Estimation of variance components for some production traits of Iranian Holstein dairy cattle using Bayesian and AI-REML methods

Sadegh Alijani*, Mehdi Jasouri, Nasrolah Pirany and Hossein Daghigh Kia

Department of Animal Sciences, Faculty of Agriculture, University of Tabriz, Tabriz, 516614766, Iran

*Corresponding author: sad-ali@tabrizu.ac.ir

ARTICLE HISTORY

Received: November 30, 2011
Revised: February 18, 2012
Accepted: April 25, 2012

Key words:
AI-REML
BAGS
Iranian Holstein dairy cattle
Variance components

ABSTRACT

The used data set included the records of 131990 Iranian Holstein dairy cattle for first three lactations that were collected from 1981 to 2008 time period by Animal Breeding Center, Iran. The traits which were considered for 305 days of lactation included milk, fat and protein yield and percentages of milk fat and protein. Variance components were estimated using average information restricted maximum likelihood (AI-REML) algorithm using AIREMLF90 software under single trait and repeatability models and Bayesian method by using a Gibbs sampling technique (BAGS) and by MTGSAM and GIBBS3F90 software by same models. The linear statistical models of the analyses included herd-year-season and lactations as fixed effects, age at calving as covariate and animal and permanent environment as random effects. The ranges of heritability estimates for lactations 1 to 3 by animal single and repeatability models using AI-REML and BAGS methods were 0.19 to 0.29, 0.17 to 0.26, 0.20 to 0.25, 0.21 to 0.25 and 0.19 to 0.35 for milk, fat and protein yield and percentage of milk fat and protein, respectively. Repeatability estimates by using BAGS method were 0.44, 0.35, 0.43 and by AI-REML method the values were 0.43, 0.34, 0.39 for milk, fat and protein yield, respectively. The results showed that estimated genetic parameter values by AI-REML analyses for all traits and lactations in both models were smaller than BAGS method. In addition, estimated heritability values for later lactations were lower in comparison with the first lactation.

INTRODUCTION

Genetic parameters that are essentials for most animal breeding programs are based on (co) variance components. Therefore, accurate estimation of these components will result in satisfaction of animal breeding programs. Based on this fact, estimation methods of (co) variance components have been greatly improved over the last four decades and development of these methods has been a constant objective of many animal breeding researchers.

In statistical science, two major approaches of conventional (or frequent) and Bayesian are used. In the mid-1980s, the maximum likelihood method was applied for estimation of (co) variance components and now the new and optimized version of Restricted Maximum Likelihood by average information (AI-REML) method is being used. This method assumes that records of traits are normally distributed (Misztal, 1994; Thompson and Mantysaari, 1999). Bayesian approach provides a complete paradigm for both statistical inference and decision making under uncertainty. Bayesian methods solve many of the difficulties faced by conventional statistical methods and extend the applicability of statistical methods. It exploits the use of probabilistic models to formulate scientific problems. To use Bayesian statistics, there is computational difficulty and secondly, Bayesian methods require specifying prior probability distributions. Markov Chain Monte-Carlo (MCMC) methods were applied to overcome the computational difficulty, and interest in Bayesian methods was renewed (Kokate et al., 2011). In the past decades the Bayesian method by using a Gibbs sampling technique (BAGS) has been mostly used for estimation of variance components in animal breeding (Wang et al., 1994; Sorensen et al., 1994; Van Tassel and Van Vleck, 1996; Magnabosco et al., 1998 and 2000). In BAGS method, sampled variance
components are taken from the inverted chi-square or its multiple dimensions Wishart distribution (Misztal, 2008).

In Gibbs sampling algorithm three parameters should be determined prior to its application. The first one is the amount of total cycles of sampling to reach reliable estimation and the second one is the number of sampling that must be discarded before the produced sampled values to include in the true posterior distribution, because these sampled values are not still correct. The third one is the interval between sampled values to obtain independent samples from the posterior distribution. Compared to REML, Bayesian methods have the advantage of allowing the inclusion of prior knowledge about unknown parameters in the analysis. Additionally, the Bayes theorem provides a solution for the finite sample size problem, since an exact a posteriori distribution exists for each large or small data set from which inferences can be drawn (Misztal, 2008; Borquis et al., 2010; Leuenberger and Wegmann, 2010). When a large data set is analyzed, a priori information tends to be overwhelmed by the likelihood function in the establishment of the posterior distribution. In this case, parameter estimates are close to those obtained by methods based on likelihood functions. However, this may not be true when the sample size is limited because the maximum likelihood procedure only possesses well-defined properties when the sample size is large enough (Kuhnern, 2009; Guillemaud, 2010; Fong et al., 2010). In traditional dairy cattle breeding, first lactation records or, by assuming correlation of unity among yield in different lactations, repeated records of more lactations were used in genetic evaluation. In the first case it is assumed that first lactation records can provide adequate information about later lactation records by considering that same genes influence on first and later lactations (Albuquerque et al., 1996). By improving in computing power and statistical modeling all lactation records are used in genetic evaluation and this has resulted in increasing accuracy of evaluation of dairy animals.

The main aim of this research was estimation of variance components of milk production traits in the first three lactations of Iranian Holstein dairy cattle using single trait and repeatability linear animal model by AI-REML and BAGS methods and determining Gibbs sampling parameters for applying BAGS method.

**MATERIALS AND METHODS**

Following editing data, the data set used in this research consisted of 254833 field records of first three lactations of Holstein cows that were collected from 1981 to 2008 in Animal Breeding Center of Iran. Total number of animals in the pedigree was 608478. Milk (MY), fat (FY) and protein yield (PY) and percentage of milk fat (FP) and protein (PP) was considered as traits analyzed for 305 days of lactation. All of the traits were adjusted to 2008 in Animal Breeding Center of Iran. Total number of animals in the pedigree was 608478. Milk (MY), fat (FY) and protein yield (PY) and percentage of milk fat (FP) and protein (PP) was considered as traits analyzed for 305 days of lactation. All of the traits were adjusted to

In the second step of data preparation, ages of animals at calving were edited and for this purpose, age range (days from birthday) of 660 to 1200 days for first lactation, 990 to 2000 days for second lactation and 1350 to 2500 days for third lactations were considered. The data set editing was done by using SAS (2003) and RENUMF90 (Misztal, 1999b) and FOXPRO 9 (2006) software.

In increasing accuracy of evaluation of dairy animals.

**Table 1: Summary information of production traits**

<table>
<thead>
<tr>
<th>Traits</th>
<th>N</th>
<th>Max</th>
<th>Min</th>
<th>Mean</th>
<th>CV (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MY1 (kg)</td>
<td>131990</td>
<td>11000</td>
<td>3000</td>
<td>7315</td>
<td>18.99</td>
</tr>
<tr>
<td>MY2 (kg)</td>
<td>82070</td>
<td>14775</td>
<td>3000</td>
<td>8132</td>
<td>21.75</td>
</tr>
<tr>
<td>MY3 (kg)</td>
<td>40773</td>
<td>16166</td>
<td>3000</td>
<td>8492</td>
<td>22.70</td>
</tr>
<tr>
<td>FY1 (kg)</td>
<td>106400</td>
<td>351</td>
<td>125</td>
<td>237</td>
<td>20.10</td>
</tr>
<tr>
<td>FY2 (kg)</td>
<td>67040</td>
<td>500</td>
<td>125</td>
<td>267</td>
<td>24.15</td>
</tr>
<tr>
<td>FY3 (kg)</td>
<td>32177</td>
<td>597</td>
<td>130</td>
<td>282</td>
<td>24.95</td>
</tr>
<tr>
<td>FY1 (%)</td>
<td>104121</td>
<td>5.94</td>
<td>2.50</td>
<td>3.32</td>
<td>13.01</td>
</tr>
<tr>
<td>FY2 (%)</td>
<td>68041</td>
<td>5</td>
<td>1.50</td>
<td>3.28</td>
<td>15.63</td>
</tr>
<tr>
<td>FY3 (%)</td>
<td>32644</td>
<td>5</td>
<td>1.33</td>
<td>3.31</td>
<td>15.58</td>
</tr>
<tr>
<td>PY1 (kg)</td>
<td>77489</td>
<td>300</td>
<td>115</td>
<td>226</td>
<td>16.09</td>
</tr>
<tr>
<td>PY2 (kg)</td>
<td>52933</td>
<td>466</td>
<td>110</td>
<td>258</td>
<td>19.19</td>
</tr>
<tr>
<td>PY3 (kg)</td>
<td>26581</td>
<td>479</td>
<td>110</td>
<td>268</td>
<td>19.84</td>
</tr>
<tr>
<td>PY1 (%)</td>
<td>80131</td>
<td>3.85</td>
<td>2.30</td>
<td>3.07</td>
<td>7.64</td>
</tr>
<tr>
<td>PY2 (%)</td>
<td>53023</td>
<td>5</td>
<td>1.84</td>
<td>3.11</td>
<td>8.54</td>
</tr>
<tr>
<td>PY3 (%)</td>
<td>25920</td>
<td>5</td>
<td>1.66</td>
<td>3.09</td>
<td>8.66</td>
</tr>
</tbody>
</table>

| N: number of recorded animals; Max: Maximum; Min: Minimum; CV: coefficient of variance. |

The analyses were conducted using a single trait and repeatability linear animal model. For univariable analyses, three lactation records were analyzed individually using a single trait linear animal model as follows (in matrix notation):

\[ y = Xb + Zu + e \]

where, \( y \) is the observation vector, in this case is one of the lactation records. \( b \) is the fixed effect vector of the HYS contemporary groups effect and regression coefficient of covariate of age of the animal at calving coefficient. \( u \) is the additive random genetic effect vector of the animal. \( e \) is the residual random effect vector. \( X \) and \( Z \) are the incidence matrices that relate the data to the fixed and random effects, respectively.

The structure of (co) variance matrix in random variables in this model is as follows:

\[
\begin{bmatrix}
y \\
u \\
e
\end{bmatrix} =
\begin{bmatrix}
V & ZG & R \\
Z' & G & 0 \\
R & 0 & R
\end{bmatrix}
\]

Where,

\[ V = ZGZ' + R. \]

After considering \( R = I \sigma^2_e \) and \( G = A \sigma^2_u \) that \( A \) is the matrix of additive relationship between individuals and \( I \) is the identically matrix in these equations, the following MME (mixed model equation) was solved for obtaining BLUP prediction of animal random effect:

\[
[XX' ZZ' + AA^{-1}k]^{-1} \begin{bmatrix}
X'y \\
Z'y
\end{bmatrix}
\]

where,

\[ k = \frac{\sigma^2_e}{\sigma^2_u} = 1 - h^2 \]

\[ \frac{\sigma^2_u}{h^2} \]

The milk production traits in three lactations of each animal were considered as repeated measurements;
therefore a repeatability animal model was applied. This model assumes a genetic correlation equal to one among the various lactation measurements. The following statistical model was used for these analyses:

\[ y = Xb + Zu + Wpe + e \]

In which \( y \), \( b \), \( u \), \( X \) and \( Z \) are defined as single trait model and \( be \) is the random effects vector of the permanent environment and \( W \) is the incidence matrix that relates the observations in \( y \) with this effect. In this model it was assumed that

\[
\begin{bmatrix}
\var(u) \\
\var(pe) \\
\sigma^2_e \\
\end{bmatrix} = 
\begin{bmatrix}
0 & 0 & 0 \\
0 & 0 & 0 \\
1 & 0 & 1 \\
\end{bmatrix}
\]

For REML method, variance components were estimated using AI-REML algorithm and by applying AIREMLF90 software (Miszta, 1999a) under single trait and repeatability model. Single traits analyzed by BAGS method using MTGSAM (Multiple Trait using Gibbs Sampling under Animal Model) software (Van tassel and Van Vleck, 1996) and for the repeatability animal model, GIBBS3F90 software (Miszta, 1999c) were used. Totally, 36 analyses were done in this research. In order to study different levels of Gibbs Sampling parameters, different strategies were applied, these strategies were used for estimation of (co) variance component for traits of 305 days milk yield (MY) and mature equivalent milk yield (ME). These strategies included different combinations of different values for each of Gibbs Sampling parameters constituting 100, 150 and 200 for thinning interval; 1000, 10000 and 30000 for bur-in period; and 100000, 150000 and 200000 for total chain length. Considering all combinations, finally 56 analyses were done.

RESULTS AND DISCUSSION

The result of applying different strategies of BAGS method showed that firstly thinning interval had no effect on the mean, mode or median of the posterior marginal distribution of genetic parameters. Secondly, all values of chain length and bur-in period had very little effect on the amount of variance components estimations in this research. These results are presented in Table 2. Based on these results in other BAGS analyses the values of 100000, 30000 and 200 were used for total chain length, bur-in period and thinning interval, respectively. The results of this BAGS analysis were reported as the results of this estimation method.

The estimated values of variance components, genetic parameters and standard errors of first three lactations of study traits that were calculated by BAGS and AI-REML method under single trait model are shown in Table 3. In Table 4, the repeatability model output of these methods is included. Some of posterior distributions of residual and additive genetic variances of different traits have been shown in Fig. 1.

Results of single trait analyses indicated that residual variance, additive genetic variance and heritability of first lactations were higher in comparison with other lactations. This might be caused by factors that do not influence the different lactations, such as management systems or environmental effects. The range of heritability in this study was obtained at 0.10 to 0.29. In both methods and in comparison to other lactations and traits the highest heritability value was for protein percentage in first lactation in contrast to the lowest value of heritability for fat yield in third lactation.

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|c|c|c|c|}
\hline
Total cycle & ME & MY & ME & MY \\
\hline
Burning Periods & 100000 & ME & MY \\
\hline
\sigma^2_u & 532980 & 532956 & 533150 & 403420 & 403403 & 403547 \\
\sigma^2_e & 1230274 & 1230284 & 1230143 & 926261 & 926267 & 926163 \\
\sigma^2_g & 532503 & 532440 & 532629 & 403016 & 403015 & 403154 \\
\hline
Total cycle & 150000 & ME & MY \\
\hline
\sigma^2_u & 1230621 & 1230657 & 1230519 & 926523 & 926548 & 926447 \\
\sigma^2_e & 531551 & 531357 & 531371 & 402344 & 402201 & 402207 \\
\sigma^2_g & 1231296 & 1231418 & 1231401 & 927030 & 927120 & 927110 \\
\hline
\end{tabular}
\caption{Variance components estimation for different values of Gibbs sampling parameters}
\end{table}

The heritability values obtained by repeatability model were relatively higher than single traits model, which probably is due to permanent environmental effects; similar to the results obtained by other studies including Palacios et al. (2007). These results showed that the genetic correlations among lactations were not equal to one. In this study repeatability of MY trait was higher than FY and PY traits in both AI-REML and BAGS methods.

According to a previous report (Boldman and Freeman, 1990), genetic parameter estimates may change with production levels, phenotypic variance and methods of estimation. However, in this research, estimated genetic parameters are similar to other results from the literature. Meyer (1984) reported heritabilities of 0.34, 0.35 and 0.28 of milk yield; 0.32, 0.33, 0.23 of fat yield and 0.24, 0.29, 0.12 of protein yield in 1st, 2nd and 3rd lactation, respectively. Dedkova and Wolf (2001) reported heritability of 0.30, 0.28 and 0.30 of milk yield; 0.24, 0.25 and 0.25 of fat yield and 0.25, 0.25 and 0.27 of protein yield in 1st, 2nd and 3rd lactation, respectively. AL-Seaf et al. (2007) obtained heritability of milk yield in Holstein dairy cows of 0.18, 0.18 and 0.14 for first three lactations, respectively. Dematawewa and Berger (1998) applied a repeatability model for first three lactation records and for milk yield, fat yield and protein yield traits of Holstein dairy cows and reported the value of these trait heritability and repeatability as 0.196, 0.177 and 0.179; 0.424, 0.410 and 0.412, respectively.

Generally, relative values of the estimated heritabilities in this study coincide with other reports such as Albuquerque et al. (1996) and Palacios et al. (2007), which concluded that the heritability of milk production traits in the first lactation was higher than of later lactations.

Conclusion: The results presented here show that the estimates of variance component and genetic parameter in BAGS method were higher than those by AI-REML method, for all traits and models. Additionally, estimates...
Table 3: Estimated parameters for first three lactations of different traits calculated by BAGS and AI-REML method under single trait model

<table>
<thead>
<tr>
<th>Traits</th>
<th>( h^2 )</th>
<th>( \hat{\sigma}^2_u )</th>
<th>( \hat{\sigma}^2_e )</th>
<th>( \hat{\sigma}^2_u )</th>
<th>( h^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>MY1</td>
<td>0.26</td>
<td>977092±8845</td>
<td>348320±11640</td>
<td>977092±8845</td>
<td>0.26</td>
</tr>
<tr>
<td>MY2</td>
<td>0.17</td>
<td>947443±9114</td>
<td>379950±20922</td>
<td>1856300±17829</td>
<td>0.17</td>
</tr>
<tr>
<td>MY3</td>
<td>0.29</td>
<td>524772±32954</td>
<td>357780±31318</td>
<td>2456000±29621</td>
<td>0.17</td>
</tr>
</tbody>
</table>

**AI-REML**

<table>
<thead>
<tr>
<th>Traits</th>
<th>( h^2 )</th>
<th>( \hat{\sigma}^2_u )</th>
<th>( \hat{\sigma}^2_e )</th>
<th>( \hat{\sigma}^2_u )</th>
<th>( h^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>MY1</td>
<td>0.29</td>
<td>386.19±12</td>
<td>581.75±15</td>
<td>550680±12941</td>
<td>0.39</td>
</tr>
<tr>
<td>MY2</td>
<td>0.26</td>
<td>305.65±9</td>
<td>207.44±12</td>
<td>305.65±9</td>
<td>0.34</td>
</tr>
<tr>
<td>MY3</td>
<td>0.29</td>
<td>1016.80±5</td>
<td>1526±6</td>
<td>1016.80±5</td>
<td>0.43</td>
</tr>
</tbody>
</table>

** качества параметров для первых трех лактаций различных признаков, рассчитанных методом BAGS и AI-REML при одном признаке**

**Fig. 1:** ПостерIOR распределений аддитивной и остаточной вариации для MY1, FY1 и PY1.

Table 4: Estimated values of variance components, genetic parameters and standard errors for first three lactations of different traits calculated by BAGS and AI-REML methods under repeatability model

<table>
<thead>
<tr>
<th>Traits</th>
<th>( \hat{\sigma}^2_u )</th>
<th>( \hat{\sigma}^2_{pe} )</th>
<th>( \hat{\sigma}^2_e )</th>
<th>( h^2 )</th>
<th>( r )</th>
</tr>
</thead>
<tbody>
<tr>
<td>MY</td>
<td>60998±11004</td>
<td>32050±5146</td>
<td>120545±3546</td>
<td>0.29</td>
<td>0.44</td>
</tr>
<tr>
<td>FY</td>
<td>598.67±6</td>
<td>207.44±12</td>
<td>1526±6</td>
<td>0.26</td>
<td>0.35</td>
</tr>
<tr>
<td>PY</td>
<td>444.73±9</td>
<td>305.65±9</td>
<td>1016.80±5</td>
<td>0.25</td>
<td>0.43</td>
</tr>
</tbody>
</table>

**References**