

Uso da Iteração nos Dados para Resolução de Equações de Modelo Misto

Paulo Sávio Lopes¹

RESUMO - Valores genéticos foram preditos usando-se a iteração nos dados em um modelo animal reduzido. O peso corporal, aos 35 dias de idade, de duas linhas de aves de corte foi medido em duas gerações, cujo arquivo de dados consistia de 1044 pais (população-base), 829 progênes-pais e 9039 progênes-não-pais. O número de iterações e o tempo de processamento, para obtenção das soluções das equações de modelo misto via iteração nos dados, foram avaliados por sete critérios de convergência (10^1 a 10^{-5}). Seis iterações foram necessárias, com 9 s de tempo, e 158 iterações, com 5 min e 38 s, para se obterem as soluções para os critérios de convergência de 10^1 e 10^{-5} , respectivamente. As correlações entre os valores genéticos preditos foram perfeitas ($r=1,00$), e não houve diferenças entre as tendências genéticas estimadas pelos critérios de convergência de 10^{-1} a 10^{-5} . Com base nestes resultados, conclui-se que a iteração nos dados pode ser eficientemente usada em microcomputadores, na avaliação genética animal, sem grande demanda de tempo e memória computacional.

Palavras-chave: avaliação genética, iteração nos dados, modelos mistos

Use of Iteration on Data to Solve the Mixed Model Equations

ABSTRACT - Breeding values were predicted by iterating on data using reduced animal model. Body weight at 35 days of age from two lines of meat-type chickens was measured in two generations and data file was consisted on 1044 parents (base population), 829 progeny parents and 9039 progeny non-parents. The number of iterative rounds and processing time for mixed model equations solutions via iterating on data were evaluated using seven convergence criteria (10^1 to 10^{-5}). A total of six rounds of iteration and 9 s of time and 158 rounds and 5 min and 38 s were required to reach the solutions for 10^1 and 10^{-5} convergence criteria, respectively. The correlation between predicted breeding values were perfect ($r=1,00$) and there were no significant differences between estimated genetic trends using 10^{-1} to 10^{-5} as a convergence criteria. Based on these results, it was concluded that iterating on data could be efficiently used for animal genetic evaluation in microcomputer without great computational requirements.

Key Words: genetic evaluation, iteration on data, mixed model

Introdução

HENDERSON (1963, 1973, 1974) propôs a metodologia de modelos mistos, para obtenção da melhor predição linear não-viesada (BLUP), dos valores genéticos dos animais. A utilização do modelo animal que possibilita a inclusão de toda informação de pedigree na análise é uma das grandes vantagens dessa metodologia. Entretanto, existe o inconveniente de o número total de equações a serem resolvidas simultaneamente em grande volume de dados demandar grande esforço computacional, quanto à memória e ao tempo de processamento.

Com o objetivo de contornar esse problema computacional na resolução das equações, diversas técnicas têm sido propostas. QUAAS e POLLAK (1980) propuseram o uso do modelo animal reduzido, em que as equações referentes aos indivíduos que não tiverem filhos são absorvidas, de maneira implíci-

ta, na construção do sistema. Outra técnica, proposta posteriormente por SCHAEFFER e KENNEDY (1986), foi o uso da iteração nos dados, que consiste em obter as soluções, sem construir as equações de forma explícita, ou seja, sem armazená-las na memória.

Outras estratégias utilizadas são algoritmos eficientes de obtenção das soluções. De modo geral, os métodos iterativos são recomendados (LOPES et al., 1993; QUAAS e POLLAK, 1980).

Comparações têm sido feitas entre essas várias alternativas, que vão desde a escolha de métodos numéricos apropriados até a utilização de algoritmos eficientes para formação e, ou, resolução dos sistemas de equações. A opção por um método ou outro vai depender da disponibilidade de computadores com maior capacidade de memória e, ou, velocidade de processamento e da precisão desejada nas soluções.

BLAIR e POLLAK (1984) verificaram que o uso do modelo animal reduzido em relação ao modelo

¹Professor da UFV e bolsista do CNPq. E.mail: plopes@mail.ufv.br

animal diminuiu em 40% o tamanho do sistema de equações; com isso, foram necessários 40% do número de iterações, sendo o tempo por iteração 20% menor.

SCHAEFFER e KENNEDY (1986) verificaram que, na utilização do modelo animal, por meio do método da iteração nos dados, gastou-se metade do tempo despendido no modelo animal reduzido, com sistema de equações armazenado na memória. Entretanto, esses autores não testaram o modelo animal reduzido com iteração dos dados, o que, certamente, reduziria ainda mais o tempo de processamento.

É importante salientar que, com a evolução na área computacional, pode haver mudança na eficiência de cada método. LOPES et al. (1994) enfatizaram que, quando se propõe apenas ordenar os animais para fins de seleção, não há necessidade de grande acurácia nas soluções, visto que o que interessa é a não-mudança na ordem de classificação dos animais. Nesse caso, não há necessidade de grande número de iterações. No entanto, quando se buscam soluções mais precisas, como é o caso de avaliação de tendências genéticas, deve-se avaliar a precisão das soluções (BLAIR e POLLAK, 1984). Nesse caso, além da memória computacional, deve-se avaliar o tempo de processamento.

O objetivo deste trabalho foi verificar o tempo de processamento e a memória computacional necessários para resolução de equações de modelo misto, utilizando-se o modelo animal reduzido e a iteração nos dados.

Material e Métodos

Utilizaram-se dados de aves de corte obtidos na Universidade Federal de Viçosa (UFV) e avaliou-se a característica peso corporal aos 35 dias de idade, medido em duas linhas, durante duas gerações. Na Tabela 1, é apresentado o número total de aves, pais (população-base), progênie-pais e progênie-não-pais.

Os dados foram analisados pelo seguinte modelo:

$$y_{ij} = GC_i + g_{ij} + e_{ij}$$

em que

y_{ij} = observação do animal j, do grupo contemporâneo i;

GC_i = efeito fixo do grupo contemporâneo i, com efeitos de linha, sexo, incubação e geração;

g_{ij} = valor genético do animal j, do grupo contemporâneo i; e

e_{ij} = erro aleatório associado a cada observação.

Utilizando-se o modelo animal reduzido (QUAAS

Tabela 1 - Número total de aves, pais, progênie-pais, progênie-não-pais e grupos contemporâneos
 Table 1 - Total number of chickens, parents, progeny-parents, progeny-non-parents and contemporary groups

Número total de aves <i>Total number of chickens</i>	10912
Número de pais (população-base) <i>Number of parents (base population)</i>	1044
Número de progênie-pais <i>Number of progeny-parents</i>	829
Número de progênie-não-pais <i>Number of progeny-non-parents</i>	9039
Número de grupos contemporâneos <i>Number of contemporary groups</i>	48

e POLLAK, 1980; MARTINS et al., 1997), descrito na forma matricial, tem-se

$$\begin{bmatrix} y_p \\ \tilde{y}_n \end{bmatrix} = \begin{bmatrix} X_p \\ X_n \end{bmatrix} \beta + \begin{bmatrix} Z_p \\ 1/2 Z_n P_{np} \end{bmatrix} g_p + \begin{bmatrix} e_p \\ Z_n \phi_n + e_n \end{bmatrix}$$

em que

y = vetor de observações das características

medidas nos animais, sendo p = pais e n = não-pais;

X = matriz de incidência de efeitos fixos;

β = vetor de efeitos fixos de grupo contemporâneo;

Z = matriz de incidência de valores genéticos;

g = vetor de valores genéticos;

P_{np} = matriz que relaciona pais com progênie;

ϕ = vetor de efeitos de amostra mendeliana; e

e = vetor de resíduos.

As soluções dos efeitos fixos e dos valores genéticos dos pais são obtidas pelo seguinte sistema de equações de modelo misto, após a absorção dos não-pais (MARTINS et al., 1997):

$$\begin{bmatrix} X' R_m^{-1} X & X' R_m^{-1} Z_m \\ Z_m' R_m^{-1} X & Z_m' R_m^{-1} Z_m + G_{pp}^{-1} \end{bmatrix} \begin{bmatrix} \beta^0 \\ \tilde{g}_p \end{bmatrix} = \begin{bmatrix} X' R_m^{-1} y \\ Z_m' R_m^{-1} y \end{bmatrix}$$

em que $R_m^{-1} = \begin{bmatrix} R_p^{-1} & \phi \\ \phi & R_{n^*}^{-1} \end{bmatrix}$,

sendo $R_{n^*} = Z_n D_n \otimes G_0 Z_n' + R_n$ e

$$Z_m = \begin{bmatrix} Z_p \\ 1/2 Z_n P_{np} \otimes I_0 \end{bmatrix}$$

As soluções dos não-pais são obtidas por (MARTINS et al., 1997):

$$\hat{g}_n = 1/2 P_{np} \otimes I_0 \hat{g}_p + D_n \otimes G_0 Z_n' R_{n^*}^{-1} \left(y_n - X_n \beta^0 - 1/2 Z_n P_{np} \otimes I_0 \hat{g}_p \right)$$

ou, ainda, por

$$\hat{g}_i = 1/2 \hat{g}_{si} + 1/2 \hat{g}_{di} + d_{ii} G_0 Z_{ii} R_{n^*}^{-1} \left(y_i - X_i \beta^0 - 1/2 \hat{g}_{si} - 1/2 \hat{g}_{di} \right)$$

em que

\hat{g}_i = vetor de valores genéticos preditos do indivíduo i ;

duo i ;

\hat{g}_{si} e \hat{g}_{di} = vetores de valores genéticos preditos do pai e da mãe de i ;

do pai e da mãe de i ;

d_{ii} = i -ésimo elemento da diagonal da matriz;

Z_{ii} = bloco diagonal da matriz Z , referente ao indivíduo i ; e

X_i = formada pelas linhas da matriz X , referentes ao indivíduo i .

Na iteração nos dados, que consiste em obter as soluções sem construir as equações de forma explícita, ou seja, sem necessidade de armazená-las na memória do computador, dividiram-se os dados em três categorias: pais (população-base), progênie-pais (indivíduos que são filhos e pais) e progênie-não-pais (filhos que não são pais) (QUAAS, 1995).

Esse método consiste em obter a solução para cada indivíduo, separadamente, a partir de três contribuições: dos pais, das progênie-pais e das progênie-não-pais. Essas contribuições foram apresentadas por QUAAS (1995) e são dadas a seguir.

1.a) Contribuição dos pais do indivíduo para a diagonal: $R_0^{-1} + x_1 G_0^{-1}$, em que G_0 e R_0 são matrizes de variâncias e covariâncias genética aditiva e residual e x recebe 2, se o animal tem dois pais; 4/3, se tem um pai; e 1, se não tem nenhum dos pais;

1.b) Contribuição dos pais para o lado direito (RHS):

$$R_0^{-1} \left(y_I - \hat{CG}_I \right) + x_I G_0^{-1} \left(1/2 \hat{s} + 1/2 \hat{d} \right);$$

em que y , \hat{CG}_I , \hat{s} e \hat{d} são, respectivamente,

observação do indivíduo, efeito estimado do grupo contemporâneo do indivíduo e valor genético predito do pai e da mãe do indivíduo.

2.a) Contribuição das progênie-pais na diagonal:

$$1/4 x_p G_0^{-1}.$$

2.b) Contribuição das progênie-pais para o lado

direito: $1/2 x_p G_0^{-1} \left(\hat{pg} - 1/2 \hat{pa} \right)$, em que \hat{pg} e \hat{pa} são,

respectivamente, valor genético predito da progênie-pai e do par acasalado.

3.a) Contribuição das progênie-não-pais na diagonal: $1/4 R_{n^*}^{-1}$, em que $R_{n^*} = d_{ii} G_0 + R_0$.

3.b) Contribuição das progênie pais para o lado direito:

$$1/2 R_{n^*}^{-1} \left(y - \hat{CG}_p - 1/2 \hat{pa} \right).$$

O valor genético predito de cada indivíduo é dado por:

$$\hat{a} = \left(\sum \text{Diagonal} \right)^{-1} \left(\sum \text{RHS} \right).$$

A solução de cada grupo contemporâneo é obtida a partir de duas contribuições, de pais e de não-pais (QUAAS, 1995), conforme a seguir.

1.a) Contribuição dos pais para a diagonal R_0^{-1} ,

1.b) Contribuição dos pais para o lado direito:

$$R_0^{-1} \left(y_I - \hat{CG}_I - \hat{I} \right),$$

em que \hat{I} é o valor genético predito do indivíduo.

2.a) Contribuição dos não-pais para a diagonal:

$$R_0^{-1},$$

2.b) Contribuição dos não-pais para o lado direito:

$$R_0^{-1} \left(y_I - \hat{CG}_I - 1/2 \hat{s} - 1/2 \hat{d} \right).$$

A solução de cada grupo contemporâneo é dada por:

$$\hat{CG}^{\text{atual}} = \hat{CG}^{\text{anterior}} + \left(\sum \text{Diagonal} \right)^{-1} \left(\sum \text{RHS} \right).$$

Para obtenção das soluções dos valores genéticos e de grupos contemporâneos utilizou-se o método iterativo de Gauss-Seidel (VARGA, 1962). Foi implementado um programa computacional em linguagem Fortran, em microcomputador Micro Pentium MMX, 200 μ Hz e 128 μ B RAM.

Avaliaram-se o número de iterações e o tempo de processamento, utilizando-se critérios de convergên-

cia de 10^1 a 10^{-5} (10 a 0,00001), dados pela diferença entre o resultado obtido numa iteração e na anterior. Obtiveram-se correlações entre os valores genéticos preditos dos animais e estimaram-se as tendências genéticas anuais por critério de convergência, a fim de determinar qual seria o mais adequado.

Resultados e Discussão

Na Tabela 2, apresentam-se o número de iterações e o tempo de processamento necessários para se obterem as soluções dos valores genéticos dos animais via iteração nos dados, utilizando-se critérios de convergência de 10^1 a 10^{-5} (10 a 0,00001). Na Tabela 3, apresentam-se as correlações entre os valores genéticos preditos dos animais, por critério de convergência.

Verifica-se que os tempos de processamento foram relativamente baixos, mesmo quando se trabalhou com critérios de convergência mais precisos, os quais variaram de 9 segundos, para o critério de 10^1 , a 3 minutos e 38 segundos, para o de 10^{-5} , tendo o número de iterações variado de 6 a 158.

Pelas correlações entre os valores genéticos preditos, observa-se que, a partir do critério de convergência de 10^{-1} , a correlação é igual à unidade, em relação aos critérios de convergência mais precisos

(10^{-2} a 10^{-5}). Esse valor corresponde, aproximadamente, ao limite mínimo de 0,099, utilizado na obtenção das soluções das equações de modelo misto, para ganho de peso em suínos, por SCHENKEL (1991).

LOPES et al. (1994 e 1995), ao utilizarem critério de convergência na obtenção da inversa da matriz ($Z'R^{-1}Z+G^{-1}$) das equações de modelo misto, dividindo-se cada elemento pela variância fenotípica da característica, consideraram o critério de convergência de 10^{-4} , para o qual foram necessárias 16 iterações na obtenção das soluções. Comparando-se os resultados deste trabalho com os desses autores, verifica-se que o critério de convergência de 10^0 , que precisou de 17 iterações, parece ser o equivalente.

Entretanto, segundo BLAIR e POLLAK (1984), quando se deseja avaliar as tendências genéticas, essa correlação entre valores genéticos ou entre ordem de classificação dos animais não seria recomendável para determinar critério de convergência. Por isso, foram estimadas as tendências genéticas anuais para cada um dos critérios utilizados, cujos resultados se encontram na Tabela 4, assim como as predições dos valores genéticos mínimo e máximo.

Verifica-se que, a partir de 10^{-3} , não houve mais mudanças nas tendências genéticas e nas predições dos valores genéticos mínimos e máximos.

Tabela 2 - Número de iterações e tempo de processamento necessários para se atingir o critério de convergência

Table 2 - Number of iteration rounds and processing time required to reach the convergence criterion

Critério de convergência <i>Convergence criterion</i>	Número de iterações <i>Rounds of iteration</i>	Tempo de processamento (s) <i>Processing time</i>
10^1	6	9
10^0	17	24
10^{-1}	45	62
10^{-2}	73	101
10^{-3}	102	141
10^{-4}	130	178
10^{-5}	158	218

Micro Pentium MMX, 200 μHz, 128 μB RAM.

Tabela 3 - Correlações entre os valores genéticos dos animais por critério de convergência

Table 3 - Correlation between predicted breeding values by convergence criterion

	10^0	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}
10^1	0,9994	0,9985	0,9984	0,9984	0,9984	0,9984
10^0		0,9998	0,9997	0,9997	0,9997	0,9997
10^{-1}			1,0000	1,0000	1,0000	1,0000
10^{-2}				1,0000	1,0000	1,0000
10^{-3}					1,0000	1,0000
10^{-4}						1,0000

Tabela 4 - Estimativas das tendências genéticas anuais e predições dos valores genéticos máximo e mínimo, por critério de convergência

Table 4 - Annual estimated genetic trends and maximum and minimum predicted breeding values by convergence criterion

Critério de convergência <i>Convergence criterion</i>	Tendência genética <i>Genetic trend</i> (g)	Valor genético máximo <i>Maximum breeding value</i> (g)	Valor genético mínimo <i>Minimum breeding value</i> (g)
10 ¹	21,24	203,53	-256,95
10 ⁰	24,12	211,29	-248,93
10 ⁻¹	25,70	216,72	-244,48
10 ⁻²	25,85	217,27	-244,15
10 ⁻³	25,87	217,33	-244,13
10 ⁻⁴	25,87	217,34	-244,13
10 ⁻⁵	25,87	217,34	-244,13

Verifica-se que, ao se utilizar o critério de 10⁰, semelhante ao proposto por LOPES et al. (1994 e 1995), a tendência genética estaria sendo subestimada em 6,76%, enquanto, ao se utilizar o de 10⁻¹, de acordo com a correlação entre os valores genéticos da Tabela 3 e os utilizados por SCHENKEL (1991), a tendência genética estaria sendo subestimada em apenas 0,66%.

Tendo em vista que valor de 10⁻¹, para o qual foram necessárias 45 iterações em apenas 1 minuto e 2 segundos, foi eficiente para se avaliarem 10.912 aves, das quais 1873 eram pais, conclui-se que essa metodologia pode ser usada em microcomputadores para avaliação genética animal, sem grande demanda de tempo e memória computacional.

Conclusões

A iteração nos dados pode ser eficientemente usada em microcomputadores, na avaliação genética animal, sem grande demanda de tempo e memória computacional.

Referências Bibliográficas

- BLAIR, H.T., POLLAK, E.J. 1984. Comparison of an animal model and an equivalent reduced animal model for computational efficiency using mixed model methodology. *J. Anim. Sci.*, 58(5):1090-1096.
- HENDERSON, C.R. 1963. Selection index and genetic expected advance. In: HANSON, W.D., ROBISON, H.F. (Eds.). *Statistical genetics and plant breeding*. Washington: NAS-NRC. p.141-163. (Publication, 982).
- HENDERSON, C.R. Sire evaluation and genetic trends. In: ANIMAL BREEDING AND GENETICS SYMPOSIUM IN HONOR OF DR. JAY L. LUSH, 1973, Blacksburg. *Proceedings...* Champaign: ASAS/ASDA, 1973. p.10-41.
- HENDERSON, C.R. 1974. General flexibility of linear model techniques for sire evaluation. *J. Dairy Sci.*, 57(8):963-972.

- LOPES, P.S., MARTINS, E.N., SILVA, M.A. et al. 1993. *Métodos de resolução de sistemas de equações lineares*. Viçosa, MG: Editora UFV. 55p.
- LOPES, P.S., SILVA, M.A., REGAZZI, A.J. et al. 1994. Critérios de convergência para resolução de equações de modelo misto. *R. Soc. Bras. Zootec.*, 23(6):1008-1020.
- LOPES, P.S., SILVA, M.A., REGAZZI, A.J. et al. 1995. Avaliação de métodos iterativos de resolução de equações de modelo misto. *R. Soc. Bras. Zootec.*, 24(3):461-472.
- MARTINS, E.N., LOPES, P.S., SILVA, M.A. et al. 1997. *Uso de modelos mistos na avaliação genética animal*. Viçosa, MG: Editora UFV. 121p.
- QUAAS, R.L. 1995. *Iterating on data*. Cornell University, Ithaca, NY. 9p.
- QUAAS, R.L., POLLAK, E.J. 1980. Mixed model methodology for farm and ranch beef cattle testing programs. *J. Anim. Sci.*, 51(6):1277-1287.
- SCHAEFFER, L.R., KENNEDY, B.W. Computing solutions to mixed model equations. In: WORLD CONGRESS ON GENETICS APPLIED TO LIVESTOCK PRODUCTION, 3, 1986, Lincoln. *Proceedings...* Lincoln, 1986, v.12, p.382-393.
- SCHENKEL, F.S. *Aplicação da metodologia dos modelos mistos na avaliação genética de suínos*. Porto Alegre, RS: UFRGS, 1991. 218p. Dissertação (Mestrado em Zootecnia) - Universidade Federal do Rio Grande do Sul, 1991.
- VARGA, R.S. 1962. *Matrix iterative analysis*. Englewood Cliffs: Prentice-Hall. 322p.

Recebido em: 08/03/99

Aceito em: 10/09/99