The REML estimation of variance components using an integrated AI-EM algorithm: experience with use of carcass data of Japanese Black cattle

Hiroaki Iwaisaki¹ and Ichiro Ashida²

¹- Department of Agro-biology, Faculty of Agriculture, Niigata University, Niigata 950-2181, Japan. Phone: +81-25-262-6885; Email: iwsk@agr.niigata-u.ac.jp; ²- Department of Health and Nutrition, University of Health and Welfare, Niigata 950-3198, Japan. Phone: +81-25-257-4414; Email: asida@nuhw.ac.jp

Abstract
In the genetic evaluation for carcass traits of the Japanese native beef cattle, we have used a two-stage estimation procedure, in which the empirical best linear unbiased prediction of breeding values is implemented using an individual animal model and variance components estimated by the restricted maximum likelihood (REML) employing an average information (AI) algorithm. The AI algorithm is efficient, as long as the successful updating of the estimate is achieved. However, the AI-REML estimate may exceed the boundary, resulting in a negative value, particularly in the cases of relatively small-sized data sets. A simple procedure can be used, in which AI algorithm is employed, but the expectation-maximization (EM) type algorithm is utilized to adjust the step direction and the size, when AI algorithm failed to increase the REML log-likelihood in certain round of iteration. The current integrated AI-EM algorithm may guarantee the final estimate within the parameter space and requires no additional memory over that for AI algorithm. Experience with relatively small data sets on carcass traits of the Japanese Black cattle indicates that the current procedure could perform well and converge fast relative to EM algorithm using the Aitken acceleration.

Key Words: AI algorithm, convergence, EM algorithm, estimation, restricted maximum likelihood, variance component

Introduction
The official genetic evaluation for carcass traits of the Japanese native beef cattle (Wagyu cattle) is carried out by prefecture (province) using an individual animal model. A two-stage estimation procedure is used, in which the restricted maximum likelihood estimation (REML; Patterson and Thompson, 1971) of variance components is first conducted using an average information (AI) algorithm (Johnson and Thompson, 1995) and then, replacing the parametric values of variance components by the REML estimates, the empirical best linear unbiased prediction of breeding values is implemented. About 300,000 animals in the pedigree file of the Japanese Black cattle are evaluated in one prefecture, while the number of animals with records to be evaluated is approximately one thousand in some prefectures.

The REML procedure using an expectation-maximization (EM) type algorithm (Dempster et al., 1977) may provide the estimates within the allowable parameter space (e.g., Harville, 1977), however the AI-REML estimate may exceed the boundary in the cases of relatively small-sized data sets. This study concerned the performance of an integrated AI-EM algorithm.

Materials and Methods
A total of 7 data sets including records of 6 carcass traits of Japanese Black cattle each were used in the current analyses. The number of animals with records and the total number of animals in the pedigree file were 1,301 and 4,389 for the smallest data set and 22,631 and 55,972 for the largest
data set, respectively. Six traits analyzed were carcass weight (CWT), ribeye area (REA), rib thickness (RT), subcutaneous fat thickness (SFT), estimated yield percent (EYP) and marbling score (MS). The operational model employed was a single-trait, additive genetic, individual animal model, and fixed effects considered were sex, or steer and female, fattening farm, year of slaughter and carcass market (discrete variables), slaughter age (linear and quadratic covariates), and degree of inbreeding (linear covariate).

A total of 42 variables were analyzed by an EM-type algorithm using the Aitken acceleration (EMa), an average information (AI) algorithm (Ashida and Iwaisaki, 1999), and an integrated AI-EM algorithm (AIe) of our concern, in which AI algorithm is used, but EM algorithm is employed to adjust the step direction and the size, when AI algorithm failed to increase the REML log-likelihood in certain round of iteration.

For AI algorithm, we used the following formula:

Let the model be \( y = Xb + \sum_{i=1}^{s} Z_iu_i + e \), where \( y \) is a vector of observations, \( b \) is a vector of fixed effects, \( u_i \) is a \( q_i \times 1 \) vector of random effects with \( s \) representing the number of random factors except for the residual term, \( e \) is a vector of random residuals, \( X \) and \( Z_i \) are incidence matrices.

It is assumed that \( Var(y) = \sum_{i=1}^{s} Z_iG_iZ_i^t\sigma_i^2 + I\sigma_0^2 \), where \( \sigma_i^2 \) and \( \sigma_0^2 \) are variance components.

Then, having \( \theta' = [\sigma_0^2 \sigma_1^2 \sigma_2^2 \cdots \sigma_s^2] \) and following Ashida and Iwaisaki (1999), an expression for the AI matrix is given as

\[
A(\theta) = \frac{1}{2} \begin{bmatrix}
    a_{00} & \left\{ \sum_{i=1}^{s} \sum_{j=1}^{s} \hat{u}_i^tG_i^{-1}C_{ij}G_j^{-1}\hat{u}_j \right\}\sigma_i^{-2}\sigma_j^{-2} \\
    Sym. \left\{ \sum_{i=1}^{s} \sum_{j=1}^{s} \hat{u}_i^tG_i^{-1}\hat{u}_j \right\} - \left\{ \sum_{i=1}^{s} \sum_{j=1}^{s} \hat{u}_i^tG_i^{-1}C_{ij}G_j^{-1}\hat{u}_j \right\}\sigma_i^{-2}\sigma_j^{-2} \end{bmatrix}^{s}
\]

with

\[
a_{00} = \left( y'y - b'X'y - \sum_{i=1}^{s} \hat{u}_i^tZ_i'y - \sum_{i=1}^{s} \hat{u}_i^tG_i^{-1}\hat{u}_i - \sum_{i=1}^{s} \hat{u}_i^tG_i^{-1}C_{ij}G_j^{-1}\hat{u}_j \right)\sigma_0^{-6},
\]

where \( r_i = \sigma_0^2 / \sigma_i^2 \), \( C_{ij} \) is the appropriate sub-matrix of a generalized inverse of the coefficient matrix of the mixed model equations, \( \hat{b} \) and \( \hat{u} \) are the solutions to the mixed model equations, and the notation \( \{s \}^{s}_{i=1} \) represents a row vector with the \( s \) elements for \( i = 1 \) to \( s \) inside the brace, \( \{s \}^{s}_{i,j=1} \) stands for a matrix of order \( s \times s \) whose elements for \( i = 1 \) to \( s \) and \( j = 1 \) to \( s \) are given in the brace, and \( \delta_{ij} \) is the Kronecker delta with \( \delta_{ii} = 1 \) and \( \delta_{ij} = 0 \) for \( i \neq j \).

Replacing the Hessian matrix by \( A(\theta) \), a set of equations for the iterative estimation can be written as

\[
\hat{\theta}_{t+1} = \hat{\theta}_t + \left[ A(\hat{\theta}_t) \right]^{-1}\partial L / \partial \theta |_{\theta = \hat{\theta}_t},
\]

where \( \hat{\theta}_{t+1} \) and \( \hat{\theta}_t \) are the estimates of \( \theta \) in the \( t+1 \)th and the \( t \)th iterations, \( A(\hat{\theta}_t) \) is the matrix \( A(\theta) \) calculated at \( \theta = \hat{\theta}_t \), and \( \partial L / \partial \theta |_{\theta = \hat{\theta}_t} \) is the gradient vector, or the vector of the first
derivatives of the REML log-likelihood function with $\theta$ replaced by $\hat{\theta}_i$.

Initial values for the variance components for fattening farm and animal were assigned as 20 and 40% of the unbiased variance calculated using raw data. Convergence was assumed when changes in the variance ratios between two successive rounds were less than $10^{-8}$.

**Results**

For 42 variables analyzed, EMa algorithm performed well and gave successfully converged estimates of farm, animal and residual variances within the allowable parameter space. With this algorithm, the number of iterations required for convergence ranged from 31 to 299. Estimates by AI algorithm converged successfully except for 3 cases, requiring less rounds of iteration compared with EMa algorithm. The cases where AI algorithm failed to give normally converged values are presented in Table 1. For these cases, we employed AIe algorithm, and consequently convergence was normally attained for all the cases. Final estimates of each variance component agreed within enough decimal places for EMa and AIe algorithms. The numbers of iterations with AIe algorithm were about 10 and obviously less than those with EMa.

**Discussion**

The estimates by the AI approach using Newton’s method may exceed the boundary, resulting in negative values of the parameters. One way would be to modify the above-mentioned estimation equations using Marquardt’s (1963) idea and replacing $A(\hat{\theta}_i)$ by $A(\hat{\theta}_i) + I \tau$ where $\tau$ is a constant, as suggested by Johnson and Thompson (1995). However, one needs to select a proper value of $\tau$ in a given setting. Parameterizations such as Cholesky transformation may also be used.

The current integrated AI-EM algorithm may guarantee the parameter estimates within the boundary. The current approach also requires no additional memory in addition to that for the AI algorithm, since the first partial derivative of the REML log-likelihood function is already used as the gradient vector in the AI algorithm. Experience with relatively small data sets on carcass traits of the Japanese Black cattle indicates that the current procedure could perform well and converge fast relative to the EM algorithm using the Aitken acceleration.

**References**


Patterson HD, Thompson R (1971) Recovery of inter-block information when block sizes are unequal. *Biometrika* 58: 545-554
## Tables Legends

### Table 1. Results of estimation with the three algorithms

<table>
<thead>
<tr>
<th>Case</th>
<th>Trait&lt;sup&gt;a&lt;/sup&gt;</th>
<th>No. of records</th>
<th>No. of animals&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Algorithm&lt;sup&gt;c&lt;/sup&gt;</th>
<th>No. of iterations</th>
<th>Estimate of variance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Farm</td>
</tr>
<tr>
<td>1</td>
<td>REA</td>
<td>1,301</td>
<td>4,389</td>
<td>AI</td>
<td>-</td>
<td>1.1336</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>DVDa</td>
<td>22.9780</td>
<td>20.5745</td>
</tr>
<tr>
<td>2</td>
<td>EYP</td>
<td>1,301</td>
<td>4,389</td>
<td>AI</td>
<td>-</td>
<td>.0202</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>EMa</td>
<td>.0202</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>SFT</td>
<td>1,323</td>
<td>3,864</td>
<td>AI</td>
<td>-</td>
<td>.0082</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>EMa</td>
<td>.0082</td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup> REA: ribeye area, EYP: estimated yield percent, SFT: subcutaneous fat thickness.

<sup>b</sup> Animals in the pedigree file.

<sup>c</sup> AI: an average information algorithm, EMa: an expectation-maximization algorithm using the Aitken acceleration, Ale: the current integrated AI-EM algorithm.