivo pelo qual, optamos por um número de gerações bem menor para ele. Ressaltamos ainda que o método SIR é de grande valia quando m (14) não está disponível impossibilitando o uso do Método da Rejeição.

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.

REFERÊNCIAS

- ANTON, H. *Cálculo um novo horizonte*, Bookman, Porto Alegre: Bookman, 2000.
- FILHO, F. F. C. *Algoritmos Numéricos*. São Paulo: LTC, 2000.
- GAMERMAN, D. *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. London: Chapman and Hall, 1997.
- MIGON, H. S.; GAMERMAN, D. *Statistical Inference: an integrated approach*. Rio de Janeiro: Arnold, 1999.
- R DEVELOPMENT CORE TEAM. *R: a language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria, 2005.

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

Anexo

Neste anexo, são apresentados os programas escritos na linguagem do *freeware* R, os quais foram usados para gerar todos os experimentos do artigo.

1. Método de Newton-Cotes

```
amostra <- rbinom(100,1,0.01)
newtoncotes <- function (linf, lsup, int,
  alfa, beta, amostra)
{
#atualização dos parâmetros da
  distribuição a posteriori
  alfalinha <- alfa + sum(amostra)
  betalinha <- length(amostra) -
    sum(amostra) + beta
  h <- (lsup-linf) / int
  soma <- 0
  i<- 1
  while (i<=int)
  {
    soma <- soma + (linf+(2*i -
      1)/2*h)*(dbeta(linf+(2*i -
        1)/2*h,
        alfalinha, betalinha))
    i <- i+1
  }
  soma <- h*soma
  soma
}
mytime <- system.time(myoutput <-
  newtoncotes(0,1,1000, 1, 1, amostra))
```

2. Método Trapezoidal

```
amostra <- rbinom(100,1,0.01)
trapezoidal <- function (linf, lsup, int, alfa,
  beta, amostra)
{
#atualização dos parâmetros da
  distribuição a posteriori
  alfalinha <- alfa + sum(amostra)
```

```
  betalinha <- length(amostra) -
    sum(amostra) + beta
  h <- (lsup-linf)/int
  fa <- linf*dbeta(linf, alfalinha,
    betalinha)/2
  soma <- 0
  i<- 1
  while (i<=int)
  {
    soma <- soma + (linf+(2*i-
      1)*h/2)*dbeta(linf+(2*i-1)*h/2,
        alfalinha, betalinha)
    i<- i+1
  }
  fb <- lsup*dbeta(lsup, alfalinha,
    betalinha)/2
  soma <- h*(fa + soma+ fb)
  soma
}
mytime <- system.time(myoutput <-
  trapezoidal(0,1,1000, 1, 1, amostra))
```

3. Regra de Simpson

```
amostra <- rbinom(100,1,0.01)
simpson <- function (linf, lsup, int, alfa,
  beta, amostra)
{
#atualização dos parâmetros da
  distribuição a posteriori
  alfalinha <- alfa + sum(amostra)
```

```

betalinha <- length(amostra) -
sum(amostra) + beta
h<-(lsup-linf)/int
fa <- linf*dbeta(linf, alfa, betalinha)
soma1<- 0
i<-1
while (i<=(int/2))
{
soma1 <- soma1 + (linf + (4*i +
1)*h/2)*dbeta(linf + (4*i + 1)*h/2,
alfa, betalinha)
i<-i+1
}
soma1 <- 4*soma1
soma2<- 0
i<-1
while (i<=(int/2))
{
soma2 <- soma2 + (linf + (4*i +
3)*h/2)*dbeta(linf + (4*i + 3)*h/2,
alfa, betalinha)
i<-i+1
}
soma2 <- 2*soma2
fb <- lsup*dbeta(lsup, alfa,
betalinha)
integral <- (h/3)*(fa + soma1 + soma2 +
fb)
integral
}
mytime <- system.time(myoutput <-
simpson(0,1,1000, 1, 1, amostra))
    
```

4. Método da Rejeição

```

amostra <- rbinom(100,1,0.01)
mrejeicao <- function (tam, alfa, beta,
amostra)
{
    
```

```

#atualização dos parâmetros da
distribuição a posteriori
alfalinha <- alfa + sum(amostra)
betalinha <- length(amostra) -
sum(amostra) + beta
tetamv <- mean(amostra)
i<-1
j<-1
teta<-vector("numeric")
while(i<=tam)
{
candidato <- runif(1,0,1)
razao <- dbeta(candidato, alfa,
betalinha)/dbeta(tetamv, alfa,
betalinha)
un <- runif(1,0,1)
if (un <= razao)
{
teta[j] <- candidato
j<-j+1
}
i<-i+1
}
teta
}
mytime <- system.time(myoutput <-
mrejeicao(100000, 1, 1, amostra))
    
```

5. Método SIR

```

amostra <- rbinom(100,1,0.01)
sir <- function (nger, alfa, beta, amostra)
{
q <- vector("numeric")
w <- vector("numeric")
teta<-vector("numeric")
cestrela <- vector("numeric")
tetaestrela <- vector("numeric")
    
```



```
#atualização dos parâmetros da
distribuição a posteriori
alfalinha <- alfa + sum(amostra)
betalinha <- length(amostra) -
sum(amostra) + beta
T <- nger/20
candidatos <- rbeta(nger,alfa,beta)
w <- dbeta(candidatos, alfa,
beta)
somaw <- sum(w)
i <- 1
tamw <- length(w)
while (i <= tamw)
{
q[i] <- w[i]/somaw
i <- i+1
}
for (i in 1:T)
{
cestrela <- sample(candidatos, nger,
replace = TRUE, q)
for (j in 1:nger)
{
tetaestrela[i] <- sample(cestrela, 1,
replace=FALSE, q)
}
}
tetaestrela
}
mytime <- system.time(myoutput <-
sir(2000, 1, 1, amostra))
```