

INFERÊNCIA BAYESIANA E SUA APLICAÇÃO NA AVALIAÇÃO GENÉTICA DE BOVINOS DA RAÇA NELORE: REVISÃO BIBLIOGRÁFICA

CARINA UBIRAJARA DE FARIA,¹ CLÁUDIO DE ULHÔA MAGNABOSCO,² ARCADIO DE LOS REYES,³
RAYSILDO BARBOSA LÔBO⁴ E LUIZ ANTÔNIO FRAMARTINO BEZERRA⁵

1. Doutoranda da Universidade Federal de Goiás, bolsista CNPq, EV/UFG, Goiânia, GO, Brasil. carina@cnpaf.embrapa.br

2. Pesquisador da Embrapa Cerrados, bolsista CNPq, Planaltina, DF, Brasil. mclaudio@cpac.embrapa.br

3. Professor titular da Universidade Federal de Goiás, EV/UFG, Goiânia, GO, Brasil. adreyesb@vet.ufg.br

4. Professor associado do Departamento de Genética da FMRP-USP, Ribeirão Preto, SP, Brasil. rayblobo@fmrp.usp.br

5. Departamento de Genética da FMRP-USP, Ribeirão Preto, SP, Brasil. lafbezerra@fmrp.usp.br

RESUMO

O método da máxima verossimilhança restrita tem sido o escolhido pelos programas de melhoramento genético para estimação dos componentes de variância e predição dos valores genéticos. No entanto, a inferência bayesiana aparece como uma alternativa de grande flexibilidade, tanto em relação aos modelos que podem ser utilizados nas análises quanto em relação às inferências que podem ser realizadas a partir dos resultados. A sua aplicação

em análises genéticas permite a obtenção de densidades posteriores das variáveis estudadas e pode ser utilizada tanto em pequenos ou grandes conjuntos de dados, não sendo necessário o conhecimento da distribuição inicial do parâmetro que se deseja estimar. No Brasil, o método da Amostragem de Gibbs, que permite uma inferência bayesiana, tem sido aplicado com êxito em diversos estudos envolvendo dados de campo de bovinos da raça Nelore.

PALAVRAS-CHAVE: Amostragem de Gibbs, modelos lineares, parâmetros genéticos.

ABSTRACT

BAYESIAN INFERENCE AND ITS APPLICATION IN THE GENETIC EVALUATION OF THE NELLORE CATTLE: BIBLIOGRAPHIC REVIEW

Restricted Maximum Likelihood method has been the choice method of the genetic improvement programs to estimate variance components the breeding values. However, the bayesian inference appears as an alternative of the great flexibility, as much in relation the models that can be used in the analyses, how in relation the inferences that can be carried from the results. Its application in genetic

analyses allows obtaining the posteriors densities of the variables and can be used in small or great data sets, not being necessary the knowledge initial distribution of the parameter. In Brazil, Gibbs sampling method, that allows a bayesian inference, has been used with success in diverse studies involving field data of Nellore cattle.

KEY WORDS: Genetics parameters, Gibbs sampling, linear models.

INTRODUÇÃO

Estimativas acuradas dos componentes de variância têm sido importantes na produção animal,

uma vez que as diferenças entre o valor verdadeiro e o valor estimado aumentam o erro de predição dos efeitos aleatórios, como, por exemplo, os valores genéticos. No entanto, para a obtenção dessas esti-

mativas acuradas dos componentes de variância, é essencial identificar um método estatístico adequado e que melhor reflita o comportamento biológico das características em estudo. Neste contexto, vários procedimentos de estimação dos componentes de variância já foram propostos para aplicação no melhoramento animal.

Primeiramente, a análise de variância (ANOVA) era o método possível de se estimar esses componentes de variância, porém requeria que os dados analisados fossem balanceados. Geralmente em dados de campo, situação bastante comum no melhoramento animal, os dados não são balanceados. Portanto, não têm uma única solução para a soma de quadrados (MAGNABOSCO et al., 2002). Apesar das restrições associadas à aplicação dos métodos da ANOVA utilizando dados não balanceados, tais métodos, por serem pouco exigentes computacionalmente, foram bastante utilizados até o início da década de 1980.

Posteriormente, apresentaram-se as equações dos modelos mistos (EMM) de HENDERSON (1953) para dados não balanceados, tornando-se conhecidos como métodos I, II e III de Henderson. Os métodos de Henderson, por sua vez, foram empregados nas décadas de 1970, sendo que o método III até finais dos anos 80. Na década de 1980, o método da máxima verossimilhança (ML) passou a ser empregado na estimação de componentes de variância, porém em pequenos conjuntos de dados, uma vez que era mais exigente, computacionalmente, do que os métodos de ANOVA e de Henderson.

Na década de 1980, foi possível a aplicação do método da máxima verossimilhança restrita (REML), sugerido por PATTERSON & THOMPSON (1971), que eliminava o viés existente no método ML, por considerar a parte dos parâmetros de localização (invariante) do modelo. Atualmente, o REML é o método de escolha para estimação dos componentes de variância e seu uso foi intensificado pela disponibilidade de programas computacionais como o MTDFREML (BOLDMAN et al., 1995).

No entanto, métodos bayesianos têm propiciado novas perspectivas a questões relacionadas à estimação de componentes de variância e parâmetros

genéticos. Os métodos de Markov Chain Monte Carlo (MCMC), dentre os quais se destaca a Amostragem de Gibbs, podem ser utilizados como uma ferramenta de forma a propiciar uma inferência bayesiana. O algoritmo de Gibbs é aplicado para gerar um valor para cada parâmetro desconhecido e apresenta fácil implementação, principalmente quando comparado a algoritmos baseados em processos não derivativos, uma vez que os resultados permitem uma inferência bayesiana que gera distribuições posteriores marginais completas, a partir das quais são obtidas as estimativas dos componentes de variância e parâmetros genéticos. Diante das considerações feitas, aborda-se a inferência bayesiana e sua aplicação no melhoramento animal por meio do método da Amostragem de Gibbs e apresentam-se os resultados obtidos em alguns estudos realizados no Brasil em bovinos da raça Nelore.

Teoria freqüentista e bayesiana

A escola freqüentista desenvolveu seus métodos nas décadas de 1930 e 1940, em que se destacaram os pesquisadores Jerzy Neyman, Egon Pearson (um dos fundadores da estatística moderna) e Abraham de Wald. Os fundamentos dos métodos se baseiam em que o valor do parâmetro que se quer estimar é um valor fixo determinado, sendo que a inferência consiste apenas em observar como se distribuiriam as estimativas do parâmetro se a experiência fosse repetida infinitas vezes.

A escola bayesiana foi fundada por Laplace, por meio de vários trabalhos publicados de 1774 a 1812, e durante o século XIX ocupou um papel preponderante na inferência científica (STIGLER, 1986). Porém, em melhoramento genético animal, o método bayesiano foi introduzido por Daniel Gianola, desenvolvendo trabalhos praticamente em todas as áreas do melhoramento animal na década de 1990. Os fundamentos do método bayesiano consistem em descrever todos os erros que podem existir em torno de um parâmetro, usando como medida do erro a probabilidade de que o parâmetro tome determinados valores. No caso em que se necessita de uma estimação pontual de dado parâmetro pode se usar a moda, média ou mediana.

Os adeptos do uso das técnicas bayesianas asseguram que o método oferece respostas a problemas até então não resolvidos em melhoramento animal. De acordo com GIANOLA et al. (1994), com o uso dos métodos freqüentistas, como é o caso do REML, obtêm-se somente estimativas pontuais da variância genética. Assim, diante de uma simples medida de erro, que só teria sentido em amostras muito grandes, e diante de dados distribuídos de forma normal e em análises em que não há o conhecimento das médias e variâncias, haveria de se utilizar as estimativas da máxima verossimilhança como se fossem os parâmetros verdadeiros, ignorando o seu erro. As análises bayesianas, porém, permitem calcular a densidade marginal posterior do parâmetro de interesse e com tal distribuição quantifica-se exatamente o erro de um parâmetro desconhecido, apresentando maior precisão dos parâmetros estimados.

Assim, as formas de apresentação dos resultados são bem diferentes entre os métodos freqüentistas e bayesianos. A escola bayesiana trabalha com probabilidades e suas estimativas são distribuições de densidade probabilística do parâmetro em estudo, e em uma amostra determinada. A escola freqüentista gera um valor estimado e um intervalo de confiança, podendo oferecer distribuições apenas quando se repetir por infinitas vezes o experimento. Mas, neste caso, a interpretação também é completamente diferente, pois estaria obtendo uma distribuição de um estimador e não de inferências sobre o parâmetro, como no método bayesiano. Vale assinalar, porém, que alguns conceitos de uso freqüente no melhoramento genético, como a diferença entre efeitos fixos e aleatórios, não são considerados na escola bayesiana, visto que o método considera todos efeitos como aleatórios (BLASCO, 2001).

A teoria bayesiana teve sua fundamentação nos trabalhos desenvolvidos por Thomas Bayes em 1761, que só foram publicados dois anos após sua morte, em 1763. Os métodos e probabilidade inversa baseiam-se no Teorema de Bayes (GIANOLA & FERNANDO, 1986), o qual estabelece que dado um vetor não observável q e um vetor de dados observáveis y , com distribuições de densidade de

probabilidade conjunta $p(\mathbf{Y}, q)$, da teoria básica de probabilidades, tem-se:

$$p(\mathbf{Y} | q) \cdot p(q) = p(\mathbf{Y}, q) = p(q | \mathbf{Y}) \cdot p(\mathbf{Y})$$

em que $p(q)$ e $p(\mathbf{Y})$ são as densidades de probabilidade marginais de q e \mathbf{Y} , respectivamente. Das igualdades acima pode ser obtido que:

$$\theta \quad \frac{\theta}{\theta}$$

Com base no argumento que $p(\mathbf{Y})$ não é uma função de q , o Teorema de Bayes é geralmente apresentado na forma:

$$p(q | \mathbf{Y}) \propto p(q) p(\mathbf{Y} | q)$$

em que $p(q)$ é a densidade de probabilidade *a priori* de q , a qual reflete o grau de conhecimento acumulado sobre os possíveis valores de q , antes da obtenção de informações contidas em \mathbf{Y} ; $p(\mathbf{Y} | q)$ é a função de verossimilhança de \mathbf{Y} , que representa a contribuição de \mathbf{Y} ao conhecimento de q ; e $p(q | \mathbf{Y})$ é a densidade de probabilidade *posterior* de q , a qual inclui o grau de conhecimento prévio sobre q “atualizado por informações adicionais” contidas em \mathbf{Y} ; μ é o operador *proporcional a*.

Mais recentemente, a introdução de métodos de Monte Carlo baseados em cadeias (seqüências) de Markov, denominados MCMC (do inglês Markov Chain Monte Carlo), tem contribuído substancialmente no sentido de viabilizar a implementação do paradigma bayesiano (SORENSEN, 1996). Os métodos MCMC constituem uma família de processos iterativos para aproximar a geração de amostras de distribuições multivariadas (em processos Monte Carlo, em propriedades das cadeias de Markov). A amostragem de Gibbs é um desses métodos.

A amostragem de Gibbs

A amostragem de Gibbs é um procedimento de integração numérica, usada na estimação das distribuições conjunta e marginal de todos os parâmetros do modelo, por meio da reamostragem de todas as distribuições condicionais da cadeia de Markov (BLASCO, 2001). É uma técnica para a obtenção indireta de variáveis aleatórias de uma distribuição (marginal), sem a necessidade de cálculo de sua densidade de probabilidade. O programa MTGSAM

(*Multiple Trait Gibbs Sampler under Animal Model*), desenvolvido por VAN TASSEL & VAN VLECK (1996), utiliza a aplicação do método Iterativo de Gauss Seidel sobre as equações de modelos mistos para a obtenção dos valores iniciais dos efeitos fixos e aleatórios.

De acordo com VAN TASSEL et al. (1995), o uso da amostragem de Gibbs apresenta várias vantagens com relação aos métodos usuais: (i) não requer soluções para as equações de modelo misto; (ii) permite a análise de conjunto de dados maiores do que quando se usa REML com técnicas de matrizes esparsas; (iii) propicia estimativas diretas e acuradas dos componentes de variância, valores genéticos e intervalos de confiança para essas estimativas e, (iv) pode ser usado sem problemas em microcomputadores e estações de trabalho.

Para a implementação do algoritmo de Gibbs, são necessários cuidados especiais com a definição das análises como a distribuição inicial, critério de convergência, período de descarte amostral, intervalo amostral e tamanho de cadeia amostral, conforme descrito por MAGNABOSCO (1997):

a) Distribuições iniciais

Um aspecto importante nas análises com o MTGSAM é a necessidade de se especificar os parâmetros que determinam a forma das distribuições iniciais assumidas para cada parâmetro estudado. A distribuição pode ser assumida como distribuição *flat* ou não informativa (não refletindo conhecimento prévio do parâmetro), simétrica (permite alguma informação do parâmetro) e aguda (*sharp*) ou informativa (refletindo conhecimento do parâmetro). Dessa forma, parâmetros de definição da forma (n) da distribuição inicial podem ser definidos conforme requerido pelo MTGSAM.

b) Critério de convergência

O MTGSAM utiliza o método Iterativo de Gauss Seidel (IGS) sobre as equações de Modelos Mistos para obter os valores iniciais dos efeitos fixos e aleatórios para iniciar a Amostragem de Gibbs (GS). O método IGS determina a solução de um sistema linear por aproximações, usando operações aritméticas a partir de dado vetor de aproximação

inicial. O aplicativo utiliza valores iniciais iguais a zero para todos os efeitos fixos e aleatórios. As iterações são repetidas até que um número máximo de iterações estabelecido ou o critério de convergência seja alcançado.

c) Período de descarte amostral

Em termos práticos, o período de descarte amostral é o número de ciclos que devem ser descartados antes que as amostras produzidas pelo GS possam ser consideradas amostras das distribuições posteriores, ou seja, a convergência ou distribuição tenha ocorrido. É importante ressaltar que programas como o GIBBSIT (RAFTERY & LEWIS, 1994) podem ser usados para determinação do período ideal de descarte para cada conjunto de dados.

d) Intervalo amostral

Para se estimar muitos parâmetros para uma distribuição, uma amostra de observações independentes se faz necessária. Amostras consecutivas obtidas de uma cadeia GS podem ser altamente correlacionadas. Dessa forma, para se obter amostras independentes ou não correlacionadas é necessário que se eliminem resultados de ciclos intermediários da cadeia amostral. O intervalo de retirada das amostras é denominado intervalo de utilização amostral e deve ser suficiente para que as amostras utilizadas não sejam correlacionadas. Para se determinar o intervalo de utilização amostral, podem ser usados programas como o GIBBSIT (RAFTERY & LEWIS, 1994), por exemplo.

e) Tamanho de cadeia amostral

O algoritmo de GS produz uma cadeia de amostras de Gibbs de tantos ciclos quanto forem especificados. Este aspecto é muito importante, porque está intimamente relacionado com a convergência das distribuições posteriores dos parâmetros. Se o número de ciclos definido pelo usuário for adequado, as médias posteriores das amostras serão estimativas válidas das distribuições posteriores dos parâmetros. Para se verificar a acurácia dessas estimativas pode ser utilizado o erro de Monte Carlo (SORENSEN et al., 1994). Esse erro decorre do

número de amostras usado da cadeia de Gibbs, sendo, portanto, inversamente proporcional ao tamanho da cadeia.

Aplicação da amostragem de Gibbs em alguns estudos realizados no Brasil em dados de campo de bovinos da raça Nelore

Os primeiros estudos no Brasil envolvendo a implementação do método bayesiano através da Amostragem de Gibbs, em melhoramento genético animal, tiveram início em 1997, em que MAGNABOSCO (1997) obteve as estimativas de componentes de variância e parâmetros genéticos utilizando dados de campo de bovinos da raça Nelore.

Em seguida, LÔBO et al. (1997), analisando dados obtidos dos arquivos do Programa de Melhoramento Genético da Raça Nelore, estimaram os componentes de variância e herdabilidades para o peso aos 120 dias de idade, utilizando o método REML e GS. Esses autores verificaram valores bem mais altos para a herdabilidade média pelo método bayesiano do que a estimação pontual fornecida pelo método REML, porém as distribuições posteriores dos parâmetros mostraram valores similares à literatura.

Posteriormente, utilizando dados de campo de bovinos Nelore para peso aos 205 dias de idade, MAGNABOSCO (2000) verificou a influência das distribuições iniciais para os componentes de variância: não-informativa e informativa. Utilizaram-se 5.000 cadeias de Gibbs e, após um período de descarte amostral de 200 ciclos, um total de 160 amostras dos parâmetros do modelo linear foi retirado em intervalo amostral de 30 ciclos. Pode ser observado na Figura 1 que a forma da distribuição inicial influenciou nas distribuições marginais posteriores de todos os parâmetros genéticos, se afastando de uma distribuição normal e, nesse caso, os valores modais foram mais adequados a serem utilizados como estimativas das densidades marginais posteriores. Além disso, o autor concluiu que, quando distribuições iniciais não são apropriadas e não

refletem o conhecimento dos valores iniciais utilizados, resultados que não espelham os verdadeiros valores biológicos podem estar sendo fornecidos pelo Amostrador de Gibbs.

Outro critério importante na implementação do GS é a convergência que deve ser alcançada para que as diferentes estimativas obtidas pelas amostras das cadeias de Gibbs não sejam distintas entre si e permitam as inferências marginais sobre os componentes de variância (WANG et al., 1993). Em outra análise, MAGNABOSCO (2000) utilizou um período de descarte amostral de 200 ciclos, e um total de 160 e 660 amostras dos parâmetros do modelo linear foi retirado, com um intervalo amostral de 20 ciclos em tamanho de cadeia amostral, de 5.000 e 20.000 ciclos, respectivamente. A Figura 2 mostra os resultados obtidos das estimativas das densidades posteriores para os parâmetros genéticos obtidos a partir de diferentes tamanhos de cadeia amostral. Os parâmetros genéticos apresentaram densidades marginais posteriores muito mais estáveis com tamanho de cadeia de 20.000, porém as distribuições marginais posteriores ainda se apresentaram distintas de uma distribuição normal.

Embora as estimativas obtidas pelo GS tenham sido similares às estimativas pontuais do método REML, MAGNABOSCO et al. (2001) utilizaram o método GS para estimar os parâmetros genéticos para peso aos 205 dias em bovinos da raça Nelore, considerando um período de descarte amostral de 2.000 ciclos, obtendo 490 e 990 amostras dos parâmetros do modelo linear retiradas com um intervalo amostral de 200 ciclos, e tamanho de cadeia amostral de 100.000 e 200.000 ciclos, respectivamente. Os autores observaram que os parâmetros genéticos apresentaram densidades marginais posteriores muito mais estáveis com tamanho de cadeia longa, aproximando-se de uma distribuição normal (Figura 3). GARCIA-CORTÉS et al. (1998) relataram que a convergência amostral dificilmente é alcançada com tamanhos de cadeia relativamente pequenos, o que pode sugerir tamanhos de cadeia de até 1.000.000 ciclos.

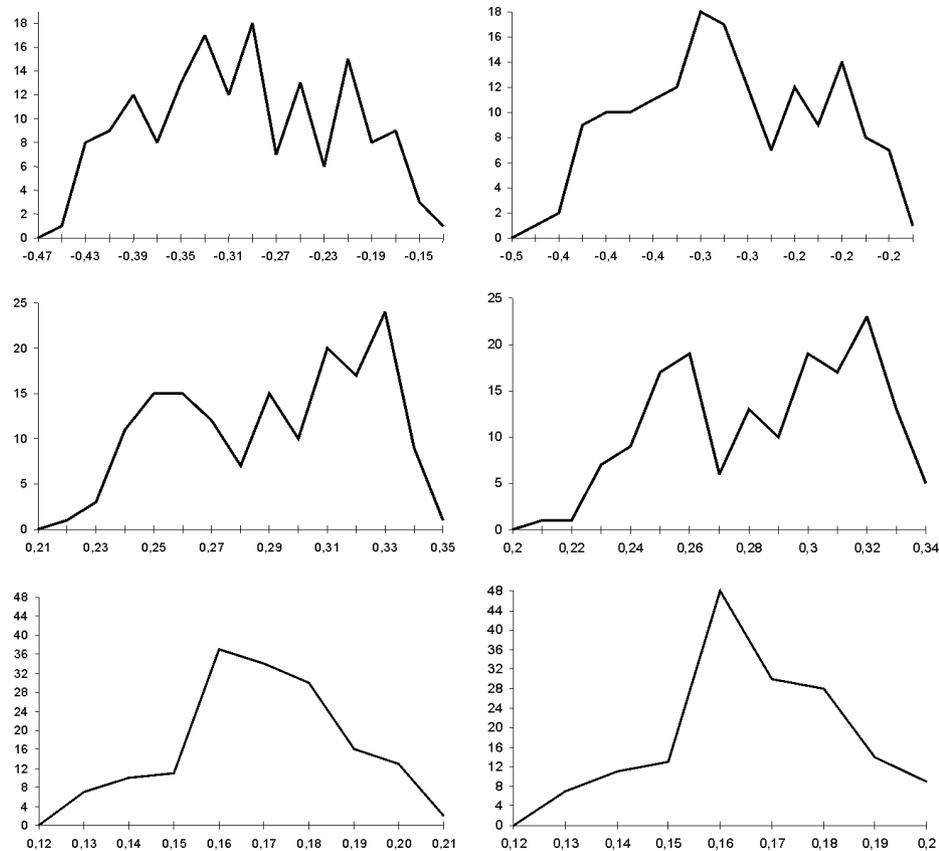


FIGURA 1. Densidades posteriores da correlação genética direta-maternal (superior), herdabilidade direta (central) e herdabilidade maternal (inferior) para peso aos 205 dias de bovinos da raça Nelore com formas da distribuição inicial ν_0 e ν_9 (MAGNABOSCO, 2000).

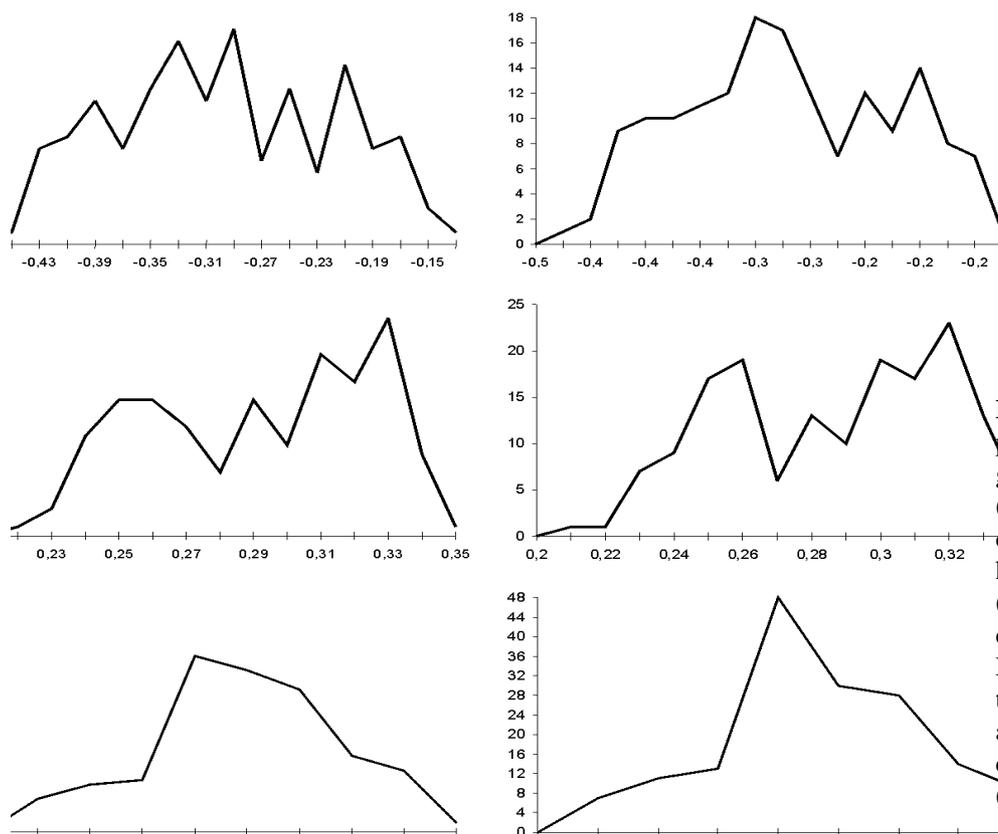


FIGURA 2. Densidades posteriores da correlação genética direta-maternal (superior), herdabilidade direta (central) e herdabilidade maternal (inferior) para peso aos 205 dias de bovinos da raça Nelore com diferentes tamanhos de cadeia amostral, 5.000 e 20.000 ciclos, respectivamente (MAGNABOSCO, 2000).

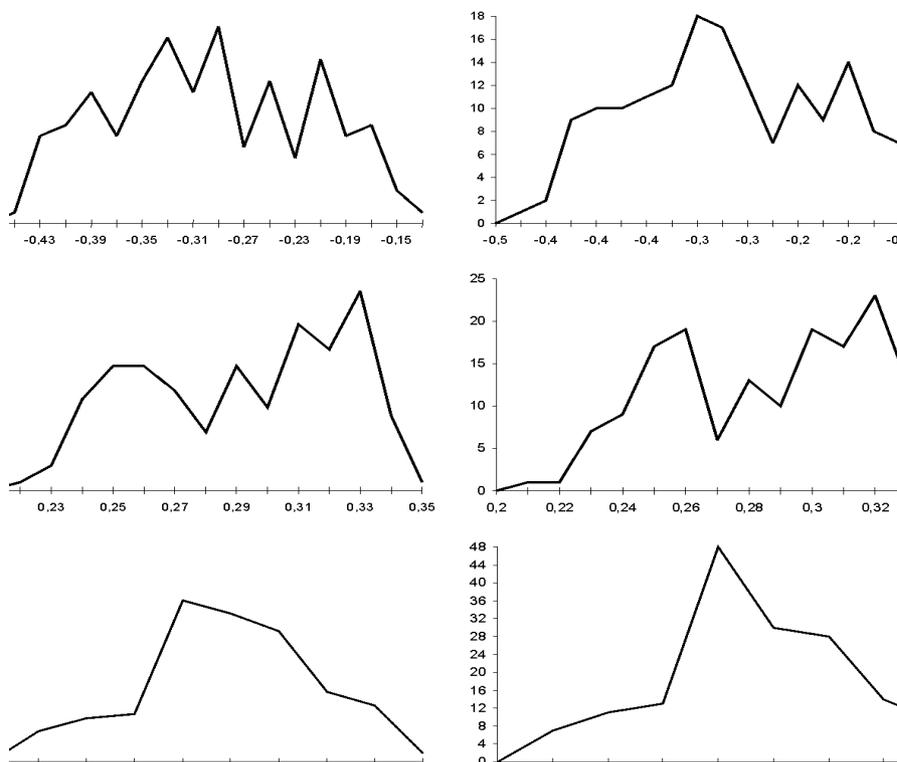


FIGURA 3. Densidades posteriores da correlação genética direta-maternal (superior), herdabilidade direta (central) e herdabilidade maternal (inferior) para peso aos 205 dias de bovinos da raça Nelore com diferentes tamanhos de cadeia amostral, 100.000 e 200.000 ciclos, respectivamente (MAGNABOSCO et al., 2001).

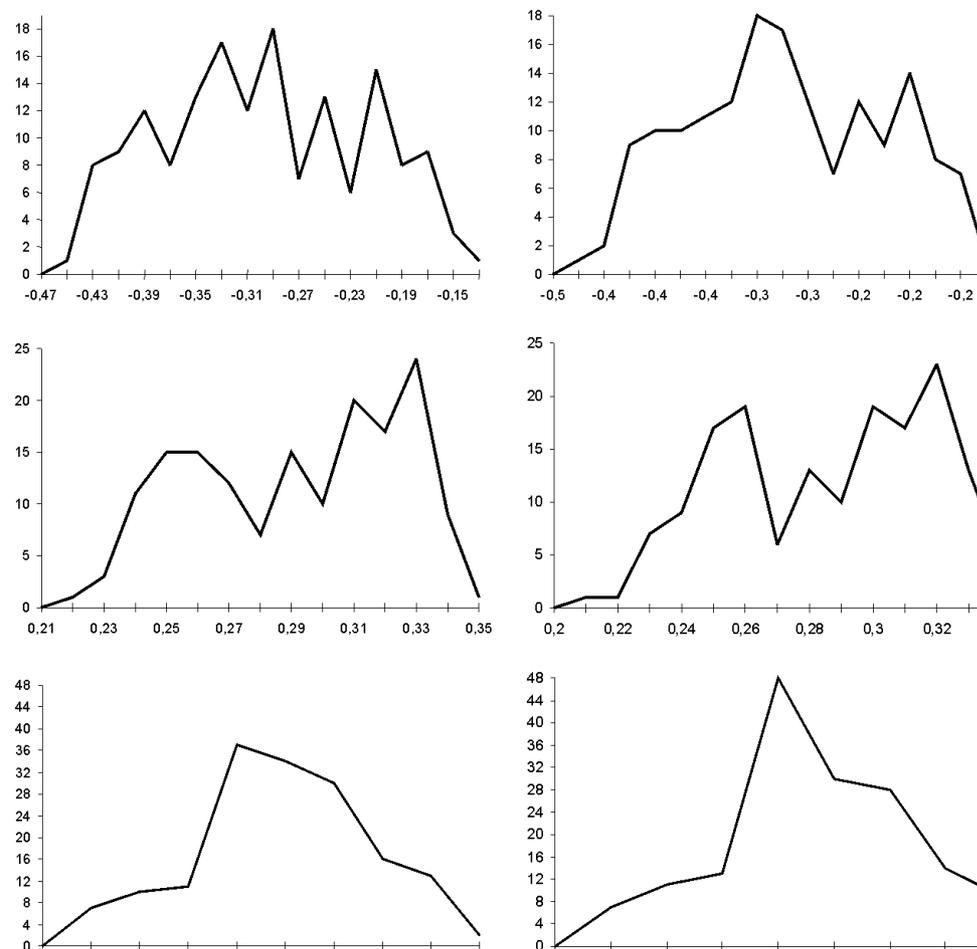


FIGURA 4. Histogramas das estimativas de densidades posteriores da variância genética aditiva direta (s_a^2), variância residual (s_e^2) e herdabilidade direta (h_d^2) para a idade ao primeiro parto (superior), peso adulto (central) e produtividade acumulada (inferior), considerando as análises unicarateres (FARIA, 2003).

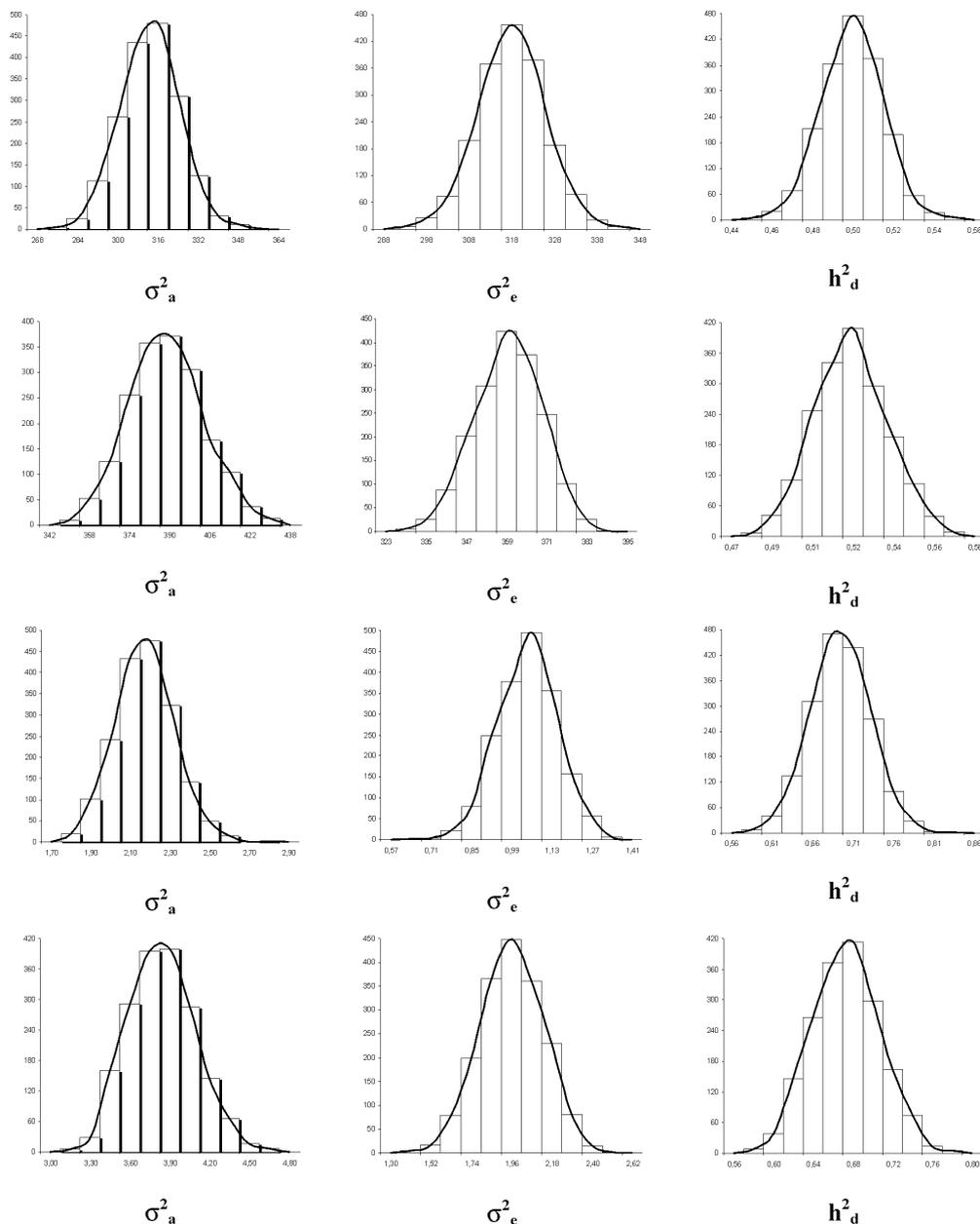


FIGURA 5. Histogramas das estimativas de densidades posteriores da variância genética aditiva direta (σ^2_a), variância residual (σ^2_e) e herdabilidade direta (h^2_d) para peso aos 365 (superior) e 450 dias de idade (centro superior), perímetro escrotal aos 365 (centro inferior) e 450 dias de idade (inferior), considerando as análises unicarateres (FARIA, 2003).

Considerando os cuidados na implementação do GS, FARIA (2003) utilizou o método para estimação dos parâmetros genéticos de características de reprodução e crescimento de bovinos da raça Nelore. Foram retiradas amostras a cada 100 ciclos após a eliminação dos 20.000 ciclos iniciais, em tamanho de cadeia de 200.000 ciclos, gerando 1.800 amostras das médias posteriores dos componentes de variância, em análises unicarateres. Os valores dos coeficientes de herdabilidade encontrados foram similares com aqueles reportados na literatura em aná-

lises genéticas utilizando o método REML. Pode ser observado na Tabela 1 que as estimativas de média, moda e mediana para os componentes de variância das características reprodutivas apresentaram-se semelhantes, sendo que, para os valores de herdabilidade, as medidas de tendência central tiveram valores idênticos. O mesmo comportamento pode ser observado nas estimativas dos componentes de variância e herdabilidade das características de crescimento, ressaltando-se que, para o perímetro escrotal aos 450 dias de idade, a moda da estimati-

va de herdabilidade foi maior que a média e a mediana, porém esta diferença foi de baixa magnitude.

Conforme citado por FARIA (2003), quanto maior o número de simulações menor será a variância da média, e as distribuições marginais posteriores das variâncias tendem à simetria, ou seja, apresentaram-se como uma distribuição normal. Apesar disso, as 1.800 amostras para cada característica não foram suficientes para que as distribuições assumissem o comportamento de uma distribuição normal;

no entanto, espera-se uma distribuição qui-quadrado invertida das densidades posteriores dos componentes de variância. As variâncias das amostras, apresentadas na Tabela 1, foram baixas para todas as características em estudo. Tais resultados são congruentes com o número de amostras utilizadas, pois, quanto maior o número de amostras obtidas, menor é a variância entre elas, considerando-se o período de descarte das amostras iniciais.

TABELA 1. Estatística descritiva dos componentes de variância e de herdabilidade para idade ao primeiro parto (IPP), peso adulto (PA) e produtividade acumulada (PAC), peso aos 365 (P365) e 450 dias de idade (P450), perímetro escrotal aos 365 (PE365) e 450 dias de idade (PE450), obtidos a partir de análises unicarateres.

Característica		Média	Moda	Mediana	s^2_k	DP	IC
IPP	s^2_a	1,85	1,90	1,85	0,0264	0,16	1,53 a 2,17
	s^2_e	5,29	5,25	5,29	0,0188	0,14	5,01 a 5,57
	h^2_d	0,26	0,26	0,26	0,0004	0,02	0,22 a 0,30
PA	s^2_a	795,25	768,88	793,17	6430,46	80,19	634,87 a 955,63
	s^2_e	1420,5	1418,5	1420,6	4597,7	67,80	1284,9 a 1556,1
	h^2_d	0,36	0,36	0,36	0,0010	0,03	0,30 a 0,42
PAC	s^2_a	105,40	105,48	105,25	175,83	13,26	78,88 a 131,92
	s^2_e	318,04	317,97	318,02	123,19	11,10	295,84 a 340,24
	h^2_d	0,25	0,25	0,25	0,0008	0,03	0,19 a 0,31
P365	s^2_a	380,80	309,28	309,03	139,52	11,81	357,18 a 404,42
	s^2_e	315,44	317,40	315,44	63,71	7,98	299,48 a 331,40
	h^2_d	0,49	0,49	0,49	0,0002	0,01	0,47 a 0,51
P450	s^2_a	384,32	385,71	383,89	220,49	14,85	354,62 a 414,02
	s^2_e	356,65	356,09	357,00	99,24	9,96	336,73 a 376,57
	h^2_d	0,52	0,52	0,52	0,0002	0,02	0,50 a 0,54
PE365	s^2_a	2,12	2,15	2,12	0,0219	0,15	1,82 a 2,42
	s^2_e	1,01	1,00	1,01	0,0107	0,10	0,81 a 1,21
	h^2_d	0,68	0,68	0,68	0,0013	0,04	0,60 a 0,76
PE450	s^2_a	3,76	3,83	3,76	0,0623	0,25	3,26 a 4,26
	s^2_e	1,91	1,85	1,91	0,0294	0,17	1,57 a 2,25
	h^2_d	0,66	0,68	0,66	0,0012	0,03	0,60 a 0,72

s^2_a : variância genética aditiva direta; s^2_e : variância residual; h^2_d : herdabilidade direta; s^2_k : variância das amostras; DP: desvio padrão; IC: intervalo de confiança de 95%. / Fonte: FARIA (2003).

Assim, quando a Amostragem de Gibbs foi corretamente implementada, as distribuições marginais posteriores se apresentam mais estáveis e tendem à normalidade (BLASCO et al., 1998; GARCIA-CORTÉS et al., 1998). Desse modo, as distribuições posteriores obtidas por FARIA (2003) comprovam que a análise bayesiana aparece como uma alternativa de grande flexibilidade nas análises genéticas gerando estimativas acuradas, oriundas de uma distribuição marginal posterior, diferente das estimativas pontuais do REML. OLIVEIRA et al. (2000), utilizando informações de peso aos 365 dias de idade em bovinos da raça Nelore, encontraram estimativas semelhantes entre aquelas obtidas pelo método REML e pela inferência bayesiana. No entanto, os autores afirmaram que a inferência bayesiana permitiu estabelecer regiões de maior densidade dos parâmetros, gerando estimativas mais precisas. As estimativas de componentes de variância e herdabilidade obtidas por FARIA (2003) para as características de reprodução e crescimento apresentaram valores análogos aos reportados pela literatura, porém os resultados foram maiores do que aqueles encontrados em estudo prévio realizado com o mesmo arquivo de dados (DIAS, 2001) em análises univariadas utilizando o método REML. No entanto, esses resultados são explicados por MAGNABOSCO (1997), ao constatar que as estimativas geradas pelo método da Amostragem de Gibbs tendem a apresentar estimativas superiores às do método REML na ordem de 10% a 15%.

Considerando análises bicarateres, MAGNABOSCO et al. (2003) compararam resultados obtidos pelo GS e REML utilizando dados de bovinos da raça Nelore para perímetro escrotal aos 365 e 450 dias de idade. As herdabilidades encontradas pelo método REML foram 0,50 e 0,55, respectivamente. Já para as análises utilizando-se o GS, as estimativas obtidas foram de 0,63 e 0,62, apresentando magnitudes superiores quando comparadas àquelas obtidas pelo REML. No entanto, as distribuições marginais posteriores apresentaram tendência à normalidade.

De acordo com FALCÃO et al. (2004), a abordagem bayesiana é capaz de produzir com exatidão as distribuições conjuntas e marginais posterior-

res para qualquer tamanho de amostra, as quais dão uma descrição completa do estado de incerteza a *posteriori* dos dados. Dessa forma, a aplicação da inferência bayesiana por meio da Amostragem de Gibbs apresentou vantagens sobre os métodos clássicos de estimação dos componentes de variância, uma vez que foi possível obter não somente as estimativas pontuais, mas também as distribuições posteriores dos parâmetros.

CONSIDERAÇÕES FINAIS

O método para estimação dos componentes de (co)variância e parâmetros genéticos de maior utilização e adotado pelos programas de melhoramento genético é a Máxima Verossimilhança Restrita. No entanto, a Amostragem de Gibbs, que em virtude de suas propriedades propicia uma inferência bayesiana, é uma metodologia adequada para aplicação no melhoramento genético animal.

De acordo com o resultados apresentados, a Amostragem de Gibbs apresentou-se como uma excelente alternativa para estimação dos componentes de (co)variância e parâmetros genéticos, pois permite a obtenção de estimativas acuradas, apresentando maior flexibilidade, pelo fato de as distribuições marginais posteriores gerarem inferências mais precisas. No entanto, são necessários certos cuidados na implementação e aplicação do algoritmo de Gibbs. Por isso, recomenda-se que mais estudos sejam desenvolvidos buscando a democratização deste método, para que mais usuários dessa poderosa ferramenta – que é a inferência bayesiana – possam aplicá-la em resultados obtidos a partir da Amostragem de Gibbs, em dados de campo.

REFERÊNCIAS

- BLASCO, A. The Bayesian controversy in animal breeding. **Journal of Animal Science**, v. 79, p. 2023-2046, 2001.
- BLASCO, A.; SORENSEN, D.; BIDANEL, J. P. Bayesian inference of genetic parameters and selection response for litter size components in pigs. **Genetics**, v. 149, p. 301-306, 1998.

- BOLDMAN, K. G.; KRIESE, L. A.; VAN VLECK, L. D.; VAN TASSELL, C. P.; KACHMAN, S. D. **A manual for of MTDFREML. A set of Programs to obtain estimates of variances and covariances (DRAFT)**. U.S.: Department of Agriculture, Agricultural Reserch Service, 1995. 115 p.
- DIAS, D. S. O. **Análise genética de características de crescimento e de reprodução de animais da raça Nelore criados na região Centro-Oeste do Brasil**. 2001. 160 f. Tese (Doutorado em Zootecnia) – Faculdade de Ciências Agrárias e Veterinárias, Universidade Estadual Paulista, Jaboticabal.
- FALCÃO, A. J. S.; MARTINS, E. N.; COSTA, C. N.; SAKAGUTI, E. S.; MAZUCHELI, J. Aplicação de métodos REML e bayesiano via amostrador de Gibbs na estimação de componentes de variância para produção de leite no estado do Paraná. In: SIMPÓSIO DA SOCIEDADE BRASILEIRA DE MELHORAMENTO ANIMAL, 5., 2004, Pirassununga. **Anais eletrônicos...** Pirassununga: SBMA, 2004. CD-ROM.
- FARIA, C. U. **Inferência bayesiana no estudo genético quantitativo de características reprodutivas e de crescimento de bovinos da raça Nelore**. 2003. 74 f. Dissertação (Mestrado) – Escola de Veterinária, Universidade Federal de Goiás, Goiânia.
- GARCIA-CORTÉS, L. A.; RICO, M.; GROENEVELD, E. Using Coupling with the Gibbs Sampler to Asses Convergence in Animal Models. **Journal of Animal Science**, v. 76, p. 441-447, 1998.
- GIANOLA, D.; FERNANDO, R. L. Bayesian methods in animal breeding theory. **Journal of Animal Science**, Champaign, v. 63, p. 217-244, 1986.
- GIANOLA, D.; RODRIGUEZ-ZAS, S.; SHOOK, G. E. The Gibbs sampler in the animal model: a primer. In: FOULLEY, J. L.; MOLENAT, H. (Ed.). SÉMINAIRE MODELE ANIMAL. INRA Departament de Genetique Animale, La Colle sur Loup, France, p. 47-56, 1994.
- HENDERSON, C. R. Estimation of variance and covariance components. **Biometrics**, v. 9, p. 226, 1953.
- LÔBO, R. B.; OLIVEIRA, H. N.; BEZERRA, L. A. F.; MARTINS, E. N.; REYES, A. de los; MAGNABOSCO, C. D. U.; RUSSO, A.R.V. Estimativa de componentes de (co)variância e herdabilidade para o peso aos 120 dias de idade na raça Nelore usando estatística bayesiana. In: REUNIÃO ANUAL DA SOCIEDADE BRASILEIRA DE ZOOTECNIA, 34, 1997, Juiz de Fora. **Anais eletrônicos...** Juiz de Fora: SBZ, 1997. CD-ROM
- MAGNABOSCO, C. D U. **Estimativas de parâmetros genéticos em características de crescimento de animais da raça Nelore usando os métodos de máxima verossimilhança restrita e amostragem Gibbs**. 1997. 89 f. Tese (Doutorado em Ciências) – Faculdade de Medicina, Universidade de São Paulo, Ribeirão Preto.
- MAGNABOSCO, C. D U. **Método alternativo para estimação dos componentes de variância: amostragem de Gibbs**. Relatório Técnico-Científico – CNPq, Planaltina: Embrapa Cerrados, 2000. 24 f.
- MAGNABOSCO, C. U.; FARIA, C. U.; REYES, A. de Los; LÔBO, R. B.; SAINZ, R. D. Inferência bayesiana para peso ao desmame em bovinos da raça Nelore. In: REUNIÓN LATINO AMERICANA DE PRODUCCIÓN ANIMAL, 17., 2001, Havana. **Anais eletrônicos...** Havana: Asociación Latinoamericana de Producción Animal, 2001. CD-ROM.
- MAGNABOSCO, C.D.U.; FARIA, C.U.; REYES, A. de los.; LÔBO, R.B.; SAINZ, R.D. **Implementação da amostragem de Gibbs na estimação de componentes de (co)variância parâmetros genéticos em dados de campo de bovinos Nelore**. Planaltina:Embrapa Cerrados, 2002. v. 1. 46 p. [Boletim de Pesquisa].

MAGNABOSCO, C. D. U.; DIAS, D. O.; FARIA, C. U.; TONHATI, H.; REYES, A. de los; LÔBO, R. B. Estudo genético quantitativo do perímetro escrotal em análise multivariada utilizando dados de campo de bovinos da raça Nelore. In: REUNIÃO ANUAL DA SOCIEDADE BRASILEIRA DE ZOOTECNIA, 40., 2003, Santa Maria. **Anais eletrônicos...** Santa Maria: SBZ, 2003. CD-ROM.

OLIVEIRA, H. N.; LÔBO, R. B.; ROSA, G. J. M.; BEZERRA, L. A. F. Análise bayesiana na estimação de componentes de (co)variância e predição de valores genéticos do peso aos 365 dias de bovinos Canchim. In: REUNIÃO ANUAL DA SOCIEDADE BRASILEIRA DE ZOOTECNIA, 37., 2000, Viçosa. **Anais eletrônicos...** Viçosa: SBZ, 2000. CD-ROM.

PATTERSON, H. D.; THOMPSON, R. Recovery of inter-block information when block sizes are unuqual. **Biometrika**, London, v. 58, n. 3, p. 545-554, 1971.

RAFTERY, A. E.; LEWIS, S. M. How many iterations in the Gibbs sampler? **Bayesian Statistics**, Oxford: Oxford University Press, 1994. p. 763-774.

SORENSEN, D. **Gibbs sampling in quantitative genetics**. Denmark: Danish Institute of Animal,

Department of Breeding and Genetics, 1996. Intern Report, n. 82.

SORENSEN, D.; WANG, C. S.; JENSEN, J.; GIANOLA, D. Bayesian analysis of genetic change due to selection using Gibbs sampling. **Genetic Selection Evolution**, v. 26, p. 333-360, 1994.

STIGLER, S. M. **The history of statistics. The measurement of uncertainty before 1900**. Boston: Harvard University Press, 1986.

VAN TASSEL, C. P.; CASELLA, G.; POLLAK, E. J. Effects of selection on estimates of variance components using Gibbs sampling and restricted maximum likelihood. **Journal of Dairy Science**, v. 78, p. 678, 1995.

VAN TASSELL, C. P.; VAN VLECK, L. D. Multiple-trait Gibbs sampler for animal models: flexible programs for Bayesian and likelihood-based (co)variance component inference. **Journal Animal Science**, v. 74, p. 2586-2597, 1996.

WANG, C. S.; RUTLEDGE, J. J.; GIANOLA, D. Marginal inferences about variance components in a mixed linear model using Gibbs sampling. **Genetic Selection Evolution**, v. 25, p. 41, 1993.

Protocolado em: 10 maio 2006. Aceito em: 29 ago. 2006.