



XVI CONGRESSO MUNDIAL DE BUIATRIA

VI CONGRESSO LATINO AMERICANO DE BUIATRIA

XVI WORLD BUIATRICS CONGRESS
XVI^{ÈME} CONGRÈS MONDIAL DE BUIATRIE
16. WELTKONGRESS FÜR BUIATRIK

13 A 17 DE AGOSTO/1990
SALVADOR/BAHIA/BRASIL

Valentini, A.; Biagi, G.; Bagliacca, M.; Greppi, G.; Buckley, B. 1037
ACCOUNTING FOR HETEROGENEOUS VARIANCES IN
ESTIMATING THE BREEDING VALUES OF BEEF CATTLE.

TOMO II

A. Valentini¹, G. Biagi², M. Bagliacca³, G. Greppi³, B. Buckley⁴

¹ Istituto Zootecnia, Università Tuscia, Viterbo, Italy

² Istituto Patologia Speciale e Clinica Medica Veterinaria,
Università di Pisa, Italy

³ Dipartimento Scienze Anat., Fisiol., e delle Produzioni Animali,
Università di Pisa, Italy

⁴ Department of Animal Sciences, University of Hawaii, USA.

INTRODUCTION

The mixed model methodologies (1) are actually widespread for estimating the breeding values of many species. Several procedures are available for solving such models, even for very large data sets, but in any case the variances associated to the random effects must be known at least at proportionality (2). In fact, for a general linear mixed model:

$$y = Xb + Zu + e$$

where y is the observation vector, X Z are the known incidence matrices respectively of the fixed (b) and random (u) unknown effects, the solution for \hat{u} (i.e. the breeding values) is found based on the mixed model equations (MME)

$$\begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z+kA^{-1} \end{bmatrix} \begin{bmatrix} b \\ u \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \end{bmatrix}$$

The scalar k is the ratio $\text{var}(e)/\text{var}(u)$ and A^{-1} is the inverse of the relationship matrix. An error in the supplied values of the variances can cause a bias of the estimates breeding values and a rank exchange of the parents when the progeny size is limited (3).

Beef cattle is breed in a great variety of environments, with different mating and feeding methods. Therefore it is logical to expect that the environmental variances change even within regions. Moreover the use of A.I. is not so common in many countries and consequently the offspring size is often limited. Furthermore also the genetic variance changes when selection is applied. For these reasons an error in the estimated breeding values of beef cattle is likely to occur if the variances used in the MME are wrong.

An obvious solution would be to partition the $Z'Z$ matrix according to a criterion for which inside the partitioned matrices the variances are supposed to be homogeneous, then to solve the MME with a method that allows the contemporary estimation of both u and k , such as the EM-REML (4). The method is correct, but unfeasible for even medium size data sets.

An other problem encountered in beef cattle is that some registered traits are strongly affected by the dam, such as the weaning weight. In these cases it is advisable to consider also

the dam effect, as happens in the Animal Model (5). But the customary absorption-like process of the dam effect cannot be carried out if the variances are unknown. The previous solution can be applied, but the computational work became, of course, heavier.

We present a method, based on the SWEEP algorithm (6), for solving the MME, that is applicable for medium size data sets and allows the contemporary estimation of the breeding values and the k-ratios of the Z'Z matrix partitioned according to the different conditions of breeding and of selection.

METHOD

One of the most common methods for inverting large square matrices is to apply the SWEEP algorithm N-1 times to the "augmented" MME, where N is the rank of the MME. The term "augmented" means that the last row (and column) of the MME is made up by $(X|Z)'y$ and $y'y$.

The components of the variance are iteratively found by

$$\text{var}(e) = y'(y - X\hat{b} - Z\hat{u}) / (N - p^*)$$

$$\text{var}_2(\hat{u}) = [\hat{u}'\hat{u} + \text{var}_1(e) \text{trace}(C_{22})] / q$$

where: p^* is the rank of X, q is the rank of $[Z'Z + kA^-]$ and C_{22} is the portion of the inverted MME corresponding to Z'Z, var_1 and var_2 are two successive estimate of the variances.

Looking at the preceding formulas it is evident that not all the inverted matrix is needed, but only the diagonal (for computing the trace) and the last row (for computing the quadratics). Therefore a considerable time saving can be achieved if a procedure is applied to find only the needed elements and disregarding the rest of the MME.

Considering the SWEEP algorithm as described by Goodnight (6):

- 1) let $D = a_{kk}$, i.e. the diagonal element the SWEEP is based on,
- 2) divide the row k by D,
- 3) for every other i^{th} row with $i < k$, let $B = a_{ik}$ and subtract $B \cdot \text{row}_k$ from row_i ,
- 4) set $a_{kk} = 1/D$.

The algorithm we found modifies the step 3, considering in a different way the rows preceding from the ones following the sweeping element:

- 3) for every i^{th} row with $i < k$ sweep only the diagonal element,
- 3') for every i^{th} row with $i > k$, let $B = a_{ik}$ and subtract $B \cdot \text{row}_k$ from row_i .

In such way there is a substantial time saving, although for arrays saved in triangular mode a temporary vector has to be created for saving the a_{ik} elements for $i > k$, but only one row at time.

NUMERICAL EXAMPLE AND DISCUSSION

The example is in part taken from Cunningham & Henderson (7).
The model is:

$$y = Z_1 u_1 + Z_2 u_2 + Xb + e$$

The Z_1 Z_2 and X incidence matrices are each made up by 2 levels.

The MME are saved in triangular way and a guess for k (1.0) has already been added to the diagonal; the fixed factor follows the random ones and one level has been zeroed for finding a generalized inverse (8):

5.00						
0.00	9.00					
0.00	0.00	7.00				
0.00	0.00	0.00	5.00			
2.00	3.00	5.00	2.00	12.00		
10.00	30.00	21.00	15.00	48.00	500.00	

After the first iteration round, the augmented MME inverted with the usual SWEEP algorithm (9) are:

0.23						
0.02	0.13					
0.05	0.04	0.23				
0.03	0.02	0.05	0.23			
-0.07	-0.06	-0.12	-0.07	0.17		
-1.64	-2.59	-1.41	-2.11	-2.23	243.00	

If they are instead inverted by our algorithm, they result:

0.23						
0.00	0.13					
0.00	0.00	0.23				
0.00	0.00	0.00	0.23			
-0.47	-0.33	-0.71	-0.40	0.17		
-1.64	-2.59	-1.41	-2.11	-2.23	243.00	

As it is evident the diagonal and the last row are conserved, while the inner elements are different. Therefore the components of the variance can be computed as well as the genetic values, although the covariances between them cannot be computed, but this is usually not needed. In our examples the convergence is reached after few iterations and the final values are: 4.35 and 2.69 for the variance of the random factors and 12.10 for the error term.

As previously mentioned, for the rows preceding the sweeping element the only element needed is the diagonal one, therefore a straightforward practice would be to put the larger block of the MME as first and to drop all its elements but the diagonal ones.

In beef cattle the obvious candidate for that block would be the incidence matrix of the cows. Thus the analysis of maternally influenced traits (such as weaning weight) would be greatly improved at the only expense to add a diagonal vector to the MME. On the other hand the customary absorption process cannot be applied, since the cow variance component is any way to be computed.

Using the proposed strategy, it is then possible to apply the EM-REML procedure to medium size data sets. If the computer runs out of memory, it is still possible to carry out the analysis, since the sweeping is done one row at time and then only one vector needs to be stored in memory.

CONCLUSION

The algorithm presented consents to contemporarily evaluate the variances and the breeding values in shorter times than the actual routines. The matrices can eventually be partitioned to account for etherogeneity of the variances and the evaluation of the breeding values and the ranking of the parents can be improved, while saving substantial computer resources.

REFERENCES

- 1) Henderson C.R.: 1973 Proc. of the Symp. in hon. of Dr. J.L. Lush. ASAS and ADSA, Champaign, IL, USA.
- 2) Henderson C.R.: 1975 Biom. 31, 423.
- 3) Nardone A., Valentini A.: 1989 Ann. Zootech. 38, 315.
- 4) Harville D.A.: 1977 J. Am. Stat. Assoc. 72, 320.
- 5) Quaas R.L., Pollak E.J.: 1980 J. of Anim. Sci. 43, 1277.
- 6) Goodnight H.J.: 1979 The American Statistician 33, 149.
- 7) Cunningham E.P., Henderson C.R.: 1968 Biom. 24, 13
- 8) Searle S.R.: 1971 Liner Models, Wiley & Sons, N.Y.
- 9) Clarke M.R.B.: 1982 Appl. Statist. 31, 166.

SUMMARY

In estimating the breeding values by mixed model methodologies it is quite customary to assume that the variances of the random components (usually the parent and the error factor) are normally distributed with the same variances across the design. Heterogeneous variances can indeed occur as consequence of selection and/or different conditions of breeding. Ignoring this fact can yield biased estimates mainly in beef cattle, where the progeny size is usually limited and a strong dam effect occurs for some registered trait. A solution can consist in a simultaneous estimation of the variances and the breeding values by EM-REML. An algorithm based on a modified SWEEP routine is presented, which allows to take into account the dam effect and the eventually heterogeneous variances, while not overwhelming the computer resources. The algorithm consists in a partial absorption of the dam effect, which is indeed estimated for computing the variance component.

ZUSAMMENFASSUNG

Bei der Bestimmung der genetischen Werte mit Methodologien, die gemischte Modelle benutzen, ist es üblich anzunehmen, daß die Varianzen der zufälligen Komponenten (normalerweise die Abstammung und der Fehlerfactor) gewöhnlich mit den selben Varianzen über das Schema verteilt sind. Heterogene Varianzen können wirklich als Folge von Selection und/oder von unterschiedlichen genetischen Bedingungen auftreten. Wird diese Tatsache nicht beachtet, können sich daraus Fehlbestimmungen ergeben; dies vor allem bei Schlactrindern, deren Zahl der Nachkommenschaft normalerweise begrenzt ist, und bei denen für einige registrierte Merkmale ein starker Einfluß mütterlicherseits auftritt. Dieses Problem kann durch eine gleichzeitige EM-REML Bestimmung der Varianzen und der genetischen Werte gelöst werden. Es wird ein Algorithmus vorgestellt, der auf einer modifizierten SWEEP-routine Basierte, die es ermöglicht, den Einfluß mütterlicherseits und eventuelle heterogene Varianzen zu berücksichtigen, ohne die Möglichkeiten des Computers zu überschreiten. Der Algorithmus besteht in einer teilweisen Absorption des Einflusses mütterlicherseits, der dagegen bei der Verarbeitung der Varianzkomponente berücksichtigt wird.

SUMARIO

Avaliando os valores genéticos através de metodologias com modelos mistos é muito usual assumir que as variações dos componentes casuais (habitualmente a ascendência e a factor de erro) são normalmente distribuídas com as mesmas variações através do esquema. As variações heterogêneas podem de facto ocorrer como consequência de seleção e/ou diferentes condições genéticas. Ignorando este facto pode levar a avaliações falsas, sobretudo em bovinos para abate, nos quais o numero da progenitura é habitualmente limitado e nos quais um forte efeito materno ocorre para alguns traços registrados. Uma solução pode consistir numa avaliação simultânea das variações e dos valores genéticos por EM-REML. Um algoritmo baseado numa SWEEP routine modificada é apresentado, o qual permite tomar em consideração o efeito materno e as eventuais variações heterogeneas, sem dominar os meios do computador. O algoritmo consiste numa absorção parcial do efeito materno o qual é de facto estimado para avaliar o componente de variação.



XVI CONGRESSO
MUNDIAL DE BUIATRIA

VI CONGRESSO
LATINO AMERICANO
DE BUIATRIA

Os trabalhos publicados nestes livros são a reprodução exata dos originais enviados pelos autores.

The scientific manuscripts published on the proceedings are the exact reproduction of the material furnished by the authors.

Edita : Interlink Consultoria & Eventos Ltd.
Av. Centenário, 2883 - Edf. Victória Center
Salas 208/209 - Chame-Chame - CEP - 40.160
Salvador-Bahia-Brasil

Imprime : Impressora Rocha Ltda
Salvador-Bahia-Brasil