



## APLICAÇÃO DE MÉTODOS REML E BAYESIANO VIA AMOSTRADOR DE GIBBS NA ESTIMAÇÃO DE COMPONENTES DE VARIÂNCIA PARA PRODUÇÃO DE LEITE NO ESTADO DO PARANÁ

ALENCARIANO JOSÉ DA SILVA FALCÃO<sup>1</sup>, ELIAS NUNES MARTINS<sup>2</sup>, CLÁUDIO NAPOLIS COSTA<sup>3</sup>,  
EDUARDO SHIGUERO SAKAGUTI<sup>4</sup>, JOSMAR MAZUCHELI<sup>5</sup>

<sup>1</sup> Prof. Adjunto, PUCPR, campus Toledo. Av. da União, 500. CEP: 85902-532. [falcao@rla01.purpr.br](mailto:falcao@rla01.purpr.br)

<sup>2</sup> Prof. Associado, Universidade Estadual de Maringá, DZO. Av Colombo 5790. 87020-900. [enmartins@uem.br](mailto:enmartins@uem.br)

<sup>3</sup> Pesquisador da EMBRAPA Gado de Leite – 36038-330 Juiz de Fora, MG. Bolsista CNPq. [cnc8@cnpgl.embrapa.br](mailto:cnc8@cnpgl.embrapa.br)

<sup>4</sup> Prof. Adjunto, Universidade Estadual de Maringá, DZO. [essakaguti@uem.br](mailto:essakaguti@uem.br)

<sup>5</sup> Professor Adjunto do Departamento de Matemática e Estatística da Universidade Estadual

**RESUMO** – Foram utilizados 40.727 registros de lactação de vacas da raça Holandesa, que pariram entre 1981 e 1992, distribuídas em 322 rebanhos, para estimação de componentes de variância pelos métodos REML e Bayesiano via amostragem de Gibbs. Utilizou-se um modelo animal contendo os efeitos de rebanho-ano, grupo genético, ordem de parto, os efeitos genéticos diretos e de ambiente permanente. Não houve evidências para falta de convergência do amostrador de Gibbs. As estimativas e desvios padrão REML e GS de herdabilidade foram  $0,26 \pm 0,001$  e  $0,28 \pm 0,014$ . O amostrador de Gibbs mostrou-se mais vantajoso, pois possibilitou estimações pontuais e intervalares dos parâmetros.

**PALAVRAS-CHAVE:** cadeias de Markov, bovinos, herdabilidade, REML

APPLYING RESTRICTED MAXIMUM LIKELIHOOD AND BAYESIAN METHODS TO ESTIMATE VARIANCE COMPONENTS FOR MILK YIELD IN PARANÁ STATE

**ABSTRACT** – Variance components were estimated by REML and Bayesian via Gibbs sampling, using 40,727 lactations of Holstein cows, calving from 1981 to 1992 in 322 herds. An animal model was applied and included the effects of herd-year, parity, genetics group, effects of additive genetic and permanent environment. There was no evidence of lack of convergence of the GS. Estimates and standard deviations of heritability using REML and GS were  $0.26 \pm 0.001$  e  $0.28 \pm 0.014$ . Gibbs sampler allows calculation point estimates and confidence interval, so it seemed more advantageous than REML to estimate variance components.

**KEYWORDS:** Dairy cattle, heritability, Markov chains, REML

### INTRODUÇÃO

Nas duas últimas décadas, o método de máxima verossimilhança restrita (REML) tornou-se o método mais utilizado na estimação de componentes de variância no melhoramento animal (Smith & Graser, 1986; Costa, 1999).

Segundo Searle (1979), do ponto de vista freqüentista é impossível derivar a distribuição amostral exata dos estimadores dos componentes de variância, exceto nos casos triviais e, assim, a melhor alternativa para contornar o problema é recorrer a resultados assintóticos. Por outro lado, a abordagem Bayesiana é capaz de produzir com exatidão as distribuições conjunta e marginal posteriores para qualquer tamanho de amostra (Zellner, 1971; Box & Tiao, 1973), as quais dão uma descrição completa do estado de incerteza *a posteriori* dos dados. No melhoramento animal, o paradigma Bayesiano e os métodos de Monte Carlo baseados em cadeias de Markov (MCMC) têm propiciado novas perspectivas a questões relacionadas à estimação de componentes de variância e parâmetros genéticos (Gianola & Fernando, 1986; 1994; Van Tassell et al., 1995; Martins et al., 2002). No entanto, para aplicação do método MCMC em problemas práticos, é necessário saber não só se o algoritmo convergirá, mas também saber quando. A lentidão na convergência dos algoritmos de Markov, devido à correlação entre os valores amostrais gerados, constitui outro desafio ainda maior na implementação do método (Gameran, 1996). Este trabalho teve como objetivos, a) estimar os componentes de variância de produção de leite, utilizando-se o método REML e procedimentos Bayesianos e b) comparar a efetividade dos dois métodos.

**MATERIAL E MÉTODOS**

Os dados utilizados foram gerados pelo Serviço de Controle Leiteiro da Associação Paranaense de Criadores de Bovinos da Raça Holandesa, e cedidos pela EMBRAPA – Gado de Leite. O conjunto de dados original foi editado com uso do sistema SAS® (SAS, 2001). Foram eliminadas as classes de rebanho-ano com menos de 15 observações, lactações inferiores a 150 e superiores a 450 dias de duração e registros de produção menores que 3050 e maiores que 13.000 kg de leite. Após todas as restrições, obteve-se um conjunto com 40.727 registros de produção de leite de 32.260 vacas, que pariram entre 1981 e 1992, distribuídas em 322 rebanhos.

As análises REML e Bayesianas via amostragem de Gibbs (GS) foram realizadas segundo um modelo animal [1] que incluiu os efeitos fixos de rebanho-ano, época de parição, composição racial das vacas (puras de origem ou superior a 31/32 da raça Holandesa), idade da vaca como covariável (efeitos linear e quadrático), e os efeitos genéticos aditivos, de ambiente permanente e residual como aleatórios. Modelo em notação matricial,  $y = X\beta + Za + Wp + e$  [1], onde  $y$  é o vetor das produções de leite ajustadas para 305 dias de lactação (PL305);  $\beta$ ,  $a$ ,  $p$  e  $e$  são os vetores de efeitos fixos, genéticos diretos, de ambiente permanente e de resíduos.  $X$ ,  $Z$  e  $W$  são matrizes de incidência dos efeitos fixos, genéticos diretos e de ambiente. Admitiu-se que  $y$ ,  $a$ ,  $p$  e  $e$  possuem distribuição conjunta normal multivariada, como segue,

$$\begin{bmatrix} y \\ a \\ p \\ e \end{bmatrix} \sim NMV \left\{ \begin{bmatrix} X\beta \\ \phi \\ \phi \\ \phi \end{bmatrix}, \begin{bmatrix} ZGZ' + WPW' + R & ZG & WP & R \\ GZ' & G & \phi & \phi \\ PW' & \phi & P & \phi \\ R & \phi & \phi & R \end{bmatrix} \right\}, \text{ em que, } G = A \sigma_a^2, P = I_p \sigma_p^2, e, R = I \sigma_e^2,$$

onde,  $A$  é a matriz de coeficientes de parentesco;  $I_p$  uma matriz identidade;  $\sigma_a^2$  a variância genética aditiva;  $\sigma_p^2$  a variância de efeito permanente de ambiente e  $\sigma_e^2$  a variância residual.

As estimativas REML dos componentes de variância (CV) foram obtidas por meio do programa MTDFREML, descrito por Boldman et al. (1995). Densidades *a priori* e pressuposições na análise Bayesiana. Para os efeitos fixos  $f(\beta) \propto cte$ . Os efeitos aleatórios pressupôs-se normalmente distribuídos:  $a | \sigma_a^2, A \sim N(0, A \sigma_a^2)$ ;  $p | \sigma_p^2 \sim N(0, I_p \sigma_p^2)$ ,  $e | \sigma_e^2 \sim N(0, I_e \sigma_e^2)$ , e os erros normais e independentes,  $y | \beta, a, p, \sigma_e^2 \sim MVN [X\beta + Za + Wp, I_n \sigma_e^2]$ . Distribuições gama invertida foram usadas para os componentes de variância, ou seja,  $\sigma_a^2 \sim \Pi(\sigma_a^2, \nu_a)$ , tal que  $f(\sigma_a^2 | \sigma_a^2, \nu_a) \propto (\sigma_a^2)^{-\frac{1}{2}(\nu_a+2)} \cdot e^{-\frac{1}{2}(\frac{\sigma_a^2}{\sigma_a^2})}$ ;  $\sigma_p^2 \sim \Pi(\sigma_p^2, \nu_p)$ , tal que  $f(\sigma_p^2 | \sigma_p^2, \nu_p) \propto (\sigma_p^2)^{-\frac{1}{2}(\nu_p+2)} \cdot e^{-\frac{1}{2}(\frac{\sigma_p^2}{\sigma_p^2})}$  e  $\sigma_e^2 \sim IX^{-2}(s_o, \nu_e)$ , tal que  $f(\sigma_e^2 | s_o, \nu_e) \propto (\sigma_e^2)^{-\frac{1}{2}(\nu_e+2)} \cdot e^{-\frac{1}{2}(s_o/\sigma_e^2)}$ . Onde  $\nu_a$ ,  $\nu_p$  e  $\nu_e$  são os graus de liberdade, e  $\sigma_a^2$ ,  $\sigma_p^2$  e  $s_o$  são os parâmetros escala das respectivas distribuições. A função densidade conjunta *a posteriori*, resultou em,

$$f(\beta, a, p, \sigma_a^2, \sigma_p^2, \sigma_e^2 | y, \nu_a, \sigma_a^2, \nu_p, \sigma_p^2, \nu_e, s_o) \propto (\sigma_a^2)^{-\frac{1}{2}(n+\nu_a+2)} \cdot \exp\left\{-\frac{1}{2}\left[a'(A^{-1}\sigma_a^{-2})a + (\nu_a - 2)(\sigma_a^2 \sigma_a^{-2})\right]\right\} \\ \cdot (\sigma_p^2)^{-\frac{1}{2}(n_i+\nu_p+2)} \cdot \exp\left\{-\frac{1}{2}\left[p'(I_{n_i}\sigma_p^{-2})p + (\nu_p - 2)(\sigma_p^2 \sigma_p^{-2})\right]\right\} (\sigma_e^2)^{-1/2(n_l+\nu_e+2)} \\ \cdot \exp\left[-\frac{1}{2}(\nu_e - 2)(s_o/\sigma_e^2)\right] \exp\left\{-\frac{1}{2}(y - X\beta - Za - Wp)' \sigma_e^{-2} (y - X\beta - Za - Wp)\right\}$$

onde  $n$ ,  $n_i$  e  $n_l$  são o número de animais, número de níveis dos efeitos de ambiente permanente e de lactações, respectivamente.

O programa MTGSAM, descrito por Van Tassell & Van Vleck (1995), foi utilizado para obtenção das cadeias de Markov. Cadeias de 250.000 iterações foram geradas com descarte inicial de 50.000 iterações e intervalo de utilização amostral de 100 iterações. A monitoração de convergência foi realizada pelo método de Heidelberg & Welch (1983).

**RESULTADOS E DISCUSSÃO**

A média observada e desvio padrão para PL305 foi 6.327±1501 kg. As estimativas REML e Bayesianas dos componentes de variância e de herdabilidade são apresentadas na Tabela 1. Os valores das herdabilidades para PL305 foram inferiores aos reportados por Ribas et al. (1993) e Rorato et al. (1994), porém próximo aos obtidos por Matos et al. (1996) e Costa (1999).

TABELA 1. Estimativas dos componentes de variância<sup>1</sup> e de herdabilidade para produção de leite no Estado do Paraná, de acordo com o método de estimação

Método	$\sigma_a^2$	$\sigma_{pe}^2$	$\sigma_e^2$	$\sigma_p^2$	$h_a^2$
Bayesiano <sup>2</sup>	355,1 ± 20,2	257,2 ± 16,4	663,3 ± 7,5	1276 ± 11,0	0,28 ± 0,014
REML	331,6	268,2	664,8	1265	0,26 ± 0,001

1 valores divididos por 1000; 2médias a posteriori e erros padrão;  $\sigma_a^2$  = variância genética,  $\sigma_{pe}^2$  = variância de ambiente permanente,  $\sigma_e^2$  = variância residual,  $\sigma_f^2$  = variância fenotípica e  $h_a^2$  = herdabilidade.

A estimativa Bayesiana da variância genética foi significativamente superior àquela obtida pelo REML. Enquanto a variância residual tendeu a ser menor para o estimador baseado no GS. Van Tassell et al. (1995), usando dados simulados para estimação de CV, observaram que as estimativas Bayesianas via GS do quadrado médio do resíduo tenderam a ser inferiores às obtidas pelos métodos REML e MIVQUE, principalmente, para dados simulados com baixa herdabilidade. A diferença deveu-se, em parte, ao impacto da *priori* usada no GS.

Os procedimentos Bayesianos permitem, além das estimativas pontuais, determinar intervalos de credibilidade (IC) para a distribuição *a posteriori* do componente de variância, sem aproximações ou o uso de pressuposições de normalidade, representando, assim, uma vantagem sobre os métodos de freqüentista, como o REML. Os IC dos componentes de variância e de herdabilidade para PL305 são mostrados nas Tabelas 2. As amplitudes dos IC e a magnitude dos desvios padrão para os componentes de variância e herdabilidade permitem deduzir que estes parâmetros foram estimados com bastante precisão.

TABELA 2. Percentis dos componentes de variância e da herdabilidade para produção de leite

Parâmetro	Percentis		
	2,5%	50%	97,5%
Variância genética	316.709	354.809	395.653
Variância de ambiente permanente	224.875	257.212	287.968
Variância residual	649.057	663.205	678.399
Herdabilidade	0,250	0,278	0,308

O amostrador de Gibbs parece ter visitado os efetivos parâmetros de variância genética e residual ao acaso, como se pode observar na Figura 1, que mostra os históricos de convergência do algoritmo. O método de Heidelberg & Welch (1983) indicou que todas as cadeias convergiram para os parâmetros de interesse.

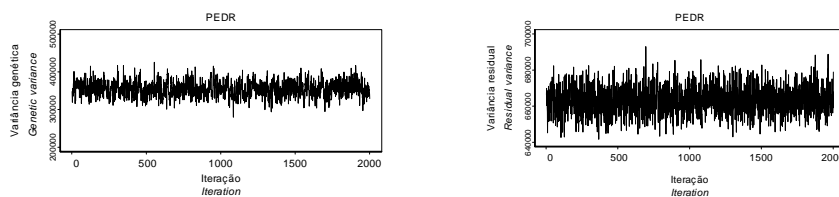


FIGURA 1. Históricos das iterações das variâncias genética e residual para a produção de leite

## CONCLUSÕES

As estimativas Bayesianas da variância genética e de herdabilidade foram ligeiramente superiores àquelas obtidas pelo método REML. A inferência Bayesiana se mostrou mais vantajosa, pois possibilitou estimações pontuais e intervalares dos parâmetros, sem aproximações e uso de pressuposição de normalidade, e uma exploração mais adequada da estrutura de dados.

## REFERÊNCIAS BIBLIOGRÁFICAS

BOLDMAN, K.G.; KRIESE, L.A.; VAN VLECK, L.D. et al. 1995. **A Manual for use of MTDFREML. A set of programs to obtain estimates of variance and covariances** [DRAFT]. USA.



- BOX, G.E.P.; TIAO, G.C.. **Bayesian inference in statistical analysis**. Reading: Addison-Wesley Publishing Co., 1973. 588p.
- COSTA, C.N. An investigation into heterogeneity of variance for milk and fat yields of Holstein cows in Brazilian herd environments. **Genetics and Molecular Biology**, v.22, n.3, p.375-381, 1999.
- GAMERMAN, D. **Simulação estocástica via cadeias de Markov**. Universidade Federal do Rio de Janeiro, 1996. 196p.
- GIANOLA, D.; FERNANDO, R.L. Bayesian methods in animal breeding theory. **Journal of Animal Science**, v.63, p.217-277, 1986.
- HEIDELBERGER, P.; WELCH, P.D. Simulation run length control in the presence of an initial transient. **Operations Research**, v.31, p.1109-1144, 1983.
- MARTINS, E. N.; ENGLER, E. O.; SAKAGUTI, E. S. et al. Bayesian analyses of heterogeneous variance among generations of Brazilian Canchim beef calves. In: 7th World Congress on Genetics Applied to Livestock Production, August 19-23. **Proceeding...**Montpellier, France:2002.
- MATOS, R. S.; RORATO, P. R. N.; FERREIRA, G. B. et al. Parâmetros genéticos para produção de leite e gordura da raça Holandesa no Estado do Rio Grande do Sul. In: REUNIÃO ANUAL DA SOCIEDADE BRASILEIRA DE ZOOTECNIA, 33, 1996, Fortaleza. **Anais...**, Fortaleza, Ceará. 1996. p.86-87.
- RORATO, P. R. N.; LÔBO, R. B.; MARTINS FILHO, R. et al. Efeito da interação genótipo-ambiente sobre a produção de leite da raça Holandesa, no Estado do Paraná. **Revista Brasileira de Zootecnia**, v.23, n.5, p.859-869. 1994.
- SAS INSTITUTE. 2000. **SAS/STAT User's guide: statistics**, versão 8.01, v.2,4. ed. Cary.
- VAN TASSELL, C. P., VAN VLECK, D. L. 1995. **A manual for use of MTGSAM. A set of FORTRAN programs to apply Gibbs sampling to animal models for variance component estimation [DRAFT]**. U.S. Department of Agriculture, Agricultural Research Service.
- VAN TASSELL, C. P.; CASELLA, G.; POLLAK, E. J. Effects on selection on estimates of variance components using Gibbs sampling and restricted maximum likelihood. **Journal of Dairy Science**, v.78, p.678-692, 1995.
- ZELLNER, A. **An introduction to Bayesian inference in econometrics**. New York: J. Willey and Sons, 1971. 431p.
- RIBAS, N. P.; RORATO, P.R.N.; LÔBO, R.B. Estimativas de parâmetros genéticos para as características de produção da raça Holandesa no Estado do Paraná. **Revista Brasileira de Zootecnia**, v.22, n.4, p.634. 1993.
- SEARLE, S. **Notes on variance component estimation. A detailed account of maximum likelihood and kindred methodology**. Biometrics Unit, Cornell University, Ithaca, NY, 1979. (Paper BU-673-M).