NEW TECHNIQUES OF PREDICTION OF BREEDING VALUE
FOR DISCONTINUOUS TRAITS

D. Gianola and J. L. Foulley

1 Invited paper presented at the 32nd Annual National Breeders Roundtable, May 6, 1983, St. Louis, Missouri

2 Department of Animal Science, University of Illinois, Urbana, Illinois 61801

3 Station de Genetique Quantitative et Appliquee, Centre National de Recherches Zootechniques, 78350 Jouy-en-Josas, France
INTRODUCTION

The objective of a program for genetic improvement of animals is to elicit favorable genetic trends for a merit function, usually referred to as the "breeding objective". Given the "breeding objective", the rate of genetic improvement per unit of time depends on population structure, testing design, field recording scheme, mating system, and method of predicting the breeding value of candidates for selection as this affects the accuracy of selection. Because all these are interrelated factors, it can be seriously misleading to discuss them individually in isolation from each other. This paper addresses exclusively statistical aspects of prediction of breeding value, with emphasis on recently developed techniques for categorical responses.

A categorical variable is a random quantity whose realization is measured as a response into one of several mutually exclusive and exhaustive classes. In the animal genetics literature, they are often referred to as quantal, binary or dichotomous when the number of categories of response (m) is 2, or as polychotomous, meristic or metameric when m > 2. The terms "discontinuous" or "quasi-continuous" have also been used for these traits (e.g., Gröneberg, 1952). There are many categorical variates of importance in animal production. Some examples are fertility, survival or death, litter size in polytocous species, liability to disease, resistance to toxicants, degree of calving difficulty, subjective scores measuring conformation or quality, and preference for certain types of mates in natural populations. Categorical variables are also ubiquitous in animal genetics. Earlier work (e.g., Wright, 1934a,b; Dempster and Lerner, 1950; Falconer, 1965, 1967) dealt with relationships between parameters of the observable or outward scale of measurement and those of a conceptual underlying distribution in which gene substitutions were assumed to occur. Contributions
Research in prediction of breeding value for categorical traits is recent and still undergoing development. The breakthroughs made by Henderson (1973, 1974, 1975, 1976) in developing a unified theory of estimation and prediction in mixed linear models for quantitative traits led investigators to either apply mixed model techniques to categorical responses (Schaeffer and Wilton, 1976; Berger and Freeman, 1978), or to modify the methodology in order to conform to the discrete structure of the data (Quaas and Van Vleck, 1980). An additional approach has been applying linear techniques to empirical logits described by mixed linear models (Gianola, 1980a, b). All these methods, perhaps with the exception of logit or probits transforms (e.g., Lush, Lamoreux and Hazel, 1948; Robertson and Lerner, 1949) have conceptual drawbacks, particularly in the light of the theory of quantitative genetics.

We describe in this paper Bayesian methods of prediction of breeding value for categorical traits developed by Gianola and Foulley (1982, 1983), and Foulley, Gianola and Thompson (1983). First, a general method of prediction of breeding value is discussed, irrespective of distributional assumptions. The method is then applied to normally distributed data and found to yield the "mixed model" equations (Henderson, 1973) which, when regarded from a "classical" viewpoint arise from a hybridization of selection index and least-squares techniques. Subsequently, the same strategy is used to develop predictors applicable to categorical traits in four different cases: binary responses, ordered polychotomies, mixtures of categorical and normal records, and multiple categorical responses. Estimation of covariance structures with categorical data is also discussed briefly.
PREDICTION OF BREEDING VALUE: A CONCEPTUAL FRAMEWORK

Genetic evaluation of candidates for selection can be viewed as a problem of predicting an unobservable random vector $\mathbf{s}$ from a vector of records $\mathbf{y}$ (Henderson, 1973). The objective of the evaluation is to order the candidates for selection on the basis of some rule, and to retain some of them as parents of subsequent generations. In addition, there is interest in making inferences about true differences in breeding value, given the data.

A reasonable optimality criterion for a ranking rule, at least in animal and plant breeding applications, would be to maximize the expectation of the average breeding value of the selected individuals. Cochran (1951) proved that if the joint density of $\mathbf{s}$ and $\mathbf{y}' = [y'_1, \ldots, y'_N]$ can be written as

$$ f(\mathbf{s}, \mathbf{y}) = \prod_{i=1}^{N} f_i(s_i, y'_i) $$

where $s_i$ is the breeding value of the $i$th candidate, and if $f_i(s_i, y'_i)$ is the same for all $i=1, \ldots, N$, then the optimal ranking rule, in the sense described above, would be ordering the $N$ individuals with

$$ \mathbb{E}(s_i | y'_i) $$

The $M < N$ selected candidates would be those with the largest values of [2]. Cochran's (1951) assumptions imply that candidates for selection are independent of each other, and that they all have the same amount of information. This situation seldom arises in animal breeding practice.

Bulmer (1980), Goffinet (1983) and Fernando (1983) have shown that ordering candidates with $\hat{\mathbf{s}} = \mathbb{E}(\mathbf{s} | \mathbf{y})$ yields an optimal ranking, in the sense of above, irrespective of whether candidates are related or not, or of differences in the amount of information available. Further, the

(over)
requirement of knowing the first moments of the distribution of $y$ can be relaxed by restricting attention to translation invariant functions of $y$ (Goffinet, 1983; Fernando, 1983). It should be noted that the unbiasedness and minimum mean square error properties of $E(s|y)$ (Henderson, 1973), albeit of statistical interest, are only incidental in the development of an optimal selection procedure.

From [2] it follows that the problem of prediction of breeding value can be formulated as one of conditional statistical inference. Parameters of interest in the conditional distribution of $s$ given $y$ are $E(s|y)$, and $Var(s|y)$. If $y$ is informative about $s$, then

$$Var(s_i|y) < Var(s_i),$$  

and

$$\text{Prob} \{ -d < s_i - s_i^* < d | y \} > \text{Prob} \{ -d < s_i - s_i^* < d \},$$  

where $d$ is a real number. Now, the conditional density of $s$ given $y$ is

$$g(s|y) = h(y|s) \cdot k(s)/\gamma(y),$$  

where $h(y|s)$ is the conditional density of $y$ given $s$, $k(s)$ is the marginal or prior density of $s$, and $\gamma(y)$ is the marginal density of $y$. Since $\gamma(y)$ does not vary with $s$, it is convenient to write [5] as

$$g(s|y) \propto h(y|s) \cdot k(s),$$  

which is Bayes theorem. In Bayesian terminology, $g(s|y)$ is referred to as the posterior density of $s$. The best selection rule is then

$$E(s|y) = \int_s s \cdot g(s|y) \, ds$$  

where the integration is over the range of $s$. In most instances, particularly in non-normal situations, this integration is difficult or impossible to execute. A possibility is to approximate $E(s|y)$, or posterior mean, by the mode of $g(s|y)$, or posterior mode. This entails a much simpler numerical
problem, i.e., maximization. If the posterior density is symmetric and unimodal, as in the normal case, $E(s|y)$ will then be identical to the posterior mode. Further, as the information on each candidate increases, the posterior density tends to normality (Cox and Hinkley, 1974) so the posterior mean and mode are the same, in the limit. We therefore suggest regarding a posterior mode as a reasonable approximation to the best ranking rule.

PREDICTION OF BREEDING VALUE: NORMALLY DISTRIBUTED TRAITS

Individual animal model with known means and variances

Suppose we wish to predict the additive genetic merit for a given trait of a set of individuals. Let the vector of their additive genetic values be $a \sim N(0, \sigma_A^2)$, where $A$ is the additive relationship matrix and $\sigma_A^2$ is the additive genetic variance which is assumed known. A set of records $y$ is available such that

$$ y = X\beta + a + e $$

where $X$ is a known matrix; $\beta$ is a vector of known parameters; $e$ is a vector of environmental effects and $\sigma_E^2$ is the environmental variance. With $\text{Cov}(a,e') = 0$, then $y \sim N(X\beta, V)$, where $V = \sigma_A^2 A + \sigma_E^2 I$. If $a$ and $y$ are distributed as joint multivariate normal, the best ranking rule would be ordering individuals with

$$ \hat{a} = E(a|y) = \sigma_A^2 V^{-1}(y-X\beta) $$

If the individuals are unrelated to each other, and if $E(y_i) = \mu_i$, then

$$ \hat{a}_i = h^2(y_i - \mu_i) $$

where $h^2$ is heritability in the narrow sense. In general, the problem is that $[9]$ is difficult to compute because the inverse of $V$ is required, and this may be difficult to obtain. However, there is a simpler way of computing $[9]$. Rearrange $[9]$ as

$$ V A^{-1}(\frac{1}{\sigma_A^2})\hat{a} = y - X\beta $$

(over)
Take $V = [A \frac{\sigma^2_A}{\sigma^2_E} + I] \sigma^2_E$, and use it in the preceding equation to obtain

$$(I + A^{-1} \frac{\sigma^2_E}{\sigma^2_A}) \hat{a} = y - X \beta,$$  \[10\]

which is convenient because it is easier calculating $A^{-1}$ than writing $A$ directly (Henderson, 1976). Also, [10] does not require the inversion of $V$. Note that as $\sigma^2_A \rightarrow \infty$, $\hat{a} \rightarrow y - X \beta$, and there is no shrinkage whatsoever. The expression in [10] is the best linear predictor of $\hat{a}$ and, under normality, the optimal ranking rule.

It is possible to obtain [10] directly from Bayes theorem as stated in [6]. Under normality, the prior density of $\hat{a}$ is

$$k(\hat{a}) = \exp \{-\frac{1}{2} \hat{a}' A^{-1} \hat{a}(1/\sigma^2_A)\}$$ \[11a\]

Given $\hat{a}$, $E(y|\hat{a}) = X \beta + \hat{a}$, and $\text{Var}(y|\hat{a}) = I \sigma^2_E$. Therefore, the conditional density of $y$ given $\hat{a}$, also under normality, is:

$$h(y|\hat{a}) = \{\exp -\frac{1}{2}(y-X \beta-\hat{a})'(I/\sigma^2_E)(y-X \beta-\hat{a})\}$$ \[11b\]

We now need to find the mode of the posterior density. The logarithm of this density, from [6] is

$$\ln [g(\hat{a}|y)] = \frac{1}{2}((y-X \beta-\hat{a})'(I/\sigma^2_E)(y-X \beta-\hat{a}) + \hat{a}' A^{-1} \hat{a}(1/\sigma^2_A))$$ \[11c\]

Differentiating [11c] with respect to $\hat{a}$, equating to zero and rearranging the resulting expression gives directly [10]. In this case, the posterior mode is identical to $E(\hat{a}|y)$, or posterior mean, because the posterior density is symmetrical and unimodal. In fact, it is multivariate normal.

The second item of interest in this conditional analysis is the posterior covariance matrix. Since $\hat{a}$ and $y$ are multivariate normal

$$\text{Var}(\hat{a}|y) = \text{Var}(\hat{a}) - \text{Cov}(\hat{a}, y') V^{-1} \text{Cov}(y, a')$$

$$= \text{Var}(\hat{a}) [I - V^{-1} \text{Var}(\hat{a})]$$

After algebra

$$\text{Var}(\hat{a}|y) = (I + A^{-1} \frac{\sigma^2_E}{\sigma^2_A})^{-1} \sigma^2_E$$ \[12\]
Note that the posterior covariance matrix is obtained by inverting the coefficient matrix in [10], and then multiplying this inverse by $\sigma_E^2$. It is also interesting to observe that $\text{Var}(\hat{a}|y) = \text{Var}(\hat{a}-a)$, i.e., the posterior covariance matrix is precisely the covariance matrix of prediction errors of Henderson (1973). To prove this, consider the well known equality

$$\text{Var}(\hat{a}-a) = \mathbb{E}\{\text{Var}(\hat{a}-a|y)\} + \text{Var}(\mathbb{E}(\hat{a}-a|y))$$  \hspace{1cm} [13]$$

With fixed $y$, $\text{Var}(\hat{a}|y) = 0$, $\text{Cov}(\hat{a},\hat{a}'|y) = 0$, $\mathbb{E}(\hat{a}|y) = \hat{a}$ and, from [9], $\hat{a} = \mathbb{E}(a|y)$. Hence, [13] becomes

$$\text{Var}(\hat{a}-a) = \mathbb{E}\{\text{Var}(a|y)\}$$  \hspace{1cm} [14]$$

However, it is clear from [12] that $\text{Var}(a|y)$ does not depend on $y$, so the average conditional variance is $\text{Var}(a|y)$. Hence

$$\text{Var}(\hat{a}-a) = \text{Var}(a|y)$$

Probability statements about true differences between the additive genetic merit of candidates for selection can be made by referring

$$z = (a_i-a, (\hat{a}_i-\hat{a},))/[\sigma^2(c_{ii}+c_{i,i'},-2c_{i,i'})]^{1/2}$$  \hspace{1cm} [15]$$

to a $N(0,1)$ distribution; $c_{ii}$, $c_{i,i'}$, and $c_{i,i'}$ are appropriate elements of the inverse of the coefficient matrix in [10].

**Sire evaluation model with unknown means and known variances**

In sire evaluation programs, interest usually centers on predicting transmitting ability from progeny records. Suppose the model for a progeny record is

$$y = Xb + Zu + e$$  \hspace{1cm} [16]$$

where $b$ is a vector of unknown parameters, $y$ is the vector of transmitting abilities to be predicted, $e$ is a residual vector and $X$ and $Z$ are known
matrices relating $\beta$ and $u$ to $y$. Assume

\[
\begin{bmatrix}
  y \\
  u \\
  e
\end{bmatrix} \sim N\left( \begin{bmatrix}
  0 \\
  0 \\
  0
\end{bmatrix}, \begin{bmatrix}
  GZ' & G & 0 \\
  G & 0 & 0 \\
  0 & 0 & R
\end{bmatrix} \right)
\]

If $E(y) = XB$ were known, the optimal ranking rule would be

\[ E(u|y) = GZ'V^{-1}(y-X\hat{\beta}) \]  \[17a\]

With unknown $\beta$, it is reasonable to replace $XB$ by its maximum likelihood estimator (MLE):

\[ \text{MLE}(XB) = X(X'V^{-1}X)^{-1}X'V^{-1}y = \hat{\beta} \]  \[17b\]

where $(X'V^{-1}X)^{-1}$ is a generalized inverse of $X'V^{-1}X$ and $\hat{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}y$. Using [17b] in [17a], and from the invariance property of maximum likelihood estimators it follows that

\[ \text{MLE}[E(u|y)] = \hat{u} = GZ'V^{-1}[y-X\hat{\beta}] \]  \[18\]

Irrespective of the distribution of $y$, $\hat{u}$ in [18] is the best linear unbiased predictor (BLUP) of $u$ (Henderson, 1973), i.e., the predictor with minimum variance in the class of linear unbiased statistics.

It is difficult or impossible to compute [18] in large data sets because the inverse of $V$ is required. Again, this is a non-trivial matrix with order equal to that of the vector of records. Fortunately, $\hat{\beta}$ and $\hat{u}$ can be obtained by solving

\[
\begin{bmatrix}
  X'R^{-1}X & X'R^{-1}Z \\
  Z'R^{-1}X & Z'R^{-1}Z+G^{-1}
\end{bmatrix}
\begin{bmatrix}
  \hat{\beta} \\
  \hat{u}
\end{bmatrix} =
\begin{bmatrix}
  X'R^{-1}y \\
  Z'R^{-1}y
\end{bmatrix}
\]

This system, known as the "mixed model" equations (Henderson, 1973, 1974, 1975).
is usually amenable to direct or iterative solution, particularly when the matrix \( R \) is diagonal or when it presents an exploitable pattern.

It turns out that \( \hat{\beta} \) and \( \hat{u} \) in \([19]\) are posterior means, and modes, in the context of a Bayesian argument (Dempfle, 1977). Take, a priori

\[
\begin{bmatrix}
\beta \\
u
\end{bmatrix} \sim N
\left(
\begin{bmatrix}
\alpha \\
0
\end{bmatrix},
\begin{bmatrix}
R & 0 \\
0 & G
\end{bmatrix}
\right)
\]

Given \( \beta \) and \( u \), \( E(y) = X\beta + Zu \) and \( \text{Var}(y) = R \). Assuming multivariate normality, the logarithm of the posterior density can be written as

\[
\ln g(\beta, u | y) = -\frac{1}{2} (y - XB - Zu)'R^{-1}(y - XB - Zu) + (\beta - \alpha)'R^{-1}(\beta - \alpha) + u'G^{-1}u
\]

Differentiating with respect to \( \beta \) and \( u \), equating to zero and rearranging yields

\[
\begin{bmatrix}
X'R^{-1}X + \Gamma^{-1} \\
Z'R^{-1}X
\end{bmatrix}
\begin{bmatrix}
\beta \\
u
\end{bmatrix} =
\begin{bmatrix}
X'R^{-1}y + \Gamma^{-1}\alpha \\
Z'R^{-1}y
\end{bmatrix}
\]

\[20\]

Note that \([19]\) differs from \([20]\) in that in the latter there is an augmentation of the \( \beta \)-equations with \( \Gamma^{-1} \) in the coefficient matrix and with \( \Gamma^{-1}\alpha \) in the right-hand sides. This represents the contribution of prior knowledge about \( \beta \). In the same way, the augmentation of the \( \hat{u} \) equations with \( G^{-1} \) represents the contribution of prior knowledge about the distribution of transmitting abilities in the population of sires. Vague prior knowledge about the likely values of \( \beta \) is tantamount to stating that \( \Gamma \to \infty \) and, therefore, that \( \Gamma^{-1} \to 0 \). Then \([20]\) reduces to \([19]\). Observe that if both \( \Gamma^{-1} \) and \( G^{-1} \to 0 \) (the latter indicating that the variance of sire effects is infinity), then \([20]\) reduces to a set of least-squares equations. This

(over)
provides a heuristic explanation for the inefficiency of least-squares predictors of breeding value: prior information about the distribution of breeding values in the population is not taken into account in a least-squares analysis.

An argument similar to the one employed in [13] and [14] yields

$$\text{Var} \left[ \begin{array}{c} \beta | y \\ u | y \end{array} \right] = \text{Var} \left[ \begin{array}{c} \hat{\beta} \\ \hat{u} - u \end{array} \right] = \begin{bmatrix} C_{\beta\beta} & C_{\beta u} \\ C_{u\beta} & C_{uu} \end{bmatrix} \sigma^2_e$$

where the matrix in the right-hand side of [21] is a generalized inverse of the coefficient matrix in [19] or [20] depending on the assumptions; \( \sigma^2_e \) is the residual variance in the model of [16].

**PREDICTION OF BREEDING VALUE: CATEGORICAL TRAITS**

A model used commonly in quantitative genetics for the analysis of categorical responses postulates a normally distributed variable in a hypothetical scale (Wright, 1934a; Dempster and Lerner, 1950; Falconer, 1965; Curnow and Smith, 1975). Gene substitutions are assumed to occur in this conceptual scale where normality is justified on the grounds of many loci with alleles of small effects acting upon the background of a non-negligible environmental component. A response in a particular category is viewed as a realized value of this conceptual variable falling into one of several mutually exclusive segments of the normal distribution; these segments correspond to intervals of the real line delimited by fixed threshold values. The model can be extended to accommodate unordered categories of response (Bock, 1975).

When responses are binary, the analysis can always be rendered univariate from a statistical point of view. However, when the categorical variable is polychotomous, the analysis can be univariate or multivariate.
depending on whether response categories are ordered or unordered; the order of categories is, of course, irrelevant with binary data. Unordered categorical responses rarely arise in animal breeding applications and are not discussed here. This is fortunate because the multivariate models needed to analyze them from the viewpoint of prediction of breeding value, are generally non-parsimonious and sensitive to incorrect specification of required genetic covariance matrices. Further, computations are awkward and the system of equations requiring solution is usually poorly conditioned and not suitable for iteration.

The threshold model and linear approaches

Ordered categorical responses are often dealt with by "scoring" in some way different categories, and then analyzing these scaled records via linear models in order to obtain "best linear unbiased predictors" (e.g., Schaeffer and Wilton, 1976; Berger and Freeman, 1978). Suppose there are R experimental units allocated to each of J conditions, and that each experimental unit is screened for a response in one of m mutually exclusive and exhaustive ordered categories. The random variable of interest in the jr th experimental unit is a random vector \( v_{jr} \) or order m x 1; this vector contains a 1 in the position corresponding to the category of response, and zeroes elsewhere. Categories are scored using a vector of weights \( a \) such that the scored response is \( a'v_{jr} \). Assume, for the sake of simplicity, that \( a \) is not data dependent. Let the conceptual underlying normal variate be \( y_{jr} \) and describe it with (16), that is

\[
y_{jr} = x'_{jr} \beta + z'_{jr} u + e_{jr} = v_{jr} + e_{jr}
\]

Now, given \( \beta \) and \( u \), the probability that the experimental unit will respond in the k th category (k=1,...,m) is
\[ \text{Prob} \{ t_{k-1} < y_j < t_k \mid \beta, u \} = \phi(t_k - u_j) - \phi(t_{k-1} - u_j), \quad [23] \]

where \( t_1 < \ldots < t_{m-1} \) are fixed thresholds, and \( \phi(.) \) is the cumulative normal distribution function.

In the observed or categorical scale

\[ E(a'y_j \mid u_j) = \sum_{k=1}^{m} a_k [\phi(t_k - u_j) - \phi(t_{k-1} - u_j)] \quad [24] \]

and

\[ \text{Var}(a'y_j \mid u_j) = \sum_{k=1}^{m} a_k^2 [\phi(t_k - u_j) - \phi(t_{k-1} - u_j)] - E^2(a'y_j \mid u_j) \quad [25] \]

Note that \( u_j \) in [22], [24] and [25] represents the structure of the model and that [25] is the residual variance. The implications of the threshold model are clear: 1) the expectation (eq. [24]) and dispersion (eq. [25]) structures of a linear model in the outward scale are score and frequency dependent; 2) the effect of changes in the explanatory variables \( (\beta, u) \) on the outward scale is not constant throughout the range of \( u_j \). For example

\[ \frac{\partial E(a'y_j \mid u_j)}{\partial u_j} = \sum_{k=1}^{m} a_k [\phi(t_{k-1} - u_j) - \phi(t_k - u_j)] z_j \]

illustrates that gene substitutions in the underlying scale affect the outward variate in a nonlinear manner, and to an extent dependent on the distribution of frequencies in the population and on the position of \( u_j \) with respect to the thresholds. The assumptions of additivity and of homoscedasticity often used in an analysis with linear models are, therefore, not tenable when the threshold model of quantitative genetics is invoked. Additional theoretical shortcomings of linear models for categorical traits are described elsewhere (Gianola, 1980a,b; Gianola, 1982).
Binary responses: one-way sire evaluation model

Suppose there are $S$ sires with $n_j$ progeny each. The data are arranged as a contingency table with $\sum_j n_j$ rows, and 2 columns corresponding to the categories of response, e.g., "alive" vs. "dead". The configuration of the contingency table is denoted symbolically by $Y$. The conceptual normal variable is modeled as

$$y_{jr} = \mu + s_j + e_{jr}$$

with $E(y_{jr} | \mu, s_j) = \mu + s_j$, and $\text{Var}(y_{jr} | \mu, s_j) = 1$. Note that the unit of measurement in the underlying scale is the residual standard deviation. The sires are assumed sampled from a normal distribution $\sim N(0, \alpha^{-1})$, where $\alpha^{-1} = h^2/(4-h^2)$, and $h^2$ is the heritability of the trait in the conceptual scale. Prior knowledge about the scalar $\mu$ is assumed to be vague.

Conditionally on the realized values of $\mu$ and $s_j$, the probability of response, e.g., "alive", for the $jr$th row of the contingency table is

$$\text{Prob} \{y_{jr} < t | \mu, s_j\} = \int_{-\infty}^{t-\mu-s_j} \phi(x)dx$$

$$= \phi(t-\mu-s_j), \quad [27]$$

where $t$ is a fixed threshold, and $\phi(.)$ and $\Phi(.)$ are the density and distribution functions, respectively, of a standard normal variate. Since the threshold is by itself of no interest, take $t=0$ and observe that

$$\phi(-\mu-s_j) = 1-\Phi(\mu+s_j)$$

The likelihood function in [6] is, therefore, product binomial
\[ h(Y|s,\mu) = \prod_{j=1}^{S} \left( \phi(-\mu-s_j) \right)^{n_{j1}} \left[ 1 - \phi(-\mu-s_j) \right]^{n_{j2}} \]

\[ \alpha = \prod_{j=1}^{S} \left[ 1 - \phi(\mu+s_j) \right]^{n_{j1}} \left( \phi(\mu+s_j) \right)^{n_{j2}} \]

where \( n_j = n_{j1} + n_{j2} \), and \( n_{j1} \) is the number of progeny of the \( j \)th sire falling in the "respondent" category. The log-posterior in [6] becomes

\[ \ln[g(\mu, s|Y)] = \sum_{j=1}^{S} \{ n_{j1} \ln[1 - \phi(\mu+s_j)] + n_{j2} \ln[\phi(\mu+s_j)] \} \]

\[ - \frac{1}{2} \left( \frac{4-h^2}{h^2} \right) s's \]

The best selection rule, \( E(s|Y) \), can be calculated by integrating \( s \ g(s|Y) \) over the range of \( s \). However, in this case the integration is impossible or difficult to execute. The posterior mean is therefore approximated by finding the posterior mode; this entails solving a nonlinear system of equations. Gianola and Foulley (1982) showed that the posterior mode can be found iterating with

\[
\begin{bmatrix}
\Sigma_{j=1}^{S} n_j w_j[i] & n_{w_{11}}[i] & n_{w_{22}}[i] & \ldots & n_{w_{SS}}[i] \\
0 & n_{w_{11}+\alpha} & 0 & \ldots & 0 \\
0 & 0 & n_{w_{22}+\alpha} & \ldots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \ldots & n_{w_{SS}+\alpha} \\
\end{bmatrix}
\begin{bmatrix}
u[i+1] \\
\vdots \\
\vdots \\
\vdots \\
\vdots \\
\end{bmatrix}
\]

\[
\begin{bmatrix}
u[i+1] \\
\vdots \\
\vdots \\
\vdots \\
\vdots \\
\end{bmatrix}
\]
where \([i]\) indicates the \(i\) \text{th} iterate. In [31]

\[
\begin{align*}
w_{jj} &= \frac{\phi^2(u+s_j)}{1-\phi(u+s_j)\phi(u+s_j)} \\
v_{jr} &= \begin{cases} \\
-\frac{\phi(u+s_j)}{1-\phi(u+s_j)} & \text{if the response is in the first category} \\
\frac{\phi(u+s_j)}{\phi(u+s_j)} & \text{otherwise.}
\end{cases}
\end{align*}
\]

Note the parallelism between [31] and the equations which result in a normal setting, e.g. [19]. In the binary case, \(w_{jj}\) measures the loss of information accruing by going from a quantitative (normal) to a discrete (binary) scale; this weighting factor, which has been used to relate "underlying" heritability to "all or none" heritability (Dempster and Lerner, 1950; Gianola, 1982) is maximum \((w_{jj}=.637)\) when \(\phi(u+s_j)=.50\).

The posterior covariance matrix can be approximated as
where \( C_{\mu\mu}, C_{\mu s}, C_{s\mu}, \) and \( C_{ss} \) are appropriate partitions of the inverse of \([31]\) after the solutions stabilize.

**Binary responses: general model**

The data appear as an \( R \times 2 \) contingency table. In the \( j \)th row of this table, there are \( n_{j1} \) experimental units classified as "responding" and \( n_{j2} \) as "not responding". The total number of experimental units in the \( j \)th row, \( n_j = n_{j1} + n_{j2} \), is assumed fixed by sampling; of course \( n_j \) may vary from row to row. The conceptual variable associated with the \( jr \)th experimental unit \((r=1, \ldots, n_j)\) is modeled as

\[
y_{jr} = x_j'\beta + z_j'u + e_{jr} \quad j=1, \ldots, R, \quad r=1, \ldots, n_j
\]

with \( E(y_{jr}|\beta, u) = x_j'\beta + z_j'u \) and \( \text{Var}(y_{jr}|\beta, u) = 1 \). In \([34]\), \( \beta \) and \( u \) are unknown vectors, and \( x_j' \) and \( z_j' \) are known incidence vectors. The matrix \( X \) resulting from horizontal concatenation of \( x_1', x_2', \ldots, x_R' \) is assumed, without loss of generality, to have full-column rank. Hence, \( \beta \) is identifiable.

A priori, \( u \sim N(0, G) \), and knowledge about \( \beta \) is assumed vague. Hence, the prior density in \([6]\) is

\[
k(u) = \exp \left[ -\frac{1}{2} u' G^{-1} u \right] \quad [35]
\]

As in \([27]\),

\[
\text{Prob} \{y_{jr} < t|\beta, u\} = \phi(t-x_j'\beta - z_j'u) = 1 - \phi(x_j'\beta + z_j'u) \quad [36]
\]

after setting \( t=0 \). The log-posterior is
\[ \ln[g(\beta, u|Y)] = \sum_{j=1}^{R} \left( n_{j1} \ln[1-\phi(x'_j\beta + z'_j u)] + n_{j2} \ln[\phi(x'_j\beta + z'_j u)] \right) - \frac{1}{2} \beta' G^{-1} \beta \]  

The posterior mode can be found (Gianola and Foulley, 1982, 1983) by iteration with

\[
\begin{pmatrix}
X'_W[i] & X'_W[i]Z \\
Z'_W[i] & Z'_W[i]Z + G^{-1}
\end{pmatrix}
\begin{pmatrix}
\hat{\beta}[i+1] \\
\hat{u}[i+1]
\end{pmatrix}
= \begin{pmatrix}
X'_W[i]y[i] \\
Z'_W[i]y[i]
\end{pmatrix} 
\]  

where

\[ y[i] = X\hat{\beta}[i] + Z u[i] + (W[i])^{-1} v[i-1] \]  

\[ W[i] = \text{Diag} \left\{ \frac{\phi^2(x'_j\beta + z'_j u) n_j}{[1-\phi(x'_j\beta + z'_j u)]\phi(x'_j\beta + z'_j u)} \right\} ; j=1,...,R \]  

and

\[ v[i] = \{ \phi(x'_j\beta + z'_j u) \left[ \frac{n_j^2}{\phi(x'_j\beta + z'_j u)} - \frac{n_j1}{1-\phi(x'_j\beta + z'_j u)} \right] \} ; j=1,...,R \]  

The posterior covariance matrix is approximately

\[
\text{Var} \begin{pmatrix}
\beta|Y \\
u|Y
\end{pmatrix} = \begin{pmatrix}
X'_W & X'_WZ \\
Z'_W & Z'_WZ + G^{-1}
\end{pmatrix}^{-1} 
\]

where \( \hat{W} \) is \( W \) evaluated at the modal values \( \hat{\beta} \) and \( \hat{u} \). Probabilities of "response" and "non-response" for different rows can be estimated, respectively, as

\[ 1 - \phi(x'_j\hat{\beta} + z'_j\hat{u}) \]  

\[ \phi(x'_j\hat{\beta} + z'_j\hat{u}) \]

(over)
Ordered polychotomies: general model

The data are organized as an R x m contingency table with typical element n_{jk}, the number of respondents in the k\text{th} category (k=1,...,m) for the j\text{th} row (j=1,...,R). The conceptual variate is modeled as in \cite{22} and the probability of response is as in \cite{23}, with t_1 < t_2 < ... < t_{m-1}; t_0 = \infty and t_m = \infty. The variables of interest are \beta, u and \tau, the vector of thresholds. A priori, u \sim N(0, \Sigma) and prior knowledge about \beta and \tau is assumed to be vague. The log-posterior is

$$\ln g(t, \beta, u | Y) = \sum_{j=1}^{R} \sum_{k=1}^{m} n_{jk} \ln [\phi(t_k - u_j) - \phi(t_{k-1} - u_j)]$$

$$- \frac{1}{2} u' \Sigma^{-1} u$$

The estimation equations are similar to \cite{38}, but including a sector for the thresholds. Details are in Gianola and Foulley (1983).

Mixtures of categorical and normal variates

Many animal breeding data sets comprise observations on categorical and quantitative variates. Furthermore, the probability of "response" for the categorical variate (Y), e.g., calving ease, may be a function of a quantitative variate (y), e.g., birth weight. This problem can be addressed in the context of the general framework discussed in this paper.

Suppose \theta is a vector of interest, and write \cite{6} as

$$g(\theta | Y, y) = h(Y, y | \theta) \cdot k(\theta)$$

Further,

$$h(Y, y | \theta) = p(Y | y, \theta) \cdot q(y | \theta)$$

so

$$g(\theta | Y, y) = p(Y | y, \theta) \cdot q(y | \theta) \cdot k(\theta)$$
Equation [47] gives the basis for finding the posterior mode of $\theta$ using a two-stage modeling approach (Foulley, Gianola and Thompson, 1983). In the first stage, a multivariate normal density is taken to represent the conditional distribution of $y$ given $\theta$. In the second stage, a multivariate normal model is used to derive the location and dispersion, given $\theta$ and $y$, of a conceptual underlying variable. The threshold model is then used to map the underlying variate onto the discrete scale. Full details are presented in Foulley, Gianola and Thompson (1983).

**Multiple categorical responses**

The situation where an experimental unit is screened for multiple categorical responses also arises in animal breeding applications. For example, calving difficulty and calf livability are considered in dairy sire evaluation programs in Canada (Cady and Burnside, 1982). Apart from computational difficulties, this problem can also be solved by finding the mode of a posterior distribution.

The data appear as a three-dimensional contingency table. There are $R$ rows corresponding to combinations of explanatory variables, and a plane representing all combinations of responses of two ordered categorical variates (e.g., calving difficulty and calf livability) with $m$ and $n$ categories of response, respectively. The number of experimental units responding in the $jkl$ th entry of the table is $n_{jkl}$, and $n_j = \sum_k n_{jkl}$ is assumed fixed by sampling; $n_j$ can vary from row to row. The models for the two normal conceptual variables in the $q$ th experimental unit are:

$$y_{jq1} = x_{j1}^T \beta_1 + z_{j1}^T u_1 + e_{jq1} = v_{j1} + e_{jq1} \quad [48a]$$

$$y_{jq2} = x_{j2}^T \beta_2 + z_{j2}^T u_2 + e_{jq2} = v_{j2} + e_{jq2} \quad [48b]$$
with
\[ E(y_{jq1}\mid \beta_1, u_1) = x_{jq1}^T \beta_1 + z_{jq1}^T u_1 \]
\[ E(y_{jq2}\mid \beta_2, u_2) = x_{jq2}^T \beta_2 + z_{jq2}^T u_2 \]
\[ \text{Var}(y_{jq1}\mid \beta_1, u_1) = \text{Var}(y_{jq2}\mid \beta_2, u_2) = 1 \]
\[ \text{Cov}(y_{jq1}, y_{jq2}\mid \beta_1, \beta_2, u_1, u_2) = r_{12} \]

The mapping from the normal to the categorical scale for the first variate is done via a set of thresholds \( \{t_1, \ldots, t_{m-1}\} \); the thresholds for the second variable are \( \{s_1, \ldots, s_{n-1}\} \). Now, given \( \beta_1, \beta_2, u_1 \) and \( u_2 \), the probability that the \( q \)th experimental unit in the \( j \)th row will respond in the \( k \)th category of the first variate and in the \( \ell \)th category of the second one, is
\[
\text{Prob}\{t_{k-1} < y_{jq1} < t_k, s_{\ell-1} < y_{jq2} < s_\ell\} = \int_{t_{k-1}}^{t_k} \int_{s_{\ell-1}}^{s_\ell} \phi(x_1, x_2) \, dx_1 \, dx_2
\]
\[
= p_{jk\ell}
\]
where \( \phi(x_1, x_2) \) is a bivariate standard normal density function with correlation \( r_{12} \). Now
\[
p_{jk\ell} = \phi(t_k - u_{j1}, s_\ell - u_{j2}) - \phi(t_{k-1} - u_{j1}, s_\ell - u_{j2}) - \phi(t_k - u_{j1}, s_{\ell-1} - u_{j2})
\]
\[+ \phi(t_{k-1} - u_{j1}, s_{\ell-1} - u_{j2}) \]
so the likelihood function is then
\[
h(Y_1, Y_2\mid \beta_1, \beta_2, u_1, u_2) = \prod_{j=1}^{R} \prod_{k=1}^{m} \prod_{\ell=1}^{n} p_{jk\ell}
\]
If prior knowledge about \( \beta_1 \) and \( \beta_2 \) is vague, and
then the log-posterior in [6] is

\[ g(B_1, B_2, U_1, U_2 | Y_1, Y_2) = \sum_{j} \sum_{k} n_{jk} \ln(P_{jk}) - \frac{1}{2} (u' u) G^{-1} \]  \[ \left[ \begin{array}{c} u_1 \\ u_2 \end{array} \right] \]

The problem of finding the posterior mode is then largely computational. Nevertheless, the equations are similar to those described by Gianola and Foulley (1983).

Unknown covariance structure

So far, we have assumed that the prior covariance matrix, G, is known. In a sire evaluation model, and with the residual variance equal to 1

\[ G = A[h^2/(4-h^2)] \]

where A is the additive relationship matrix. Hence, the problem of unknown G reduces to one of unknown heritability. In principle, \( h^2 \) could be estimated in the outward scale using an arbitrary scoring procedure, and this estimate could be transformed to the underlying scale using well known formulae (Dempster and Lerner, 1950; Vinson et al., 1976; Gianola, 1982). This is perhaps the simplest course of action as Van Vleck (1972) found that these expressions, at least in the binary case, yield reasonable approximations to the true underlying heritability provided that the covariances between relatives are estimated from half-sib family structures.

Alternatively, equations [38] and [39] suggest a "generalized" linear model with variance

\[ \text{Var}[y^{[i]} | \theta^{[i]}] = Z' G Z + (W^{[i]})^{-1} \]
Hence, it is possible that some of the methods available for normal data, e.g., restricted maximum likelihood, could be mimicked here. Note that the residual variance does not need to be estimated in this model as it is equal to 1, by definition.

CONCLUSION

This paper presents an overview of recently developed techniques of estimation of breeding value for categorical traits. A Bayesian framework is adopted to address the problem of genetic evaluation of candidates for selection with the aim of maximizing the expected breeding value of selected individuals. In all instances considered here, the mode of a posterior density is regarded as a reasonable approximation to the posterior mean or best ranking rule. The methods discussed are intrinsically nonlinear and require iterative implementation. In theory, they are superior and more informative than their linear model counterparts, where the assumptions needed to deal with categorical data are clearly violated. Work is needed to establish the trade-offs between the additional computing costs stemming from an iterative implementation, and the benefits accruing from a more accurate evaluation of candidates for selection.

ACKNOWLEDGEMENTS

The senior author wishes to acknowledge the Holstein Association, Brattleboro, Vermont, for supporting his work in the area of categorical data in animal breeding.
REFERENCES


-151- (over)


Wright, S. 1934a. An analysis of variability in number of digits in an inbred strain of guinea pigs. Genetics 19:506
Wright, S. 1934b. The results of crosses between inbred strains of guinea pigs differing in number of digits. Genetics 19:537
Question to Dr. Daniel Gianola

Title: "New Techniques for Prediction of Breeding Value for Discontinuous Traits"

Question from Walter R. Harvey: In the binomial case, under what conditions would you expect BLUP and the procedures described to yield widely different results? Also, with ordered categories how many classes does one need to simply analyze the data as a continuous variable?

Answer: The question can be answered at two levels: theoretical and empirical. From a modeling viewpoint, the methodology described here can be regarded as an attempt to approximate the best ranking rule. On the other hand, linear approaches, e.g., "BLUP", disregard distributional aspects of the problem and, therefore, should be expected to be less powerful or informative. I deliberately refer to this latter method with quotes because in a categorical situation with an underlying variate following a mixed model, the covariance structure required to complete a linear model formulation cannot be specified, i.e., BLUP does not exist.

At the empirical level, it is not possible to state unequivocally that the methods described here are uniformly better for ranking purposes than linear approaches, e.g., "BLUP". Ranking results depend on the incidence and heritability of the trait in question, the model used for the underlying variate, the number of categories of response and, of course, on the idiosyncrasies of a given data set! Albert Meijering from Zeist, Holland, has compared "BLUP" with the procedure described here using Monte Carlo methods. With binary responses the two methods gave virtually the same ranking of sires (under a number of sampling scenarios) except when the model for the underlying variate was mixed, and the distribution of experimental units by cells (combination of levels of "fixed" factors) was highly unbalanced. In this case the nonlinear procedure was generally better than the linear approach. In some situations ($h^2=0.20$; incidence = 1%) this superiority was 12% in terms of "efficiency" of selection.

The cost of processing data is very small relative to the total cost of a breeding program (testing facilities, field personnel, etc.). New methodologies simply affect the way that data are processed. Hence, small increases in accuracy are usually cost-effective because the total cost of an evaluation is affected in a slight manner, and these new evaluations have multiplicative industry-wide effects. If this were not the case, we would still be evaluating dairy sires via the daughter-dam comparison!