Combining Disparate Estimates of Genetic Correlations

***Hossein Jorjani*[1]*, Bert Klei*[2] *and Ulf Emanuelson*[1]**
[1]*Interbull Centre, Box 7023, S - 750 07, Uppsala, Sweden,*
[2]*Holstein Association USA, Inc, 1 Holstein Place, Brattleboro VT 05301-0808, USA,*

Abstract
A simple method to obtain weighted bended genetic correlation matrices is proposed and the use of the suggested algorithm is demonstrated by a small example. The conclusion is that the weighted bending method should replace un-weighted bending method practiced hitherto in Interbull evaluations.

Introduction

Genetic correlations among countries for the various traits evaluated by Interbull are estimated for groups of countries at a time, mainly because of computational demands of estimating them. The practice is slightly different for production and udder health traits on one hand and conformation traits on the other hand. For production and udder health traits, a large number of country combinations involving two, three, four and occasionally five countries are used. For conformation traits, countries are first divided into smaller groups of 2-4 countries based on number of common bulls and ¾ sibs that exist among them. Correlations for conformation traits among countries are then estimated by using two groups of countries at a time with USA and CAN as link providers.

Irrespective of the trait, the above procedure leads to the existence of more than one estimate of correlation for some country combinations. These multiple estimates need to be first combined into one single estimate, and thereafter, all correlations among all countries need to be combined into one single correlation matrix (START) for each trait and breed. For the START matrix one can use the average of all existing correlations (AVE), the maximum of all existing correlations (MAX), or a combination of these two that yields the least non-positive definite (non-PD) matrix (*i.e.* the smallest eigenvalue is least negative).

The START matrix is, almost invariably, non-PD and needs to be bended by a simple iterative method (to be described shortly in the next section) to obtain the correlation matrix that will be used for breeding value estimation (FINAL). The main disadvantage of the current bending method (defined as **un-weighted bending**) is that all elements of START are treated as equally reliable and each and every one of them is prone to be changed in the bending process. Intuitively, this is not a desirable property. One popular method to prevent more reliable estimates from changing is “eye-balling”, which can basically be defined as the visual inspection of the correlation matrix at the end of each round of iteration in the bending process and restricting changes in the more reliable estimates to a “minimum and acceptable level”. However, this is a subjective method and depends on whose “eye” is doing the inspection.

Our aim in this study was to develop a formal method for incorporation of reliabilities of the point estimates into the bending process (defined as **weighted bending**).

Method

Let **V** be a matrix of (co-)variances comprising separately estimated values. **V** is a non-PD matrix. Further, let **W** be a matrix of weighting factors for elements of **V**. The bending process then would be as follows:

1) Determine matrix of eigenvectors, $\mathbf{U}_n$, and diagonal matrix of eigenvalues, $\mathbf{D}_n$, of $\mathbf{V}$. Hence, $\mathbf{V}_n = \mathbf{UDU}'$, where n denotes iteration number;
2) Replace $\mathbf{D}_n$ with $\mathbf{\Delta}_n$, where $\delta_{i,i}=d_{i,i}$, for $d_{i,i} \geq 0.0$ and $\delta_{i,i} = \varepsilon$, otherwise. Set the value of ε to a small positive real number;
3) Calculate a new covariance matrix: $\mathbf{V}_{n+1} = \mathbf{V}_n - [\mathbf{V}_n - \mathbf{U}_n\mathbf{\Delta}_n\mathbf{U}_n'] \otimes \mathbf{W}$;
4) Repeat until $\mathbf{V}_{n+1}$ is PD.

It can easily be seen that in the case of un-weighted bending $\mathbf{W}$ is a matrix with all elements equal to 1, i.e. $\mathbf{W} = \mathbf{J}$.

For extension to correlation matrices one must account for some special properties of $\mathbf{R}$, I) trace of $\mathbf{R}$ (tr($\mathbf{R}$)), and consequently tr($\mathbf{D}$), are equal to the order of $\mathbf{R}$ and II) diagonal elements of $\mathbf{R}_{n+1}$ (and $\mathbf{U}_n\mathbf{\Delta}_n\mathbf{U}_n'$) must be equal to unity. To accommodate property I, add a step 2.1 for scaling elements of $\mathbf{\Delta}$ by the following factor tr($\mathbf{D}$) / tr($\mathbf{\Delta}$). To accommodate for property II add a step 3.1 as $r_{i,j} = r_{i,j} / \mathrm{sqrt}(r_{i,i} * r_{j,j})$. Alternatively, for step 3.1, one can switch back and forth between $\mathbf{V}$ and $\mathbf{R}$ and recalculate correlations from covariances at each round.

Example

To demonstrate the effects of choosing different values for the $\mathbf{W}$ matrix the following matrices will be used. Let $\mathbf{R}$ be equal to:

$$\begin{bmatrix} 1.00 & 0.95 & 0.80 & 0.40 & 0.40 \\ & 1.00 & 0.95 & 0.80 & 0.40 \\ & & 1.00 & 0.95 & 0.80 \\ & & & 1.00 & 0.95 \\ & & & & 1.00 \end{bmatrix}$$

And let $\mathbf{W}$ be either equal to $\mathbf{J}$ or equal to the reciprocal of number of bulls used in the estimation of correlations:

$$\begin{bmatrix} 1000 & 500 & 20 & 50 & 200 \\ & 1000 & 500 & 5 & 50 \\ & & 1000 & 20 & 20 \\ & & & 1000 & 200 \\ & & & & 1000 \end{bmatrix}$$

The vector of eigenvalues for $\mathbf{R}$ is equal to $\mathbf{D}_1' =$ [3.995, 0.985, 0.236, -0.031, -0.185]. By using $\mathbf{W}=\mathbf{J}$ and $\varepsilon=10^{-4}$ a PD correlation matrix is obtained after 1 rounds of iteration and the FINAL correlation matrix, $\mathbf{R}_1$, is equal to:

$$\begin{bmatrix} 1.00 & 0.87 & 0.77 & 0.42 & 0.36 \\ & 1.00 & 0.90 & 0.70 & 0.42 \\ & & 1.00 & 0.90 & 0.77 \\ & & & 1.00 & 0.87 \\ & & & & 1.00 \end{bmatrix}$$

However, using the weighting matrix equal to the reciprocal of number of bulls leads to a PD correlation matrix after 1713 rounds of iteration and the resulting FINAL correlation matrix, $\mathbf{R}_{1713}$, is equal to:

$$\begin{bmatrix} 1.00 & 0.94 & 0.78 & 0.38 & 0.39 \\ & 1.00 & 0.93 & 0.49 & 0.41 \\ & & 1.00 & 0.74 & 0.62 \\ & & & 1.00 & 0.93 \\ & & & & 1.00 \end{bmatrix}$$

Comparison of results of un-weighted and weighted bending for this constructed example and for real correlation matrices (results not shown) indicate that the average change in correlations is of the same magnitude. However, in the weighted bended matrices there are fewer / smaller changes for the more reliable estimates and more / larger changes for the less reliable estimates.

Discussion

Computational cost of implementing weighted bending is quite negligible, because of the small size of the matrices involved.

In the lack of standard errors of the correlations we considered number of common bulls to be a good approximation to the reliability values. Further, for the sake of the present example it would suffice to use number of common bulls.

In the light of inevitability of the need to perform a bending of the non-PD correlation matrices, the question to raise is whether we are ready to accept changes even in those correlations that have a high reliability. It seems quite logical that this is not the case and we would prefer to restrict changes in the more reliable estimates to a minimum.

Comparison of the results of the two alternative weighting matrices indicates that the proposed algorithm does exactly as we intuitively would like it to do, that is to prevent reliable correlations to change dramatically. Based on these results we suggest that this new algorithm to be adopted in the Interbull evaluations.

Acknowledgements

HJ would like to acknowledge financial support from USDA/NAAB.