MIXED MODEL METHODOLOGY UNDER GENETIC MODELS WITH A SMALL NUMBER OF ADDITIVE AND NON-ADDITIVE LOCI

ASKO MÄKI-TANILA[1] and BRIAN W. KENNEDY

Department of Animal and Poultry Science
University of Guelph, Guelph
Ontario N1G 2W1, CANADA

SUMMARY

The general properties of linear predictors are discussed for genetic models with a small number of (non-)additive loci. The usefulness of mixed model approach is studied by simulations making allowance for selection. The BLUP estimates are satisfactory except when selection limit is reached and when an inadequate model is used under dominance. The ways to improve the estimation and utilization of dominance values are discussed.

INTRODUCTION

It may not be an exaggeration to suggest that every locus contributes in one way or another to any given quantitative trait, although the effects of most loci are vastly unequal, there being a small number of major loci and a large number of minor loci. Genes with considerable effects on a metric trait have been found, e.g. dwarf in poultry, halothane in pigs, and booroola in sheep.

We know that when the variation in a trait is governed by a large number of loci or alternatively by a locus with a large number of alleles without epistatic interaction and linkage disequilibrium between loci and when the environmental deviations are independent of the genotypic values and normally distributed, we end up having a normal phenotypic distribution. Hence linear methods are adequate. The question we would like to address is how do mixed model techniques (or, in general, procedures making linear assumptions about various components which might affect phenotype) behave when the genetic variation in a trait is due to a finite, or small, number of loci. The earlier studies (Robertson, 1977; Bulmer, 1980; Mäki-Tanila, 1982) suggest to deal with the answer under four different topics.

(i) In the presence of dominance an animal's breeding value or additive value and genotypic value are not linearly related. Largest departures from linearity are due to alleles which are almost completely recessive and very rare. If the allele with a smaller effect is recessive, the regression is curved upwards. Non-linearity vanishes naturally very quickly (proportionally to 1/n) as the number of loci (n) goes up. One aspect which interests us here with respect to mixed model technique is, does the inclusion of dominance effect in the linear model have any corrective effect.

---

[1] Present address: Dept. of Animal Breeding, Agricultural Research Centre, SF-31600 Jokioinen, FINLAND.

(ii) A more serious reason for non-linear relationship between additive value and phenotypic value is the departure of genotypic distribution from normality. In this case the observed response can be larger or smaller than is linearly predicted, i.e. linear predictors produce biassed values. We keep the earlier assumption about the normality of environmental deviations. Although the relationship between additive value and genotypic value is in the absence of dominance always linear, the linearity with respect to phenotypic value depends on the skewness of genotypic distribution. If allele frequencies are intermediate (around .5), linearity will hold while extreme frequencies lead to skewed distributions and non-linearity. For similar reasons dominance causes non-linearity. The genotypic skewness decreases proportionally to $1/\sqrt{n}$. The degree of curvature depends also on the magnitude of environmental variance or on the ratio of genotypic (σ_g^2) to phenotypic (σ_p^2) variance. Without going into details (c.f. Mäki-Tanila, 1982) we summarize the results in Table 1 (0 for complete or approximate linearity, + and - for positive and negative curvilinearity, respectively, and their number for the degree of curvature) for the case where there are few loci with a similar state of alleles. The frequency (q) and dominance (d, -/+ 1 = complete recessivity/dominance, 0 = complete additivity) is expressed in terms of the allele with a smaller effect.

Table 1

		σ_g^2 / σ_p^2		
q	d	1	.5	.1
.1	-1	0	--	---
	-.9	++	-	--
	-.7	+	-0	-
	0	0	-0	-0
	.7	-0	-0	-0
	.9	-0	-0	-0
	1	0	-0	-0
.5	-1	0	-0	-0
	-.9	0	-0	-0
	-.7	+0	-0	-0
	0	0	0	0
	.7	-0	-0	-0
	.9	0	-0	-0
	1	0	-0	-0

The largest deviations from linearity follow when the allele is rare and completely recessive and the proportion of genotypic variance is low.

(iii) Whatever genetic model we have (allowing for a very small n, dominance, epistasis, linkage), the additive value of an offspring is in terms of his parent's breeding values (p and m to subscript sire's and dam's value respectively) $a_p/2 + a_m/2 + e$ where e varies depending on the heterozygosity of the parents. If the genetic variation is mediated by a large number of loci, e is normally distributed with mean 0 and variance $\sigma_a^2(1-F)/2$ (σ_a^2 is the additive

variance and $\bar{F}$ is the mean of parents' inbreeding coefficients). Hence we can express the variance-covariance matrix of breeding values as $\underset{\sim}{T}\underset{\sim}{U}\underset{\sim}{T}'\sigma_a^2 = \underset{\sim}{A}\sigma_a^2$ (Thompson, 1977). $\underset{\sim}{A}$ is the relationship matrix, $\underset{\sim}{T}$ is a matrix describing the flow of genes from one generation to the next and $\underset{\sim}{U}$ is a diagonal matrix with an element $(1-F_i)/2$ for the ith animal.

If the source of genetic variation is a small number of loci, the variance of e is no longer independent of parents' breeding values (Bulmer, 1980). Although the heteroscedasticity of the regression is in most cases negligible, it is of interest to study the validity of the use of the relationship matrix in the mixed model context under extreme genetic models.

(iv) Finally, how will mixed model methods behave under the major activity in animal breeding, that is, under selection when the variation is due to a small number of loci. Sorensen and Kennedy (1984) have found that if we have a full knowledge of the relationship matrix, by averaging 'BLUP estimates' we get unbiassed estimates of selection response with smaller sampling errors than the straightforward (least-squares) phenotypic generation means.

SIMULATION AND STATISTICAL METHODS

The following parameters describe the genetic model: n = number of (unlinked) loci affecting the trait, d = degree of dominance such that -1, 0, and +1 represent complete recessivity, additivity, and dominance of the allele with a larger effect, q = initial frequency of the allele. The expected value of the additive variance at the generation 0 is made to equal 10 at each run. A normally distributed environmental deviation is added to genotypic value according to the initial ratio σ_g^2/σ_p^2. At each generation 2 males and 8 females are used to produce 8 offspring of both sexes. On the male side either random sampling or truncation selection upwards on phenotype is used. The selection is started at generation 1. In all runs simulation is carried over until the 3rd generation. The simulation consists of 100 replicates for each set of parameters.

The generation means were estimated as simple generation means from phenotypic values or starting from a mixed linear model. In the latter the phenotypic value (y_i) for an ith individual can be written (e.g. Henderson, 1985)

$$y_i = \mu + m_i + e_i$$

where u is the base population mean and m is the total genetic value of animal i, and e the residual. In matrix form the model is

$$\underset{\sim}{y} = \underset{\sim}{1}\mu + \underset{\sim}{I}\,\underset{\sim}{m} + \underset{\sim}{e}$$

$E(\underset{\sim}{m}) = 0$, $Var(\underset{\sim}{m}) = \underset{\sim}{A}\,\sigma_a^2 + \underset{\sim}{D}\,\sigma_d^2$ and it is assumed that $Var(\underset{\sim}{e}) = \underset{\sim}{I}\,\sigma_e^2$. In the analyses the true values for the variances are used.

When there is no inbreeding, the construction of $\underset{\sim}{D}$, the dominance relationship matrix, is fairly simple (e.g. Falconer, 1981). In our analyses $\underset{\sim}{D}$ was computed modifying the algorithm reported by Smith

and Allaire (1985). Let us have individuals X and Y of which, say, Y is older with a sire S. If we denote by subscripts m and p the genome halves originating from dam and sire, respectively, we can form a genomic table for these individuals from the rule illustrated by an example

$$P(X_m-Y_p) = (P(X_m-S_p) + P(X_m-S_m))/2$$

By definition the dominance relationship

$$P(X_p-Y_p)\ P(X_m-Y_m) + P(X_p-Y_m)\ P(X_m-Y_p).$$

RESULTS

Random mating. The results are summarized in Table 2 for n = 2 and $\sigma_g^2/\sigma_p^2 = .5$. When there is no dominance, the linear predictors give satisfactory estimates with smaller sampling errors. The results do not differ from the infinitesimal ones (Sorensen & Kennedy, 1984). In the case of dominance we obtain inbreeding depression; in all runs (including selection) the degree of inbreeding is .14 + .003 at the 3rd generation. The analyses based on models having only the additive effect grossly underestimate the depression, while its estimates are better when the dominance effect is also considered. In both cases the sampling errors are smaller than those of the simpler estimates.

Table 2. The genetic means and their LS and BLUP estimates for three generations of random mating. In MM(a) model the dominance effect is ignored while it is included in MM(a+d). n = 2; a) gen. 1 and b) gen. 3.

q	d	true mean	LS estimate	MM(a) est.	MM(a+d) est.
.5	0	a)-.11 (1.47)	-.09 (1.63)	-.11 (1.51)	
		b)-.25 (1.87)	-.08 (2.15)	-.11 (1.89)	
	1	.24 (1.41)	.27 (1.63)	.09 (1.51)	.13 (1.50)
		-.17 (2.14)	-.14 (2.45)	-.05 (2.05)	-.07 (2.14)
.9	0	-.06 (1.33)	-.07 (1.51)	-.16 (1.22)	
		1.41 (.84)	1.42 (1.22)	1.59 (.98)	
	1	-.07 (2.35)	-.20 (3.16)	-.42 (2.42)	-.32 (2.54)
		-.82 (4.09)	-1.04 (4.55)	-.59 (2.99)	-.79 (3.67)

Selection. With complete additivity we would expect from the earlier results (Table 1) that linear predictors would slightly over(under)estimate the response when the allele frequencies are high (low). The simulation results (Table 3) do support this, although the biases are fairly small. Had the selection lasted more generations the predicted response would have deviated more and not plateaued even after the selection limit had been reached (R. Fernando, personal communication). This bias is further augmented by dominance. In addition, when the recessive allele is very rare, most of the genetic variation is due to dominance, and because of

either larger or smaller than the predicted one (Hill, 1969). In Table 3 this is clearer for the case where q = .1 and d = -1 whilst the opposite case leads to fixation very quickly if the number of loci causing variation is very small (Table 4). The addition of dominance value makes a considerable improvement in predicting the total genetic value.

Table 3. The genetic means and their LS and BLUP estimates for three generations of selection. n = 2, σ_g^2/σ_p^2 = .5.

q	d	true mean	LS estimate	MM(a) est.	MM(a+d) est.
.1	-1	.13 (2.54)	.10 (3.29)	.69 (3.39)	.26 (2.96)
		3.36 (8.38)	3.38 (8.36)	2.05 (5.04)	3.00 (6.79)
	0	.21 (1.57)	.10 (1.93)	.26 (1.76)	
		3.03 (2.94)	3.05 (3.15)	2.95 (2.81)	
.5	-1	-.14 (1.32)	-.27 (1.64)	.14 (1.50)	.00 (1.48)
		2.79 (2.34)	2.81 (2.77)	2.52 (2.16)	2.65 (2.31)
	0	.19 (1.30)	.24 (1.54)	.17 (1.36)	
		2.30 (1.71)	2.19 (1.88)	2.26 (1.69)	
	1	-.13 (1.54)	-.27 (1.79)	-.23 (1.37)	-.27 (1.44)
		1.13 (1.34)	1.16 (1.83)	1.35 (1.45)	1.36 (1.53)
.9	0	-.06 (1.33)	-.07 (1.51)	-.16 (1.51)	
		1.41 (.84)	1.42 (1.22)	1.59 (1.22)	
	1	-.30 (3.00)	-.17 (3.29)	-.42 (2.30)	-.44 (2.67)
		-.20 (2.68)	-.44 (3.54)	.15 (2.25)	.11 (2.72)

Table 4. The genetic means and and their LS and BLUP estimates for three generations of selection. σ_g^2/σ_p^2 = .5.

q	d	n	true mean	LS estimate	MM(a) est.	MM(a+d) est.
.1	-1	8	.16 (2.59)	.23 (3.33)	2.00 (3.48)	.94 (3.17)
			7.05 (6.89)	7.06 (7.00)	4.35 (4.71)	6.02 (5.90)
		30	-.48 (2.22)	-.98 (3.14)	2.03 (2.58)	.40 (2.61)
			7.94 (5.10)	7.94 (5.38)	4.61 (3.34)	6.57 (4.29)
.9	1	8	-.26 (2.21)	-.22 (2.99)	-1.10 (2.36)	-.82 (2.48)
			-1.38 (3.34)	-1.57 (3.68)	-.71 (2.46)	-.86 (2.90)
		30	.26 (2.48)	-.23 (3.03)	-1.44 (2.19)	-1.04 (2.42)
			-2.99 (3.87)	-3.00 (4.20)	-1.22 (2.41)	-1.75 (3.14)

DISCUSSION

If we have the right model and know the true parameters, the linear predictors work satisfactorily even if the number of loci governing the variation is very small. The only major exception is due to fixation. Although a common practical problem, a completely different question is the use of wrong models and parameters. We may conclude that the assumptions for the use of relationship matrices do not seem to be grossly violated even with a small number of loci. The accuracy of estimating dominance values can be improved by carrying out heavier inbreeding and thus making more links between individuals over generations. The closest practical applications may be found in poultry. As in another example of non-linearity, i.e. quadratic indexes, the optimum way to utilize the dominance values or, in general, non-additive values might be to predict them for progeny from all possible mating pairs and do selection on the outcome (c.f. Jansen & Wilton, 1985).

REFERENCES

BULMER, M.G. (1980) The Mathematical Theory of Quantitative Genetics. Oxford University, Oxford.

FALCONER, D.S. (1981) Introduction to Quantitative Genetics (2nd edn). Longman, London.

HENDERSON, C.R. (1985) Best linear unbiased prediction of nonadditive genetic merits in noninbred populations. J.Anim.Sci. 60, 111-117.

HILL, W.G. (1969) The rate of selection advance for non-additive loci. Genet.Res. 13, 165-173.

JANSEN, G.B. & J.W. WILTON (1985) Selecting mating pairs with linear programming techniques. J.Dairy Sci. 68, 1302-1305.

MÄKI-TANILA, A. (1982) The validity of the heritability concept in quantitative genetics. PhD Thesis, Edinburgh University.

ROBERTSON, A. (1977) The non-linearity of offspring-parent regression. In Proc. Int. Conf. Quantitative Genetics. - ed. E. Pollak, O. Kempthorne and T.B. Bailey. Iowa State Univ., Ames, Iowa. Pp. 297-304.

SORENSEN, D.A. & B.W. KENNEDY (1984) Estimation of response to selection using least-squares and mixed model methodology. J.Anim. Sci. 58, 1097-1106.

THOMPSON, R. (1977) The estimation of heritability with unbalanced data. Biometrics 33, 497-504.