BAYESIAN MODEL SELECTION: AN APPLICATION TO GENETIC EVALUATION OF THE LATXA DAIRY SHEEP

A. Legarra\textsuperscript{1}, P. López-Romero\textsuperscript{2} and E. Ugarte\textsuperscript{1}

\textsuperscript{1}NEIKER A.B., Apdo. 46, 01080 Vitoria-Gasteiz, Spain
\textsuperscript{2}INIA, 28040 Madrid, Spain

INTRODUCTION
Model selection is a frequent concern in animal breeding (Bidanel, 1998; Thompson, 2001). Nevertheless, there is not, and probably there will not be, a method “of choice” in this sense. Our interest in model choice is motivated by the model that has been used since 1990 for the Latxa genetic evaluation based on the standardised 120 days milk yield (MY) (Ugarte \textit{et al.}, 1996). This model includes two contemporary groups. Despite the model seems biologically correct, it is somewhat counter-intuitive, and it can lead to confusion. For this reason, we have tested this model vs. two simpler models. Two strains exist in the Latxa breed: Black-faced and Blond-faced, following separate breeding programs. Although both strains were studied, only the Blond-faced results are presented in this work. The study for Black-faced Latxa gave similar results.

A Bayesian framework was followed because of its simplicity and more appealing results. In this context one can affirm that one model is \(x\) times more likely than the other. Several criteria were used, and most of them have been formerly used in animal breeding. They have revealed easy to program within a typical Gibbs sampling scheme.

MATERIAL
Models and data. An animal model for repeated measures was used: \(y = \mathbf{b} + \mathbf{a} + \mathbf{p} + e\), where \(\mathbf{b}\), \(\mathbf{a}\) and \(\mathbf{p}\) are respectively systematic (“fixed”), additive and individual random permanent effects, following usual assumptions. Models differed in the structures of \(\mathbf{b}\) and \(\mathbf{X}\).

The model currently used in the BLUP genetic evaluation is as follows: \(\text{MY} = \text{flock-month of lambing-age at parity-number of parity (3,174 levels) + flock-year (1,440 levels) + number of alive born lambs (3 levels)+ interval between lambing and first milk recording (8 levels) + additive genetic effect + permanent random effect + residual. This model (FMAP for short) is thought to be biologically correct, because management practices are such that an interaction between feeding, age, and month of lambing could exist. There are two main problems in the model: (1) the first effect is regarded to be the same between different years, and this is a rather strong assumption, and (2) some contemporary groups can be very small, which could lead to sampling errors in their estimates.

The second model (FYM) included the systematic effects flock-year-month of lambing (2,856 levels) + age at parity-number of parity (9 levels) + other terms. Thus, it is assumed that no interaction between age-parity and contemporary group exists, but an interaction between year and season was added. The third model (FY) is a simplistic one and includes the effects flock-year
(1,440 levels) + month of lambing (6 levels) + age at parity-number of parity (9 levels) + other terms. This was assumed not to be correct, but it was included as a reference.

As genealogy is not fully known (17% animals in data with both sire and dam known, 33% with only dam known) all models included 18 genetic groups, based on selection path and year of birth. All data available for MY (average = 120.61 l, std.dev. = 49.89) in the Latxa milk recording database from 1984 to 2000 were used. This correspond to 105,206 data, 42,433 animals in data and 45,233 animals in pedigree, including 338 rams with an average of 22 daughters with productive register, and 11,694 ewes in data with daughter(s) in data.

Criteria for selecting models. There is a huge literature on Bayesian model selection. We have used two main criteria that have been formerly used in animal breeding: Bayes Factors (Sorensen, et al., 2000; López-Romero et al., 2001) and predictive distributions (Varona et al., 1997; Rekaya, 1997). None of the model selection criteria is claimed to be the best one, moreover, the use of more than one is usually recommended.

Bayes Factors. (BF) This is the classical Bayesian tool for model choice. It is the ratio of the marginal densities of the data (MDD) between two competing models.

\[ BF = \frac{p(y \mid Model 1)}{p(y \mid Model 2)} \]

If prior probabilities for each model are assumed to be the same, the BF is the posterior odds between models, \( p(\text{Model 1} \mid \mathbf{y}) / p(\text{Model 2} \mid \mathbf{y}) \), which is a very intuitive way of choosing among alternative models.

Predictive distributions. From a leave-one-out cross-validation approach it has been suggested to obtain the distribution of one data \( y_r \) given the rest of data \( \mathbf{y} \setminus \mathbf{r} \). The model leading to the “best” prediction of \( y_r \) is the “best” one. This is measured through the so-called checking functions which measure difference between prediction \( Y_r \mid \mathbf{y} \setminus \mathbf{r} \) and observation \( y_r \) with \( r = 1, 2 \ldots \) number of data.

Consider the error \( e_r = y_r - Y_r \). The first checking function considers the expectation of the squared error, that is, a measure of the accuracy: \( d_1 = E(e_r^2) \); the second one considers a probabilistic measure of the existence of a systematic trend to over- or under-estimate, a measure of bias: \( d_2 = E[p(e_r < 0) - 0.5] \). Both functions should tend to 0.

Methods. All these criteria can be implemented within a usual genetic parameter estimation using Gibbs sampling (See Sorensen, 2000; Gelfand,1996; Gelfand et al., 1992; Rekaya, 1997). However, some tricks are needed which will not be presented here for the sake of simplicity. Separate Gibbs samplings are needed for each model. Vague, but proper, priors were used for all parameters. The same analyses were also carried out considering “fixed” effects as random with a (fixed) variance of 500, but conclusions were the same and they will not be presented here. For each model, 5 different chains were run, using different random seeds. 105,000 iterations were run for each chain, and the first 5,000 were discarded as burn-in.
RESULTS.

**Variance components estimates.** They were obtained as a sub-product, and are presented in table 1. Estimate for residual variance is a measure of the model fit. This is known to favour more complex models. Nevertheless, FYM, which is an intermediate model in complexity, gives the lowest estimate. Estimates are similar to those previously found (Ugarte et al., 1996).

<table>
<thead>
<tr>
<th></th>
<th>FMAP</th>
<th>FYM</th>
<th>FY</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean 95% HPD</td>
<td>Mean 95% HPD</td>
<td>Mean 95% HPD</td>
</tr>
<tr>
<td>var(a)</td>
<td>312 289-336</td>
<td>310 287-334</td>
<td>302 279-327</td>
</tr>
<tr>
<td>var(p)</td>
<td>231 205-256</td>
<td>234 209-259</td>
<td>257 230-282</td>
</tr>
<tr>
<td>var(e)</td>
<td>851 842-861</td>
<td>830 821-839</td>
<td>876 866-886</td>
</tr>
<tr>
<td>h²</td>
<td>0.22 0.21-0.24</td>
<td>0.23 0.21-0.24</td>
<td>0.21 0.19-0.23</td>
</tr>
</tbody>
</table>

Results for the different model choice criteria considered are shown in table 2. As all chains led to very similar results, only results for one chain in each model are shown here.

**Bayes Factors.** The value of the MDD is very low. The reason is that there is a great amount of data and parameters to be fitted in every model. The Log Bayes Factor of two different models is: Log BF = (log MDD model 1 – log MDD model 2). BF indicates FYM as the best model.

**Checking functions.** Checking functions (d1 and d2) led to different conclusions. Following d1 (accuracy in predicting y_r given y-a), the best model is FYM. Following d2 (systematic error in the same prediction), FMAP is the best one. Similar results were observed by Frey et al. (1997), using cross-validation deleting small groups of data: those models well fitted for bias were not those well fitted for accuracy.

<table>
<thead>
<tr>
<th></th>
<th>FMAP</th>
<th>FYM</th>
<th>FY</th>
</tr>
</thead>
<tbody>
<tr>
<td>log(MDD)</td>
<td>-504,965</td>
<td>-503,610</td>
<td>-506,451</td>
</tr>
<tr>
<td>d1</td>
<td>1,503</td>
<td>1,412</td>
<td>1,447</td>
</tr>
<tr>
<td>d2</td>
<td>0.1396</td>
<td>0.1408</td>
<td>0.1436</td>
</tr>
</tbody>
</table>

**Stability.** Due to numerical reasons, the estimate of the MDD is completely dominated by the least likely sample of parameters, \( \theta' = \left( b' a' p' \sigma_a^2 \sigma_p^2 \sigma_e^2 \right) \) obtained along the Gibbs sampling process. However, this does not imply that MDD estimate is unstable. A trace plot of the estimated log MDD for different models (2 chains for each model) along the Gibbs sampling is shown in figure 1. From this figure it can be concluded that, after some iterations, estimates are stable, and that different chains in the same model produce the same result.
CONCLUSION
Bayesian model selection is easy to implement in classical animal breeding mixed models.
The best-fitted model was FYM according to two of three criteria. In spite of the high value of the
Bayes Factor against the rest of models, predictive distributions did not show such a strong
evidence for any of the models.

Figure 1. Plot of the estimates of the marginal density of the data along the Gibbs sampling.
Three different models (FMAP, FYM and FY) and two Gibbs sampling chains (1 and 2) for
each model are shown

ACKNOWLEDGEMENTS
Legarra and P. López-Romero thank INIA and MCYT for their scholarships. We thank L. Varona
and R. Rekaya for their help.

REFERENCES
J.M. Bernardo, A.P. Berger, A.P. Dawid, and A.F.M Smith, Oxford Univ. Press, Oxford,
UK
Gilks, S. Richardson and D.J. Spiegelhalter, Chapman & Hall, London, UK
Sorensen, D (1999) « Gibbs sampling in quantitative genetics ». Internal report no. 82. Danish
Institute of Animal Sciences, Foulum, Denmark.