BAYESIAN INFERENCE IN SIRE EVALUATION IN ANIMAL BREEDING

E. Kaya Başar[1], E. Karaman[1], M.Z. Firat[1]

Second Bayesian Young Statisticians Meeting (BAYSM 2014)
Vienna, September 18-19, 2014

[1]Department of Animal Science, Akdeniz University, TURKEY.
ebrukaya@akdeniz.edu.tr, emrekaraman@akdeniz.edu.tr,
mzfirat@akdeniz.edu.tr

Abstract

In this study, it was aimed to provide a short description of how Bayesian inference is used in animal breeding for sire evaluation. For this purpose, total milk yields of 4787 cows belonging to 580 sires were analyzed through a linear mixed effects model. Bayesian analyses were done through the MCMCglmm package of R software. Beside the estimates of fixed effects and heritability of the total milk yield, a graphical representation of running means of breeding values was presented.

Keywords: Total milk yield, lactation, breeding value.

1 Introduction

In animal breeding, genetic evaluation for dairy sires and cows for production traits has depended for many years on the analysis of total milk yield (TMY) [2]. Genetic and environmental factors together affect the amount of milk that can be produced by a cow during lactation. These include age at calving, days of lactation for first test, pedigree status and Holstein proportion etc. [2]. Estimation of these factors is crucial for an efficient management and for making selection decisions through the estimation of breeding values accurately.

For many years, Bayesian method of estimation of model parameters in linear mixed effects models (LMM) is being used by researchers from various disciplines including animal science [7], [8], [5], [2], [1], [4]. This study aims to present how Bayesian methodology is used to estimate the factors affecting the TMY, and its use in sire evaluation.

2 Material and Methods

The structure of the data set is displayed in Table 1. The total milk yield records of Holstein-Friesian cows were used in this study. The sires having less than three daughters were discarded from the full data set. The resulting data set consisted of records from 4787 daughters of 580 sires. The number of daughters per sire varied between 3 and 31; and the average number of records per sire was 8.25. The number of herd-year-month (HYM) categories was 2183 and the average number of records per herd-year-month was 2.19 with a maximum number of records of 15. The milk yields were expressed in kg.

Table 1. The structure of the data

Number of sires	580
Total number of daughters	4787
Number of hym categories	2183
Mean Holstein proportion	0.41
Pedigree daughters	3132
Non-pedigree daughters	1655
Mean number of daughters per hym	2.19
Mean number of daughters per sire	8.25

For the total milk yield, following mixed effects model was used:

$$\mathbf{Y} = \mathbf{H}\boldsymbol{\alpha} + \mathbf{C}\boldsymbol{\beta} + \mathbf{D}\mathbf{s} + \mathbf{e}$$

where $\mathbf{Y}$, $\boldsymbol{\alpha}$,$\boldsymbol{\beta}$, $\mathbf{s}$ and $\mathbf{e}$ represent matrices of TMY $(N \times 1)$, herd-year-month effects $(g \times 1)$, regression coefficients $(c \times 1)$, sire effects $(s \times 1)$ and random error effects $(N \times 1)$, respectively. $\mathbf{H}$ and $\mathbf{D}$ denote design matrices relating the herd-year-month group and sire; and contains the values of covariates that are age at calving (AC), pedigree status (PS), Holstein proportion (HP) and days of lactation for first test (DL).

Following prior distributions were assumed for model parameters:

$$f\left(\boldsymbol{\alpha}, \boldsymbol{\beta}\right) \sim sabit$$

$$f\left(s\right) \propto \left(\sigma_s^2\right)^{-\frac{1}{2}s} \exp\left(-\frac{\mathbf{s}'\mathbf{s}}{2\sigma_s^2}\right)$$

$$f\left(\sigma_s^2 \mid \nu_s, s_s^2\right) \propto \left(\sigma_s^2\right)^{-\frac{1}{2}(\nu_s+2)} \exp\left(-\frac{\nu_s s_s^2}{2\sigma_s^2}\right)$$

$$f\left(\sigma_e^2 \mid \nu_e, s_e^2\right) \propto \left(\sigma_e^2\right)^{-\frac{1}{2}(\nu_e+2)} \exp\left(-\frac{\nu_e s_e^2}{2\sigma_e^2}\right)$$

The full conditional posterior distributions of $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $\mathbf{s}$, σ_s^2 and σ_e^2 which are obtained from the joint posterior probability density function are given, respectively, as follows:

$$f\left(\boldsymbol{\alpha} \mid \boldsymbol{\beta}, \mathbf{s}, \sigma_s^2, \sigma_e^2, \mathbf{y}\right) \sim N\left(\left(\mathbf{H}'\mathbf{H}\right)^{-1} \mathbf{H}'(\mathbf{y} - \mathbf{C}\boldsymbol{\beta} - \mathbf{D}\mathbf{s}), \left(\mathbf{H}'\mathbf{H}\right)^{-1} \sigma_e^2\right)$$

$$f\left(\boldsymbol{\beta} \mid \boldsymbol{\alpha}, \mathbf{s}, \sigma_s^2, \sigma_e^2, \mathbf{y}\right) \sim N\left(\left(\mathbf{C}'\mathbf{C}\right)^{-1} \mathbf{C}'(\mathbf{y} - \mathbf{H}\boldsymbol{\alpha} - \mathbf{D}\mathbf{s}), \left(\mathbf{C}'\mathbf{C}\right)^{-1} \sigma_e^2\right)$$

$$f\left(\mathbf{s} \mid \boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma_s^2, \sigma_e^2, \mathbf{y}\right) \sim N\left(\left(\mathbf{D}'\mathbf{D} + \mathbf{I}_s \frac{\sigma_e^2}{\sigma_s^2}\right)^{-1} \mathbf{D}'(\mathbf{y} - \mathbf{C}\boldsymbol{\beta} - \mathbf{D}\mathbf{s}), \left(\mathbf{D}'\mathbf{D} + \mathbf{I}_s \frac{\sigma_e^2}{\sigma_s^2}\right)^{-1} \sigma_e^2\right)$$

$$f\left(\sigma_s^2 \mid \boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{s}, \sigma_e^2, \mathbf{y}\right) \sim IG\left(\frac{\mathbf{s}+\nu_s}{2}, \frac{\mathbf{s}'\mathbf{s}+\nu_s s_s^2}{2}\right)$$

$$f\left(\sigma_e^2 \mid \boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{s}, \sigma_s^2, \mathbf{y}\right) \sim IG\left(\frac{N+\nu_e}{2}, \frac{(\mathbf{y}-\mathbf{H}\boldsymbol{\alpha}-\mathbf{C}\boldsymbol{\beta}-\mathbf{D}\mathbf{s})'(\mathbf{y}-\mathbf{H}\boldsymbol{\alpha}-\mathbf{C}\boldsymbol{\beta}-\mathbf{D}\mathbf{s})+\nu_e s_e^2}{2}\right)$$

Posterior estimates of the expected breeding values for the TMY trait were calculated using the formulae given as,

$$E\left(A_i \mid \theta, x_i\right) = 2\sigma_s^2 \left(n_i\sigma_s^2 + \sigma_e^2\right)^{-1} \sum \left(x_{ij} - \alpha_{h(ij)} - \boldsymbol{\beta}' c_{ij}\right)$$

A_i denotes the breeding value for sire i measured relative to its expectation without selection.

Analyses were performed using the MCMCglmm package [4] of R program [7]. A single chain of size 500,000 iterations was run. The initial 50,000 iterations were discarded as a burn-in period and every 250th iteration was recorded. In total, 1,800 samples were used to estimate the features of marginal distributions.

3 Results

Results obtained from fitting a LMM to the TMY data using Bayesian method of estimation are given in Table 2. Figure 1 provides a graphical representation of the change in average breeding values estimated through Bayesian methodology. It can be seen from this figure that top 50 sires have a breeding value of about 150kg on average. An increase in the number of sires to be selected results in a decrease in the average gain.

Table 2. Estimations of Bayesian method

Gibbs			
Mean	2.5%	97.5%	Sample
30.55	19.78	41.49	1800
3.88	0.56	7.21	1800
-110.4	-261.04	23.68	1800
175.78	-12.95	377.08	1800
41720	7371	75682	1800
430858	406442	457339	1800

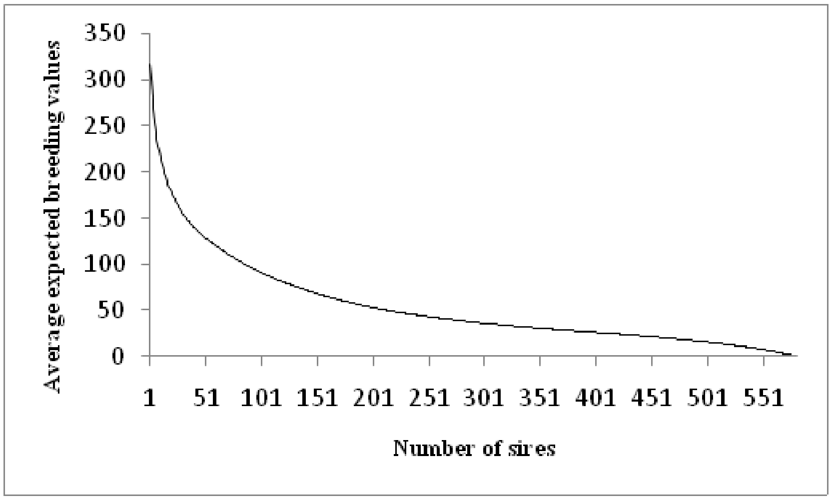

Figure 1: Plot of average posterior expected breeding values against the number of sires

References

[1] Blasco, A., Miriam Piles, M., Varona, L. (2003). A Bayesian analysis of the effect of selection for growth rate on growth curves in rabbits. Genet. Sel. Evol., 35: 21-41.

[2] Firat, M.Z.(2001). Sürü yıl-ay etkilerini şansa bağlı varsayan dengesiz karışık bir modelde kontrol günü süt verimlerinin bayesian analizi. Turk. J. Vet. Anim. Sci., 25, 327-333.

[3] Firat, M.Z., Theobald, C.M., Thompson, R. (1997). Univariate analysis of test day milk yields of British Holstein Friesian heifers using Gibbs sampling. Acta Agriculturae Scandinavica, Section A–Animal Science. 47: 213-220.

[4] Hadfield JD (2010). "MCMC methods for Multi-response Generalised Linear Mixed Models: The MCMCglmm R Package." Journal of Statistical Software,33(2), 1-22.

[5] Kumar, S., Rao, A.R., Bhatia, V.K. (2004). Bayesian estimation of heritability in animal breeding experiments under 2-way nested classification. J. Ind. Soc. Agril. Statist., 58(3): 352-362.

[6] Lee, C., Wang, C.D. (2001). Bayesian inference on variance components using gibbs sampling with various priors. Asian-Aust. J. Anim. Sci., 14(8): 1051-1056.

[7] R Development Core Team R (2010). A language and environment for statistical computing, reference index version 2.12.1. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

[8] Su, G., Sorensen, P., Sorensen, D. (1997). Inferences about variance components and selection response for body weight in chickens. Genet. Sel. Evol., 29: 413-425.

[9] Theobald, C.M., Firat, M.Z., Thompson, R. (1997). Gibbs sampling, adaptive rejection sampling and robustness to prior specification for a mixed linear model. Genet. Sel. Evol., 29: 57-72.