

SIMULAÇÃO DE CARACTERÍSTICAS CORRELACIONADAS

Michele Porto Pires, Adam Tati Harth Utsunomiya, Rafael Keith Ono, Ricardo da Fonseca –
UNESP/DRACENA – Faculdade de Zootecnia - Laboratório de Computação Científica Aplicada à Zootecnia
(LuCCA-Z)

A simulação de dados desempenha um importante papel nas pesquisas científicas, particularmente no melhoramento genético animal, minimizando os gastos na montagem do experimento.

Marby e See (1990), Analla et al.(1995), Carneiro et al.(2002) e Fonseca(2003) simularam dados para a avaliação genética de características quantitativas. Meuwissen et al. (1995) usou simulação para analisar aspectos de avaliação genética de características categóricas, Israel e Weller (1998) e Israel e Weller (2002) geraram genes candidatos e efeitos de Qtls por meio da simulação. Entretanto, nenhum dos estudos citados descreveu ou enfocou o algoritmo implementado.

Embora esta técnica seja intensamente utilizada nos projetos de pesquisa, as informações e discussões sobre os métodos adotados são escassas.

Informações mais detalhadas e a discussão sobre os algoritmos de simulação proporcionará dados de qualidade superior e desempenho de processamento melhorado, além do enriquecimento da literatura científica sobre o assunto.

Portanto, o objetivo do presente trabalho foi analisar um algoritmo para gerar características geneticamente correlacionadas utilizando a simulação de dados.

Para gerar valores fenotípicos para duas ou mais características geneticamente correlacionadas determinadas por l_i locos bialélicos, pressupõe-se, por motivos de simplicidade, a ausência de desvios de dominância e interações epistáticas. Também é assumido que informações sobre a matriz de (co)variâncias genéticas aditivas entre as características (G), o número de locos para cada característica(l_i) e as frequências iniciais dos alelos para cada locos (p e q) são conhecidas.

A idéia da solução apresentada é representar os valores fenotípicos das características correlacionadas por meio de uma função linear de n características não correlacionadas (z_i), de forma que a estrutura de dependência entre elas seja fornecida pelos coeficientes obtidos de uma decomposição de Cholesky, os quais permitirão que z_i seja transformada em g_i (os valores genotípicos em unidades reais).

Os valores genotípicos para cada característica e cada indivíduo serão armazenados em um vetor g , cujos elementos serão transformados em variáveis aleatórias independentes (z_i) com média 0 e variância 1 e armazenados em um vetor z , onde a matriz de (co)variâncias de z é uma matriz identidade chamada H . Dessa forma, o primeiro valor genotípico (a) pode ser obtido da expressão da variância para l locos:

$$1 = 2pq\sum l_i$$

em que l_i é o número de locos que determinam a característica i .

Para que z seja convertido em g , o artifício matemático utilizado é a aplicação da decomposição de Cholesky em G :

$$G = LL'$$

em que,

L é uma matriz triangular inferior e L' sua transposta (triangular superior).

A conversão é feita por:

$$g = Lz$$

O algoritmo segue abaixo:

1. Executar a Decomposição de Cholesky sobre a matriz G , para obter os coeficientes das funções lineares.

2. Para cada indivíduo e cada característica a ser simulada;

2.1. Calcule o valor genético aditivo (a) por : $\sqrt{\frac{1}{2pqli}}$

2.2. Armazene a ;

2.3. Para cada loco da i -ésima característica até n ;

2.3.1. Defina u_j de forma que para A_1A_1 , A_1A_2 e A_2A_2 os valores são a , 0 e $-a$ respectivamente;

2.3.2. Armazene u_j em uma lista w de tamanho l_i ;

2.4. Some os elementos de w para obter z_i ;

2.5. Armazene z_i ;

2.6. Faça $g_i = l_i z_i$;

g_i é o valor genético de cada animal para a característica i

2.7. Amostre o desvio de ambiente (e) de uma distribuição normal com média zero e variância σ_e^2

2.8. Armazene e ;

2.9. Faça $y_i = g_i + e$

y_i é o valor fenotípico da característica i

Apoio: FAPESP/CNPq

Referências

- Analla, M., Sanchez-Palma, A., Muoz-Serrano, A. e Serradilla, J. (1995). Simulation analysis with blup methodology of different data structures in goat selection schemes in Spain. *Small Ruminant Research*, 17:51-55.
- Carneiro, P., Corrêa, F., Carneiro, Júnior, J., Euclides, R., Pereira, C., Torres, R. e Silva, M. (2002). Oscilação genética em diferentes tamanhos de população sob seleção. *Arq. Bras. Med. Vet. Zootec.*, 54:84-92.
- Falconer, D. e Mackay, T. (1996). *Introduction to Quantitative Genetics*. Prentice Hall, 4 ed.
- Fonseca, R. (2003). *Use of Information on identified genes to reduce the selection bias on genetic evaluation*. Tese de doutorado, Universidade Federal de Viçosa, Viçosa-MG.
- Israel, C. e Weller, J. (1998). Estimation of candidate gene effects in dairy cattle populations. *J.Dairy.Sci.*, 81:1653-1662.
- Israel, C. e Weller (2002). Estimation of quantitative trait loci effects in dairy cattle populations. *J.Dairy.Sci.*, 85:1285-1297.
- Mabry, J. e See, M. (1990) Selection with the animal model versus selection within contemporary groups for swine. *J.Dairy.Sci.*, 73:2657-2665.
- Meuwissen, T., Engel, B. e van der Werf, J. (1995). Maximizing deletion efficiency for categorical traits. *J.Anim.Sci.*, 73:1933-1939.
- van Vleck, L., Pollak, E. e Oltenacu, E. (1987). *Genetics for the Animal Sciences*. W.H. Freeman and Company.