1 Draft for Master Thesis Project: "Generalized Linear Mixed Models In Genetic Evaluations"

Selection of livestock animals based on predicted breeding values has led to tremendous improvements of product quality and of production efficiency in livestock production. The breeding value of an animal is an unobservable quantity which assesses the value of the genetic potential of a given animal. Because, breeding values cannot be observed, they must be predicted based on phenotypic observations of animals. The prediction of breeding values is based on Best Linear Unbiased Predictions (BLUP) of random effects in a linear mixed effects model. The linear mixed effects model explains observations as a linear function of fixed effects and random breeding values. Estimates of fixed effects and predictions of random breeding values are solutions of the so-called mixed model equations which are proposed by (Henderson 1982). The quality of the ranking of the predicted breeding values is assessed by comparing average phenotypic performances of offspring from top-ranking and bottom-ranking parents. Average phenotypic performances of offspring from top-ranking parents must be significantly higher than average phenotypic performances of offspring from bottom-ranking parents.

Strictly speaking, the solutions of the mixed model equations are only valid, if the observations and the breeding values both follow multivariate normal distributions. There are references such as (Negussie, Strandén, and Mäntysaari 2008) which show that the predicted breeding values are also valid as ranking criterion for parents when phenotypes are not normally distributed. Such references are used as justification to use the linear mixed effects model framework together with the solutions from the mixed model equations for traits that do clearly not follow a normal distribution. Extreme cases of such data are observations that show a binary distribution. From a livestock breeding point of view, it is not clear what the distribution of the predicted breeding values should look like and how such an extreme distribution of observations affect the ranking of breeding animals according to the predicted breeding values.

Recently, breeding organisations are more interested in the potential of improving their populations with respect to health traits or with respect to animal behavior. Examples of such traits for which Qualitas recently developed a routine evaluations process are

- Twin and multiple birth in cattle
- Early-life calf survival in dairy cattle
- Carcass conformation in beef cattle

The results of the above three evaluations showed some problems that were previously not found when using the same procedure for other traits.

1. Goodness of fit of the mixed linear effect model as assessed by criteria such as AIC and BIC was very low.
2. The standard deviation of the predicted breeding values was very low
3. As a consequence of 2, the standard errors of prediction for the predicted breeding values were very high.
4. The ranking of the animals according to the predicted breeding value has a low quality as measured by the top-bottom comparisons.

The aim of this project is to assess the benefit of using different types of models such as the generalized linear mixed model for genetic evaluations of trait response variables showing binary or categorical distributions. There are a few studies such as (Hoeschele et al. 1986), (Tempelman 1998), (Koenig et al. 2005) or (Vazquez et al. 2009) that have already tried such models. To the best of our knowledge no routine evaluations have been implemented using the class of generalized linear mixed models. Important pre-requisites for the implementation of a routine evaluation is to be able to completely characterize the properties of a genetic evaluation using generalized linear mixed model. Furthermore, it is also important to evaluate different possible software solutions for a routine evaluation pipeline.

The benefits of using generalized linear mixed models is evaluated using a simulated dataset which has approximately the same structure as the datasets that are analysed during the routine evaluations. In a second phase of the project, datasets that come from real-world genetic evaluations should als be evaluated.

References

Bates, Douglas, Martin Mächler, Bolker Benjamin M., and Steven C. Walker. 2015. "Fitting Linear Mixed-Effects Models Using lme4." *Journal of Statistical Software* 67 (1). https://doi.org/10.18637/jss.v067.i01.

Henderson, Charles R. 1982. "Analysis of Covariance in the Mixed Model: Higher-Level, Nonhomogeneous, and Random Regressions." *Biometrics* 38 (3): 623–40. https://doi.org/doi:10.2307/2530044.

Hoeschele, Ina, J. L. Foulley, J. J. Colleau, and D. Gianola. 1986. "Genetic evaluation for multiple binary responses." *Genetics Selection Evolution1* 18 (3): 299–320. https://doi.org/DOI:10.1186/1297-9686-18-3-299.

Koenig, S., A. R. Sharifi, H. Wentrot, D. Landmann, M. Eise, and H. Simianer. 2005. "Genetic parameters of claw and foot disorders estimated with logistic models." *Journal of Dairy Science* 88 (9). Elsevier: 3316–25. https://doi.org/10.3168/jds.S0022-0302(05)73015-0.

Negussie, Enyew, Ismo Strandén, and Esa A. Mäntysaari. 2008. "Genetic analysis of liability to clinical mastitis, with somatic cell score and production traits using bivariate threshold-linear and linear-linear models." *Livestock Science* 117 (1): 52–59. https://doi.org/10.1016/j.livsci.2007.11.009.

Tempelman, Robert J. 1998. "Generalized Linear Mixed Models in Dairy Cattle Breeding." *Journal of Dairy Science* 81 (5): 1428–44. https://doi.org/10.3168/jds.S0022-0302(98)75707-8.

Vazquez, A I, D Gianola, D Bates, K A Weigel, and B Heringstad. 2009. "Assessment of Poisson, logit, and linear models for genetic analysis of clinical mastitis in Norwegian Red cows." *Journal of Dairy Science* 92 (2). Elsevier: 739–48. https://doi.org/10.3168/jds.2008-1325.