

Integrating Phenotypic Measurement Models With Animal Models

S.M. van den Berg^{*}, F. Fikse[†], P. Arvelius[†], C.A.W. Glas^{*}, E. Strandberg[†]

Introduction

Behaviour is becoming an important phenotype in animal breeding (Kanis, De Greef, Hiemstra et al. (2005)). However, compared to such phenotypes as milk yield, behaviour is usually more difficult to quantify reliably. Often, a measure is constructed based on a number of indices. For instance, dog behaviour can be studied in several test situations in which the dog's responses are coded on a scale or as present/absent. Such indices are generally added or averaged to form a raw score that is then analysed using traditional quantitative methods, treating the scores as interval measures.

For a number of reasons, this approach is suboptimal. For one, there is always an amount of measurement error that is not taken into account which ends up in the environmental variance component. When a score is based on a small number of situations, the error variance component can be considerable. Distinguishing between measurement error variance and 'true' environmental variance helps identifying measurement problems. Second, there are distributional problems, in that the total or averaged scores are often skewed in practice. Such distributional violations of the standard animal model can lead to serious bias in inference regarding heritability (Van den Berg, Glas, and Boomsma (2007)).

Such measure-related problems can be solved when the phenotypic measurement process is explicitly modelled when estimating heritability. We propose to use measurement models from modern test theory and integrate them with animal models. Such measurement models are based on Item Response Theory (IRT). In IRT, observed behaviours are modelled as functions of a latent variable (the trait in question) and severity of behaviour. An animal can score high or low on the latent variable, and a response can refer to common behaviour (a dog barking to a stranger) or to less common behaviour (biting the owner). In an IRT model, the probability of a particular behaviour in one situation (say, a dog barking to a stranger), can be modelled as a logistic function of the animal's (unobserved) score on the latent trait, and unknown situation parameters. With a sufficient number of dogs and test situations, the animal latent scores and situation parameters can be estimated. Moreover, an IRT approach helps improving measurement tools, for instance helping to decide whether to include more common or less common behaviours in the testing procedure.

A standard analysis of estimates of latent score based on IRT would yield approximately the same results as an analysis on total scores. However, if instead the genetic analysis is based on the latent trait, while concurrently estimating the phenotypic measurement model, the heritability coefficient is automatically corrected for measurement error that is due to the

^{*} University of Twente, Institute for Behavioural Research, P.O. Box 217, 7500 AE Enschede, The Netherlands.

[†] Swedish University of Agricultural Sciences, Animal Breeding & Genetics, Box 7023, 750 07 Uppsala, Sweden.

measurement instrument itself (i.e., its internal consistency). In other words, by taking into account the probabilistic nature of behavioural measurement, one can estimate the heritability as if the trait were estimated with a perfect measurement instrument (i.e., an infinite amount of test situations). Such an integration of genetic models and phenotypic measurement models has been developed previously for human twin families (Van den Berg, Glas and Boomsma (2007); Van Leeuwen, Van den Berg, and Boomsma (2008)). Here we present a method for animal pedigree data, and apply it to fearfulness in a dogs.

Material and methods

Sample. Data come from the Dog Mentality Assessment (DMA; Svartberg & Forkman 2002). For 944 Irish soft-coated wheaten terriers there were complete behavioural data on a 7-item subscale for fearfulness (items 17, 19[‡], 20, 22, 23[‡], 24 and 33). These dogs belonged to 514 full sib families and were progeny of 288 fathers and 377 mothers. Of these parents, 63 fathers and 95 mothers were themselves tested. For more than 55% of the tested dogs, at least one parent was tested. Untested relatives (1448 dogs) were included to facilitate estimation of genetic parameters. Maximum inbreeding coefficient was 0.25 (two dogs), with 0.0009 on average.

Statistical analyses. For the introduction and illustration of IRT models, a simple IRT model for dichotomous (0,1) data was applied. To this end, each original item, consisting of 5 response categories, was dichotomized using median splits, such that a dog either had a score of ‘0’ or ‘1’, reflecting ‘below median’ and ‘median or above’, respectively. Let $Y_{ij} = y_{ij} \in \{0, 1\}$ be the dichotomous random variable. The probability of outcome $Y_{ij} = 1$ was modelled as

$$p(Y_{ij} = 1 | \theta_i, \beta_j) = \frac{\exp(\theta_i - \beta_j)}{1 + \exp(\theta_i - \beta_j)},$$

where θ_i is the latent trait score for dog i and β_j is the location of item j . Parameters of interest are the genetic and environmental variance components σ_A^2 and σ_E^2 for parameter θ . In order to estimate these variance components, θ_i was modelled as a function of parental genotypes $a_{Sire(i)}$ and $a_{Dam(i)}$, a Mendelian sampling term with variance σ_{MS}^2 , that in turn depends on parental inbreeding coefficients F_{sire} and F_{dam} , such that $\theta_i \sim N(a_i, \sigma_E^2)$, where

$$a_i \sim N\left(\frac{1}{2}(a_{Sire(i)} + a_{Dam(i)}), \sigma_{MS}^2\right)$$

$$\sigma_{MS}^2 = \frac{1}{2}\left(1 - \frac{1}{2}(F_{Sire(i)} + F_{Dam(i)})\right)\sigma_A^2.$$

Breeding values for founders were modelled as normally distributed with expectation 0 and variance σ_A^2 and with an inbreeding coefficient of 0. The model was implemented in the BUGS software package (Lunn, Thomas, Best et al. (2000); Damgaard (2007)) with priors $\sigma_A^2 \sim InvGamma(0.01, 0.01)$, $\sigma_E^2 \sim InvGamma(0.01, 0.01)$ and $\beta \sim MVN(\mathbf{0}, 100\mathbf{I})$. This

[‡] These items were reverse-coded.

method was compared with an analysis without an IRT model, modeling the average score on the dichotomised seven items directly.

Results and discussion

Table 1 shows genetic parameter estimates. The parameters for the IRT approach are based on 2 independent MCMC chains of 1586 samples each, after a burn-in of 1000. The analysis took 15.8 hours on one 3 Ghz processor. Figure 1 shows rapid convergence, but slow movement through the posterior space. The item parameters are highly correlated in the posterior space (all correlations 0.84-0.85) which suggests that the average score on the latent trait, identified through the expected genetic value of 0 for founder animals, is ill-defined. With IRT the heritability estimate and its variance are larger than in the traditional approach. However, the difference in heritability estimates with and without IRT is not as dramatic as observed earlier in a twin study (Van den Berg, van Leeuwen, and Boomsma (2007)).

These results illustrate that due to imperfect measurement instruments, heritability estimates for actual behavioural traits are generally biased downwards, as are the posterior variances of the heritability. The probabilistic nature of measurement (i.e., observing behaviour only in a limited number of situations) can be taken into account by integrating IRT models in the statistical modeling. In this approach the heritability estimate reflects the coefficient that *could* be obtained if the measurement were perfect (i.e., with an infinite amount of situations observed, cf. Van den Berg, van Leeuwen, and Boomsma (2007)). In a breeding context however, the heritability based on total scores would be most useful, as one can only use such total scores for selection: that heritability estimate is relevant for predicting genetic gain. Nevertheless, the IRT approach presented here is able to show to what extent genetic progress can be improved if measurement tools were improved. For example, a preliminary IRT analysis on the terrier data showed that the measurement tool could be improved by including more common fear-related behaviours, as most terriers scored rather low in all test situations, leading to poor discrimination between them.

It should be noted that rigorous testing of the method is still required. Next, the method can be easily extended to include systematic environmental effects of, for instance, sex and age, and to apply more complicated IRT models. For instance, several IRT models exist for polytomous data (ordinal or categorical data), such as we have in the DMA. In addition, IRT models allow for the inclusion of item discrimination parameters, which are the IRT analogues of factor loadings in traditional factor analysis.

Table 1: Posterior means and standard deviations of genetic parameters with and without an IRT model.

Parameter	Standard analysis	Analysis with IRT model
σ_E^2	0.055±0.004	0.855±0.137
σ_A^2	0.024±0.004	0.553±0.137
h^2	0.30±0.04	0.39±0.09

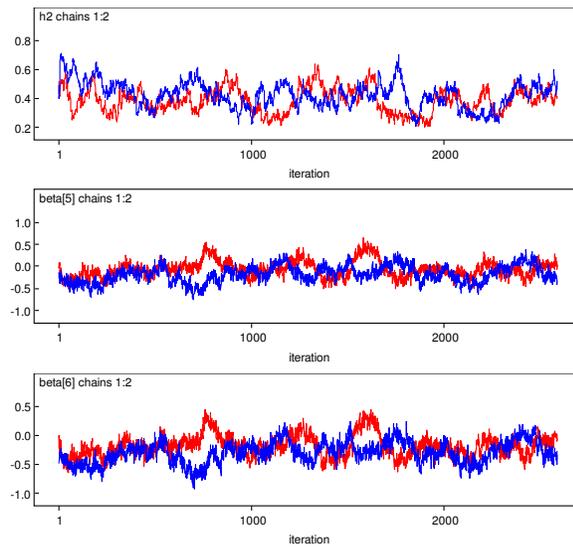


Figure 1: Posterior samples for the heritability coefficient and two item parameters.

Conclusion

The IRT approach presented here is able to show to what extent genetic progress can be improved if measurement tools were improved. If indeed a measure with a higher heritability seems worthwhile, Item Response Theory serves as a comprehensive framework for improving measurement.

References

- Damgaard, L. (2007). *J. Anim. Sci.*, 85:1363–1368.
- Kanis, E., De Greef, K.H., Hiemstra, A., et al. (2005). *J. Anim. Sci.*, 83:948–957.
- Lunn, D.J., Thomas, A., Best, N., et al. (2000). *Stat. Comp.*, 10:325–337.
- Svartberg, K. and Forkman, B. (2002). *Appl. Anim. Behav. Sci.*, 79:133–155.
- Van den Berg, S.M., Glas, C.A.W., and Boomsma, D.I. (2007). *Behav. Genet.*, 37:604–616.
- Van Leeuwen, M., Van den Berg, S.M., and Boomsma, D.I. (2008). *Learn. Ind. Differ.*, 18:76–88.