

Estudo da prevalência da tuberculose: uso de métodos bayesianos

Study of the prevalence of tuberculosis using Bayesian methods

Resumo

Neste artigo, apresentamos estimadores bayesianos para a prevalência de tuberculose usando métodos computacionais de simulação de amostras da distribuição a posteriori de interesse. Em especial, consideramos o uso do amostrador de Gibbs para simular amostras da distribuição a posteriori, e daí encontramos, em uma forma simples, inferências precisas para a prevalência de tuberculose. Em uma aplicação, analisamos os resultados do exame de Rx do tórax no diagnóstico da tuberculose. Com essa aplicação, verificamos que os estimadores bayesianos são simples de se obter e apresentam grande precisão. O uso de métodos computacionais para simulação de amostras como o caso do amostrador de Gibbs tem sido recentemente muito utilizado para análise bayesiana de modelos em bioestatística. Essas técnicas de simulação usando o amostrador de Gibbs são facilmente implementadas e não exigem muito conhecimento computacional, podendo ser programadas em qualquer software disponível. Além disso, essas técnicas podem ser consideradas para o estudo da prevalência de outras doenças.

Palavras-chave: Prevalência de tuberculose. Análise bayesiana. Amostrador de Gibbs.

Jorge Alberto Achcar

Faculdade de Medicina de Ribeirão Preto
Universidade de São Paulo
Ribeirão Preto, SP CEP 14049-900

Antonio Ruffino Netto

Faculdade de Medicina de Ribeirão Preto
Universidade de São Paulo

Abstract

In this paper we present Bayesian estimators of the prevalence of tuberculosis using computational methods for simulation of samples of posterior distribution of interest. We especially considered the Gibbs sampling algorithm to generate samples of posterior distribution, and from these samples we obtained accurate inferences for the prevalence of tuberculosis. In an application, we analyzed the results of lung X-ray tests in the diagnosis of tuberculosis. With this application, we verified that Bayesian estimators are more accurate than some existing estimators usually considered by health researchers. The use of computational methods for simulation of samples as the case of the Gibbs sampling algorithm is becoming very popular for Bayesian analysis in biostatistics. These simulation techniques using the Gibbs sampling algorithm are easily implemented and do not require great computational expertise and usually can be performed using available existing software. We could also consider these techniques for studying the prevalence of other diseases.

Key Words: Prevalence of tuberculosis. Bayesian analysis. Gibbs sampler.

Introdução

Na epidemiologia de qualquer doença infecto-contagiosa é sempre importante a localização e o diagnóstico das fontes de infecção para submetê-la ao tratamento adequado. No caso da tuberculose pulmonar existem vários testes para diagnosticar a doença. No estudo presente, a título de exemplo, vamos analisar os resultados do exame de Rx do tórax no diagnóstico da enfermidade citada.

Para calcular a prevalência da tuberculose em N pessoas submetidas a 3 exames radiológicos do tórax, o que seria equivalente, submetidos a um único exame radiológico lido por 3 observadores diferentes, segundo tenham apresentado 0, 1, 2 ou 3 exames (leituras) do Rx classificadas como positivas. Assim, teríamos a seguinte distribuição de freqüências (ver a Tabela 1):

Tabela 1 - Distribuição de freqüências (Rx positivo)

Table 1 - Frequency distribution (positive Rx)

X: número de Rx positivos	Número de pessoas classificadas
0	N_0
1	N_1
2	N_2
3	N_3
total	N

Definindo-se:

- D: o evento representando uma pessoa realmente com a doença (tuberculose);
- p: probabilidade de uma pessoa ser realmente doente ($p=P(D)$);
- p_1 : probabilidade de uma pessoa doente (D) apresentar Rx positivo (+) no primeiro exame;
- p_2 : probabilidade de uma pessoa não doente (\bar{D}) apresentar Rx positivo (+) no primeiro exame.

Também vamos assumir p_1 e p_2 constantes para os 3 exames (suposição de independência para o primeiro, segundo e terceiro exames).

Seja X uma variável aleatória representando o número de Rx positivos entre os 3 testes aplicados a cada paciente.

Assim, temos duas situações com as seguintes distribuições de probabilidade condicionais:

(1) Seja X o número de Rx positivos entre os doentes (D); a distribuição de probabilidade da variável aleatória X é dada por uma distribuição binomial $b(3, p_1)$ com função de probabilidade dada por

$$P\{X=x / D\} = \binom{3}{x} p_1^x (1 - p_1)^{3-x} \quad (1)$$

onde $x=0, 1, 2$ e 3 ;

(2) Seja X o número de Rx positivos entre os não-doentes (\bar{D}); a distribuição de probabilidade da variável aleatória X é dada por uma distribuição binomial $b(3, p_2)$ com função de probabilidade dada por

$$P\{X=x / \bar{D}\} = \binom{3}{x} p_2^x (1 - p_2)^{3-x} \quad (2)$$

onde $x=0, 1, 2$ e 3 ;

A probabilidade não condicional da variável aleatória X (número de testes positivos) é dada por,

$$P(X=x) = P(X=x, D) + P(X=x, \bar{D}) \\ = P(X=x/D)P(D) + P(X=x/\bar{D})P(\bar{D}) \quad (3)$$

Como $P(D) = p$ e $P(\bar{D}) = 1-p$, temos:

$$P(X=x) = p \binom{3}{x} p_1^x (1 - p_1)^{3-x} + (1-p) \binom{3}{x} p_2^x (1 - p_2)^{3-x} \quad (4)$$

onde $x=0, 1, 2, 3$.

Assim, temos:

$$P(X=x) = \{ p(1 - p_1)^3 + (1-p)(1 - p_2)^3 \}, \text{ se } x=0;$$

$$\{ 3pp_1(1 - p_1)^2 + 3(1-p)p_2(1 - p_2)^2 \}, \text{ se } x=1;$$

$$\{ 3pp_1^2(1 - p_1) + 3(1-p)p_2^2(1 - p_2) \}, \text{ se } x=2;$$

$$\{ pp_1^3 + (1-p)p_2^3 \}, \text{ se } x=3.$$

Pela fórmula de Bayes, podemos atualizar a probabilidade de um indivíduo ter a doença dado o número de testes positivos:

$$P(D / X=x) = P(X=x / D)P(D) / P(X=x) \quad (6)$$

onde $x=1, 2, 3$.

Assim, observamos que

$$P(D / X=x) = pp_1^x (1 - p_1)^{3-x} / [pp_1^x (1 - p_1)^{3-x} + (1-p)p_2^x (1 - p_2)^{3-x}] \quad (7)$$

onde $x=0, 1, 2, 3$.

Podemos notar que $p = P(D)$ é a probabilidade a priori de uma pessoa ter tuberculose e $P(D / X=x)$ é a probabilidade a posteriori de uma pessoa ter tuberculose, dado que foi observado x resultados positivos entre os 3 testes de Rx.

Como ilustração, vamos considerar os dados introduzidos por Ruffino-Neto¹ apresentados na Tabela 2, relativos às frequências de resultados positivos de $N = 10000$ pessoas submetidas a 3 testes de Rx de tórax.

Como o objetivo do pesquisador de saúde pública seria a estimação dos parâmetros desconhecidos p, p_1 e p_2 , usualmente se resolve o sistema de equações obtidas a partir de (5) com as frequências relativas apresentadas na Tabela 2:

$$\{ p(1 - p_1)^3 + (1-p)(1 - p_2)^3 \} = 9600/10000;$$

$$\{ 3pp_1(1 - p_1)^2 + 3(1-p)p_2(1 - p_2)^2 \} = 150/10000; \quad (8)$$

$$\{ 3pp_1^2(1 - p_1) + 3(1-p)p_2^2(1 - p_2) \} = 50/10000;$$

$$\{ pp_1^3 + (1-p)p_2^3 \} = 200/10000.$$

Após considerável álgebra, encontramos a solução do sistema (8) dada por

$$p = 0,0290; p_1 = 0,7130 \text{ e } p_2 = 0,0008.$$

A partir desses estimadores, é possível prever o número esperado de doentes $N_1 P(D/X=x)$, onde $P(D/X=x)$ é dado por (7).

Tabela 2 - Distribuição de freqüências para testes positivos entre N=10000 pessoas (Rx positivo)

Table 2 - Frequency distribution for positive tests among N = 10,000 individuals (positive Rx)

X: número de Rx positivos (+)	Número de pessoas classificadas
0	$N_0 = 9600$
1	$N_1 = 150$
2	$N_2 = 50$
3	$N_3 = 200$
total	$N = 10000$

Tabela 3 - Número esperado de doentes entre os 10000 indivíduos testados a partir do número de Rx positivos.

Table 3 - Expected number of sick individuals among the 10,000 tested individuals from the number of positive Rx tests.

X: número de Rx positivos (+)	Número de classificados	P(D / X=x)	Número esperado de doentes
0	9600	0,0007	6
1	150	0,6871	103
2	50	0,9997	50
3	200	1,0000	200
total	10000	-----	-----

Na tabela 3, temos os números esperados de doentes com tuberculose entre os N_i indivíduos com i testes de Rx positivos. O método das freqüências relativas usado para a obtenção dos estimadores pontuais para p , p_1 e p_2 em geral não apresenta bons estimadores (ver por exemplo, Bickel e Doksum⁸). Além disso, há necessidade da obtenção do erro-padrão associado ao estimador obtido.

É importante salientarmos que outros estimadores para p , p_1 e p_2 podem ser obtidos usando-se outros métodos de inferência estatística, entre os quais destacamos os métodos bayesianos (ver por exemplo, Press²). O uso de métodos Bayesianos via simulação MCMC (Monte Carlo em Cadeias de Markov) apresenta grande precisão e simplicidade para analisar dados médicos (ver por exemplo, Gelfand e Smith³).

Análise bayesiana do modelo

Vamos considerar X como o número de testes de Rx positivos entre os 3 testes aplicados a cada indivíduo. Neste caso, observamos que a variável aleatória X tem uma

distribuição de probabilidade dada por uma mistura de duas distribuições binomiais:

$$P(X = x) = \sum_{j=1}^2 \lambda_j \binom{3}{x} p_j^x (1 - p_j)^{3-x} \quad (9)$$

onde $x=0, 1, 2, 3$; $\lambda_1 + \lambda_2 = 1$; $p_1 = P(\text{Rx positivo} / D)$ e $p_2 = P(\text{Rx positivo} / \bar{D})$.

Considerando-se uma amostra aleatória de N indivíduos, a função de verossimilhança para λ_1 , p_1 e p_2 (ver, por exemplo, Mood, Graybill e Boes⁴) é dada por

$$L(\lambda_1, p_1, p_2) = \prod_{i=1}^N \sum_{j=1}^2 \lambda_j \binom{3}{x_i} p_j^{x_i} (1 - p_j)^{3-x_i} \quad (10)$$

Para uma análise Bayesiana do modelo, vamos considerar as seguintes distribuições a priori para λ_1 , p_1 e p_2 :

- (i) $\lambda_1 \sim \text{Beta}(a, b)$, onde a e b são conhecidos;
- (ii) $p_1 \sim \text{Beta}(c_1, d_1)$, onde c_1 e d_1 são conhecidos; (11)
- (iii) $p_2 \sim \text{Beta}(c_2, d_2)$, onde c_2 e d_2 são conhecidos;

onde Beta (a,b) denota uma distribuição beta com média $a/(a+b)$ e variância $ab/[(a+b)^2(a+b+1)]$. Também vamos assumir independência a priori entre λ_i, p_1 e p_2 .

Observamos em (11) que se $a=b=1$, temos uma distribuição uniforme $U(0,1)$. Este caso deve ser considerado quando não temos informação a priori sobre os parâmetros λ_i, p_1 e p_2 do modelo, isto é, assumimos uma priori não informativa para λ_i, p_1 e p_2 . Outras distribuições a priori também poderiam ser consideradas para λ_i, p_1 e p_2 . A escolha das distribuições a priori (11) foi fundamentada observando-se a variação dos parâmetros (valores entre 0 e 1).

Usando a fórmula de Bayes (ver, por exemplo, Box e Tiao⁵), encontramos a distribuição a posteriori conjunta para λ_i, p_1 e p_2 dado os dados $\mathbf{x}'=(x_1, x_2, \dots, x_N)$:

$$\Pi(\lambda_i, p_1, p_2/\mathbf{x}) \propto \lambda_1^{a-1} (1-\lambda_1)^{b-1} p_1^{c_1-1} (1-p_1)^{d_1-1} p_2^{c_2-1} (1-p_2)^{d_2-1} \cdot \left[\prod_{i=1}^N \sum_{j=1}^2 \lambda_j \binom{3}{x_i} p_j^{x_i} (1-p_j)^{3-x_i} \right] \quad (12)$$

Para encontrarmos amostras de λ_i, p_1 e p_2 da distribuição a posteriori conjunta (12), utilizamos o amostrador de Gibbs (ver, por exemplo, Gelfand e Smith³).

Para simplificação do algoritmo, introduzimos variáveis artificiais ou latentes (ver, por exemplo, Tanner e Wong⁶) para eliminarmos o produto de somatórios na distribuição a posteriori conjunta (12).

Para isso, definimos $\mathbf{m}'_i = (m_{1i}, m_{2i})$, $i=1, \dots, N$ onde $m_{1i} + m_{2i} = 1$ e m_{1i} tem uma distribuição de Bernoulli com probabilidade de sucesso h_{1i} dada por,

$$h_{1i} = \frac{\lambda_1 \binom{3}{x_i} p_1^{x_i} (1-p_1)^{3-x_i}}{\sum_{j=1}^2 \lambda_j \binom{3}{x_i} p_j^{x_i} (1-p_j)^{3-x_i}} \quad (13)$$

onde $i=1, \dots, N$.

Assim, obtemos,

$$\Pi(\mathbf{m}'_i / \lambda_i, p_1, p_2, \mathbf{x}_i) \propto h_{1i}^{m_{1i}} (1-h_{1i})^{m_{2i}} \quad (14)$$

onde $m_{2i} = 1 - m_{1i}, i=1, \dots, N$.

Isto é,

$$\Pi(\mathbf{m}'_1, \mathbf{m}'_2, \dots, \mathbf{m}'_N / \lambda_1, p_1, p_2, \mathbf{x}) \propto \prod_{i=1}^N h_{1i}^{m_{1i}} (1-h_{1i})^{m_{2i}} \quad (15)$$

Como $m_{2i} + m_{1i} = 1$, temos,

$$P(\mathbf{m}'_1, \mathbf{m}'_2, \dots, \mathbf{m}'_N / \lambda_1, p_1, p_2, \mathbf{x}) \propto (16)$$

$$\lambda_1^{r_1} (1-\lambda_1)^{r_2} p_1^{s_1} (1-p_1)^{3r_1-s_1} p_2^{s_2} (1-p_2)^{3r_2-s_2} / \left\{ \prod_{i=1}^N \sum_{j=1}^2 \lambda_j \binom{3}{x_i} p_j^{x_i} (1-p_j)^{3-x_i} \right\}$$

onde $r_1 = \sum_{i=1}^N m_{1i}; r_2 = \sum_{i=1}^N m_{2i}; s_1 = \sum_{i=1}^N m_{1i} x_i$ e $s_2 = \sum_{i=1}^N m_{2i} x_i$.

Combinando-se (12) com (16), temos:

$$\Pi(\lambda_1, p_1, p_2 / \mathbf{x}, \mathbf{m}) \propto \lambda_1^{a+r_1-1} (1-\lambda_1)^{b+r_2-1} p_1^{c_1+s_1-1} (1-p_1)^{d_1+3r_1-s_1-1} p_2^{c_2+s_2-1} (1-p_2)^{d_2+3r_2-s_2-1} \quad (17)$$

A partir de (17), encontramos as distribuições condicionais a posteriori necessárias para o amostrador de Gibbs:

- (i) $\lambda_i/p_1, p_2, \mathbf{x}, \mathbf{m} \sim \text{Beta}(a+r_1, b+r_2)$
- (ii) $p_1/\lambda_1, p_2, \mathbf{x}, \mathbf{m} \sim \text{Beta}(c_1+s_1, d_1+3r_1-s_1)$ (18)
- (iii) $p_2/\lambda_1, p_1, \mathbf{x}, \mathbf{m} \sim \text{Beta}(c_2+s_2, d_2+3r_2-s_2)$

A partir das distribuições condicionais a posteriori (18), gerar amostras da distribuição a posteriori conjunta para λ_i, p_1 e p_2 seguindo o seguinte algoritmo:

- (i) Dado um valor inicial $\lambda_1^{(0)}, p_1^{(0)}$ e $p_2^{(0)}$, gerar N observações m_{1i} da distribuição de Bernoulli com probabilidade de sucesso h_{1i} dada em (13);
- (ii) Gerar $\lambda_1^{(1)}, p_1^{(1)}$ e $p_2^{(1)}$ das distribuições condicionais a posteriori $\Pi(\lambda_1 / p_1, p_2, \mathbf{x}, \mathbf{m}), \Pi(p_1 / \lambda_1, p_2, \mathbf{x}, \mathbf{m})$ e $\Pi(p_2 / \lambda_1, p_1, \mathbf{x}, \mathbf{m})$.
- (iii) Repetir (i) e (ii) até conseguir uma distribuição estacionária.

Aplicação com os dados de prevalência de tuberculose

Vamos considerar os dados de preva-

lência da tuberculose apresentados na Tabela 2. Assumindo ignorância a priori sobre os parâmetros λ_1 , p_1 e p_2 , considerar $a=b=c_1=d_1=c_2=d_2=1$ nas distribuições a priori (11) para λ_1 , p_1 e p_2 . Neste caso, estamos considerando distribuições a priori uniformes para os parâmetros. Na tabela 2, temos 9600 valores $x_i = 0$; 150 valores $x_i = 1$; 50 valores $x_i = 2$; e 50 valores $x_i = 3$, o que totaliza $N = 10000$ observações. Assim, em cada ciclo do amostrador de Gibbs, geramos 9600 valores m_{ii} da distribuição de Bernoulli com probabilidade de sucesso $h_{ii} = \lambda_1(1 - p_1)^3 / [\lambda_1(1 - p_1)^3 + \lambda_2(1 - p_2)^3]$, $i=1, \dots, 9600$; 150 valores m_{ii} da distribuição de Bernoulli com probabilidade de sucesso $h_{ii} = \lambda_1 p_1(1 - p_1)^2 / [\lambda_1 p_1(1 - p_1)^2 + \lambda_2 p_2(1 - p_2)^2]$, $i=9601, \dots, 9750$; 50 valores m_{ii} da distribuição de Bernoulli com probabilidade de sucesso $h_{ii} = \lambda_1 p_1^2(1 - p_1) / [\lambda_1 p_1^2(1 - p_1) + \lambda_2 p_2^2(1 - p_2)]$, $i=9751, \dots, 9800$ e 200 valores m_{ii} da distribuição de Bernoulli com probabilidade de sucesso $h_{ii} = \lambda_1 p_1^3 / [\lambda_1 p_1^3 + \lambda_2 p_2^3]$ para $i=9801, \dots, 10000$.

Como valor inicial para o algoritmo, consideramos $\lambda_1^{(0)} = 0,03$; $p_1^{(0)} = 0,7$ e $p_2^{(0)} = 0,001$. É importante salientar que outros valores iniciais poderiam ser considerados.

Daí, geramos $\lambda_1^{(1)}$, $p_1^{(1)}$ e $p_2^{(1)}$ das distribuições condicionais a posteriori dadas em (18). Na Tabela 4, temos os sumários a posteriori para os parâmetros de interesse, considerando-se 1000 amostras geradas pelo amostrador de Gibbs. Na geração das amostras, simulamos cadeias com 1200 amostras e descartamos as primeiras 200 amostras para eliminar a influência dos valores iniciais para λ_1 , p_1 , p_2 . A convergência do algoritmo foi verificada a partir de métodos gráficos e de outros métodos existentes (ver, por exemplo, Gelman e Rubin⁷).

Na Tabela 4, também temos intervalos de credibilidade bayesianos com probabili-

dade de cobertura igual a 0,95. Outros valores para os hiperparâmetros foram considerados, obtendo-se inferências bayesianas similares para λ_1 , p_1 , p_2 . Considerando-se $a=b=c_1=d_1=c_2=d_2=10$, encontramos as médias a posteriori para λ_1 , p_1 e p_2 dadas por 0,02637; 0,90837 e 0,00530, respectivamente.

É importante salientar que a programação computacional para o amostrador de Gibbs não requer grande conhecimento de programação e pode ser feita usando-se linguagens usuais de programação. Neste exemplo, usamos o software MINITAB (Ryan e Joiner¹⁰), que não exige muito conhecimento de programação.

Observamos que os resultados Bayesianos da Tabela 4 apresentam grande precisão e não dependem de resultados assintóticos. Além disso, temos na Tabela 4 intervalos de credibilidade para os parâmetros do modelo, o que pode ser de grande interesse prático para pesquisadores de saúde pública. Esses estimadores são obtidos das 1000 amostras geradas pelo amostrador de Gibbs (estimativas de Monte Carlo das médias a posteriori).

Também podemos usar as amostras geradas de Gibbs para achar estimativas de Monte Carlo das probabilidades a posteriori $P(D / X=x)$ dadas por (7). Essas estimativas são apresentadas na Tabela 5. As estimativas bayesianas para $P(D / X=x)$ foram encontradas a partir de estimadores de Monte Carlo baseados nas 1000 amostras simuladas de Gibbs e dadas por:

$$\hat{P}(D / X=x) = \left(\frac{1}{1000} \right) \sum_{s=1}^{1000} \lambda_1^{(s)x} p_1^{(s)x} (1 - p_1^{(s)})^{3-x} / [\lambda_1^{(s)x} p_1^{(s)x} (1 - p_1^{(s)})^{3-x} + (1 - \lambda_1^{(s)}) p_2^{(s)x} (1 - p_2^{(s)})^{3-x}] \quad (19)$$

para $x=0, 1, 2$ e 3 .

Tabela 4 - Sumários a posteriori (dados da prevalência da tuberculose)

Table 4 - Posterior summaries (tuberculosis prevalence data)

parâmetro	média a posteriori	desvio-padrão a posteriori	intervalo de credibilidade
λ_1	0,02547	0,00163	(0,02229;0,02851)
p_1	0,92243	0,01123	(0,89804;0,94342)
p_2	0,00507	0,00043	(0,00426;0,00596)

Na Tabela 5, também temos as estimativas dos valores esperados de doentes, dado o número de Rx positivos.

Observamos que estimadores precisos foram obtidos usando-se o método bayesiano via amostrador de Gibbs. Além disso, podemos obter, de forma bem simples, intervalos de credibilidade para as probabilidades condicionais de doentes dados os resultados dos 3 testes de Rx. Intervalos de credibilidade com probabilidade de cobertura igual a 0,95 são apresentados na Tabela 6 para $P(D / X=x)$ e para os números esperados de doentes.

Algumas notas conclusivas

O uso de métodos bayesianos pode ser de grande interesse prático para a obtenção de boas inferências na área de epidemiologia. Os métodos computacionais recentes de Monte Carlo em Cadeias de Markov com

ênfase no amostrador de Gibbs levam a resultados de grande precisão e simplicidade, e modelos mais complexos, possivelmente incluindo alguns fatores de risco poderiam ser considerados sem muita dificuldade computacional adicional. Além disso, os métodos bayesianos podem incorporar opinião de especialistas da área de saúde pública, o que pode levar a melhores inferências. É importante salientar que um dos métodos clássicos de inferência mais utilizados nas aplicações é apresentado pelo método de máxima verossimilhança onde as inferências obtidas em geral são baseadas em resultados assintóticos. Observamos também que, para o modelo considerado neste artigo, podemos ter dificuldades práticas na maximização da função de verossimilhança (10), pois o modelo consiste de uma mistura de duas distribuições binomiais (ver por exemplo, Titterton et al⁹).

Tabela 5 - Estimativas bayesianas para os números esperados de doentes entre os 10000 testados

Table 5 - Bayesian estimators for the expected number of sick individuals among the 10,000 tested

X: número de Rx positivos (+)	Número de classificados	$P(D / X=x)$	Número esperado de doentes
0	9600	0,0000132	0,1269
1	150	0,0288268	4,3240
2	50	0,984951	49,2475
3	200	1,000000	200
total	10000	-----	-----

Tabela 6 - Intervalos de credibilidade 95% para $P(D / X=x)$ e para os números esperados de doentes

Table 6 - 95% credible intervals for $P(D / X=x)$ and for the expected numbers of sick individuals

X: número de Rx positivos (+)	Número de classificados	Intervalos de credibilidade para $P(D / X=x)$	Intervalos de credibilidade para os números esperados de doentes
0	9600	(0,000005;0,000028)	(0,04585;0,27156)
1	150	(0,015192;0,049765)	(2,27876;7,46472)
2	50	(0,97728;0,990509)	(48,8641;49,5255)
3	200	(0,99999;1,000000)	(199,998;199,999)

Referências

1. Ruffino-Netto A . Estudo da Prevalência da Tuberculose, *Revista de Ciência e Cultura*, 1977; 29: 1429-33.
2. Press SJ. *Bayesian Statistics: Principles, Models and Applications*. New York: John Wiley; 1989.
3. Gelfand AE, Smith AFM. Sampling Based Approaches to Calculating Marginal Densities, *J Am Stat Assoc* 1990; 85: 398-409.
4. Mood AM, Graybill FA, Boes DC. *Introduction to the Theory of Statistics*. London: McGraw-Hill; 1974.
5. Box GEP, Tiao GC. *Bayesian Inference in Statistical Analysis*. New York: Addison-Wesley; 1973.
6. Tanner MA, Wong WH. The Calculation of Posterior Distributions by Data Augmentation. *J Am Stat Assoc* 1987; 82: 528-50.
7. Gelman A, Rubin DB. Inference from Iterative Simulation Using Multiple Sequences. *Stat Sci*, 1972; 7: 457-511.
8. Bickel PJ, Doksum KA. *Mathematical Statistics: Basic Ideas and Selected Topics*. San Francisco: Holden Day; 1977.
9. Titterington DM, Smith AFM, Makov UE. *Statistical Analysis of Finite Mixture Distributions*. New York: Wiley; 1985.
10. Ryan B.F.; Joiner BL. *Minitab Handbook*, Belmont, California: Duxbury Press; 1994.

Recebido em: 03/07/2003

Versão final reapresentada em: 15/10/2003

Aprovação em: 12/11/2003