

REQUERIMENTOS PARA IMPLEMENTAÇÃO DA AMOSTRAGEM DE GIBBS NA ESTIMAÇÃO DE PARÂMETROS GENÉTICOS PARA PESO AO DESMAME NA RAÇA NELORE: INTERVALO DE UTILIZAÇÃO AMOSTRAL

Cláudio de U. Magnabosco¹, Carina U. de Faria², Arcadio de los Reyes³, Raysildo B. Lôbo⁴, Vanessa Barbosa⁵, Roberto Sainz⁶

¹ Pesquisador da Embrapa Cerrados/Arroz e Feijão, Bolsista do CNPq, Caixa Postal 08223, Planaltina, DF,

² Mestranda da Universidade Federal de Goiás, Bolsista da Embrapa Cerrados, Caixa Postal 08223, Planaltina, DF,

³ Professor Titular do Departamento de Produção Animal da EV/UFG, Caixa Postal, 131, Goiânia, GO,

⁴ Professor Associado, FMRP-USP, Ribeirão Preto, SP,

⁵ Zootecnista, Bolsista CNPq, Embrapa Cerrados, Caixa Postal 08223, Planaltina, DF,

⁶ Professor da University of Califórnia, Davis/EUA.

INTRODUÇÃO

Os parâmetros genéticos são obtidos com base em estimativas de componentes de (co)variância, que devem ser muito bem estimados. Dessa maneira, os métodos para estimação desses componentes têm sofrido um constante aprimoramento ao longo dos anos, em especial, na última década.

Recentemente, o uso de métodos bayesianos surgiu como uma opção para a solução de problemas relacionados à avaliação do mérito genético em populações animais. Conforme descrito por MAGNABOSCO et al. (2000), o algoritmo de Amostragem de Gibbs (GS) é um método capaz de gerar amostras aleatórias de estimativas de parâmetros, a partir de um conjunto de observações, obtendo-se dessa forma as denominadas distribuições posteriores dos parâmetros, dado um conjunto de observações, que é proporcional ao produto da probabilidade dos parâmetros e aquela das observações. A amostragem de Gibbs por ser um método de Monte Carlo requer intensas integrações numéricas, no entanto, é de fácil implementação, principalmente se comparado a algoritmos baseados na Máxima Verossimilhança Restrita.

MATERIAL E MÉTODOS

Para a implementação da Amostragem de Gibbs, considerou-se um conjunto de dados obtidos de arquivos do Controle de Desenvolvimento Ponderal da Associação Brasileira de Criadores de Zebu - ABCZ e foram analisados 29.245 registros de peso aos 205 dias de idade (P205) de animais da raça Nelore criados a pasto, nascidos no período de 1980 a 1993, oriundos de rebanhos localizados nas regiões Sudeste e Centro-Oeste do Brasil. Para as análises utilizou-se o modelo linear misto seguinte:

$$Y = Xb + Z_1a + Z_2m + Z_3p + e$$

onde y é o vetor da variável dependente (P205), b o vetor dos efeitos fixos, X a matriz de incidência que associa b com y , a representa o vetor dos efeitos genéticos aditivos diretos, Z_1 a matriz de incidência que associa a com y , m o vetor das contribuições maternas (efeito genético aditivo maternal), Z_2 a matriz de incidência que associa m com y , p o vetor das contribuições de ambiente permanente, Z_3 a matriz de incidência que associa p com y e finalmente e o vetor dos efeitos residuais aleatórios.

Nesse estudo foi assumido que b é um vetor $k \times 1$ dos efeitos fixos, o qual, do ponto de vista bayesiano, é na realidade um vetor de efeitos aleatórios, com valores iniciais de distribuição que representam nenhum conhecimento prévio sobre os valores verdadeiros. Considera-se X como sendo uma matriz de posto completo, Z_1 , Z_2 e Z_3 , são as conhecidas matrizes de incidência que relacionam o vetor de dados y com os efeitos diretos, maternas e de ambiente permanente incluídos no modelo. A distribuição condicional do vetor de dados y , dados os parâmetros de localização e os de dispersão, é apresentada como tendo uma distribuição normal multivariada.

$$y | b, a, m, p, s^2 \sim N [Xb + Z_1a + Z_2m + Z_3p, I_s s^2]$$

onde, s^2 é a variância dos efeitos residuais e para os outros efeitos aleatórios incluídos.

Na implementação do algoritmo de Gibbs foram assumidos três diferentes intervalos de utilização amostral ($d = 30$, $d = 150$, $d = 750$) em esquema de cadeia longa, com o intuito de avaliar as densidades posteriores obtidas para os componentes de (co)variância e parâmetros

OBJETIVOS

Determinar o intervalo de utilização amostral mais adequado para a obtenção de distribuições marginais posteriores dos parâmetros genéticos para peso aos 205 dias de idade em bovinos da raça Nelore.

RESULTADOS

Para essas análises usando o algoritmo de Amostragem de Gibbs (GS) foi utilizado um esquema chamado de cadeia longa. Depois de um período de descarte amostral de $k = 20.000$ ciclos, um total de $m = 2.667$, 534 e 107 amostras dos parâmetros do modelo linear foram retiradas com intervalos de utilização amostral $d = 30$, 150, 750 ciclos, respectivamente, com o tamanho de cadeia amostral de 100.000 ciclos e considerando $v = 9$. O aplicativo MTGSAM (Multiple Trait using Gibbs Sampling under Animal Model) foi usado para calcular as médias posteriores dos componentes de (co)variância obtidas a partir das amostras da cadeia de Gibbs, bem como, para a estimação de correlações e herdabilidades direta e maternal. Foi verificado que o procedimento de se adotar diferentes intervalos de utilização amostral, pouco ou nada influenciou as estimativas de médias, modas e medianas obtidas das distribuições marginais posteriores dos parâmetros genéticos (Tabela 1). No entanto, as densidades marginais posteriores representaram-se graficamente distintas para os diferentes intervalos de utilização amostral estudados (Figura 1). Amostras obtidas de intervalos abaixo de 30 ciclos podem produzir diferenças nas estimativas devido à correlação serial entre estas amostras, o que causa uma dependência indesejável entre essas estimativas obtidas do processo de reamostragem de Gibbs (MAGNABOSCO et al., 2000; JENSEN et al., 1994). Foi observado que a utilização de intervalos de utilização amostral muito grandes pode levar à obtenção de densidades marginais posteriores apresentando uma distribuição com tendência à forma bimodal, i.e., existe uma dissociação entre as amostras geradas pelo método GS, conforme observado no intervalo de utilização amostral de 750 ciclos. Esses resultados são apresentados na Figura 1. Neste estudo foi considerada uma distribuição inicial (a priori) que representava certo grau de conhecimento dos valores iniciais dos dados ($v = 9$), e desta forma, não houve necessidade de intervalos de utilização amostral maiores que 30 ciclos, que gerou distribuições posteriores estáveis e com uma distribuição normal. Na Tabela 1 pode ser observado que as medidas de tendência central estimadas suportam essa afirmação. Os resultados obtidos por LIN e BERGER (2001) concordam com os desse estudo, pois os autores relatam que quando existe informação suficiente sobre os dados o intervalo de utilização amostral pode ser representado por um pequeno número de ciclos no processo de amostragem. SEWALEM e JOHANSSON (2000), em estudo recente, recomendaram um intervalo de utilização amostral de 10 ciclos em análises com informação suficiente sobre a estrutura dos dados de campo.

Tabela 1. Estimativas das distribuições marginais posteriores dos parâmetros genéticos para peso aos 205 dias de idade (P205) considerando intervalos de utilização amostral de 30, 150, 750 ciclos e $v = 9$.

Intervalo Amostral	Parâmetros	Média	Moda	Mediana	Variância	Desvio Padrão	IC
$d = 30$ 2.667 amostras	h^2_d	0,27	0,27	0,27	0,0009	0,03	0,21 a 0,33
	h^2_m	0,19	0,18	0,19	0,0003	0,02	0,15 a 0,23
	r_{dm}	- 0,33	- 0,34	- 0,33	0,0043	0,07	- 0,47 a - 0,19
$d = 150$ 534 amostras	h^2_d	0,27	0,27	0,27	0,0009	0,03	0,21 a 0,33
	h^2_m	0,19	0,19	0,19	0,0003	0,02	0,15 a 0,23
	r_{dm}	- 0,33	- 0,37	- 0,34	0,0043	0,07	- 0,47 a - 0,19
$d = 750$ 107 amostras	h^2_d	0,28	0,27	0,27	0,0010	0,03	0,22 a 0,34
	h^2_m	0,19	0,20	0,19	0,0003	0,02	0,15 a 0,23
	r_{dm}	- 0,32	- 0,34	- 0,33	0,0047	0,06	- 0,46 a - 0,18

h^2_d , h^2_m , r_{dm} são as herdabilidades direta e maternal e a correlação genética entre efeitos direto e maternal respectivamente. IC é o Intervalo de confiança de 95%.

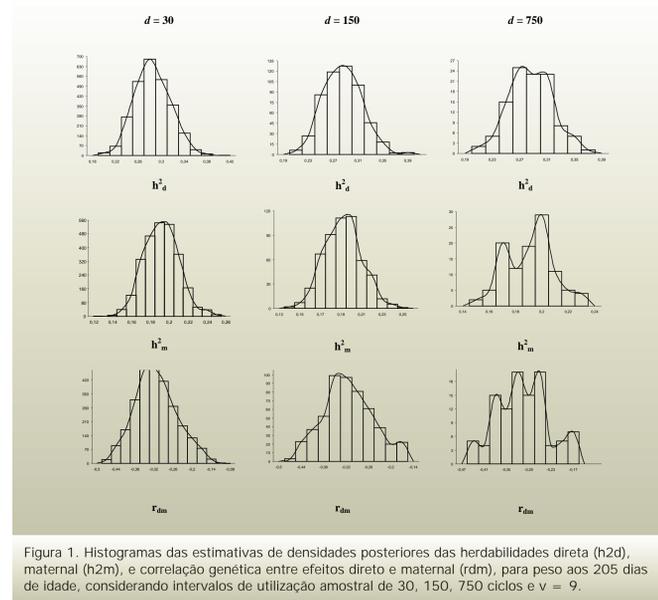


Figura 1. Histogramas das estimativas de densidades posteriores das herdabilidades direta (h^2_d), maternal (h^2_m), e correlação genética entre efeitos direto e maternal (r_{dm}), para peso aos 205 dias de idade, considerando intervalos de utilização amostral de 30, 150, 750 ciclos e $v = 9$.

CONCLUSÕES

O intervalo de utilização amostral é de fundamental importância para que as amostras das estimativas de parâmetros genéticos, geradas no processo de reamostragem de Gibbs, não apresentem correlação serial, i.e., dependência. O intervalo de utilização amostral que melhor representou a distribuição das densidades posteriores foi de 30 ciclos, considerando dados de campo com boa estrutura.

REQUERIMENTOS PARA IMPLEMENTAÇÃO DA AMOSTRAGEM DE GIBBS NA
ESTIMAÇÃO DE PARÂMETROS GENÉTICOS PARA PESO
AO DESMAME NA RAÇA NELORE: INTERVALO DE UTILIZAÇÃO AMOSTRAL



MINISTÉRIO DA AGRICULTURA,
PECUÁRIA E ABASTECIMENTO



*Empresa Brasileira de Pesquisa Agropecuária
Embrapa Cerrados
Ministério da Agricultura, Pecuária e do Abastecimento
BR 020, km 18, Rodovia Brasília/Fortaleza, Planaltina, DF
Telefone: (61) 388- 9898 Fax: (61) 388- 9879*

MINISTÉRIO DA AGRICULTURA,
PECUÁRIA E ABASTECIMENTO

