INTRODUCTION

Accurate prediction of breeding values is an important part of animal breeding programs. Best linear unbiased procedures developed by Henderson (see Henderson, 1973) are currently being used to evaluate sires in the northeastern United States, Canada, and some parts of Europe. The method requires knowledge of genetic parameters of the base population, before forces such as genetic drift or selection started operating with consequent changes of genetic variances and covariances. In this paper we report methods that can be used to obtain estimates of base population parameters, when data are available from populations that have undergone several cycles of random mating or selection.

RANDOM MATING

In a random mating population of effective size N, the expected value of the additive genetic variance at generation t is (Falconer, 1981):

\[ \mathbb{E}(\sigma_A^2) = (1-1/(2N))^t \sigma_A^2 = (1-F_t)^2 \sigma_A^2 \]  

where \( \sigma_A^2 \) is the additive genetic variance in the base population and \( F_t \) is the inbreeding coefficient. When an analysis is carried out within generations, a traditional estimator such as the variance component between half-sibs has expectation equal to \( 1/4(1-F_t)^2 \sigma_A^2 \).

Data may be available from a control population and in this section we show how we can use the complete vector of observations comprising the whole period of random mating to obtain an estimator of \( \sigma_A^2 \).

1. The Statistical Model

We assume that the observations can be described by the following mixed model:

\[ y_{ijk} = u + g_i + s_{ij} + e_{ijk} \]  

with \( i=1, \ldots, t; j=1, \ldots, b; k=1, \ldots, c; \) and \( n=tbc \).

In (2), \( y_{ijk} \) is the observation on the \( ijk \) animal, \( u \) is the overall mean; \( g_i \) is a fixed generation effect; \( s_{ij} \) is a random

*Department of Animal and Poultry Science, University of Guelph, Guelph, Ontario, Canada N1G 2W1
within generation sire effect and \( e_{ijk} \) is a random error. In matrix notation, we write (2) as:

\[
y = Xb + Zu + e
\]

where \( y \) denotes the \( nxl \) vector of observations, \( b \) and \( u \) are the vectors of fixed and random effects respectively, \( X \) and \( Z \) are design matrices and \( e \) represents the vector of random errors. It is usually assumed that \( \text{Var}(u) = I\sigma^2_u \), \( \text{Var}(e) = I\sigma^2_e \) and

\[
\text{Var}(y) = ZZ'\sigma^2_s + I\sigma^2_e
\]

where \( \sigma^2_s \) is the variance component between half sib progeny groups, \( \sigma^2_e \) is the error variance and \( I \) is the identity matrix.

In attempting to estimate the base population additive genetic variance from a control population that has undergone several cycles of random mating, allowance must be made for the additional correlated structure that develops among the observations as a result of genetic drift. A model equivalent to (2) is:

\[
y = u + g + a_1 + e_{ij}
\]

with \( i=1, \ldots, t \); \( j=1, \ldots, b \); where \( y_{ij} \) is the \( ij \)th observation, \( u \) and \( g_i \) are as in (2), \( a_{ij} \) is the additive genetic effect of the \( ij \)th animal and \( e_{ij} \) is the environmental effect. The variance-covariance matrix of the vector of all observations, \( y \), is

\[
\text{Var}(y) = ZGZ'\sigma^2_A + I\sigma^2_e
\]

where \( G \) is an \( n x n \) matrix of additive genetic relationships among the observations, \( I \) is an identity matrix of order \( n \) and \( \sigma^2_A \) and \( \sigma^2_e \) are the additive genetic variance in the base population and the environmental variance (assumed constant each generation) respectively.

2. Methodology.

To obtain an estimator of \( \sigma^2_A \) we use an extension of Henderson's method 3 (Henderson, 1953). In our case we compute the total sum of squares, \( y'y \), \( R(u,g,s) \) and \( R(u,g) \); in taking the expectations of these reductions we assume \( \text{Var}(y) \) is given by (4). (See Searle, 1971 for the \( R(.) \) notation). The error sum of squares is,

\[
\text{SSE} = y'y - R(u,g,s) - R(s/u,g) = R(u,g,s) - R(u,g).
\]

Hence we solve the following equations which are functions of the variance components in (5)

\[
E[R(s/u,g)] = \sum_{i=1}^{l} \left( \sigma^2_A + (n-tb)\sigma^2_e \right) - \sigma^2_A - (t-1)\sigma^2_e
\]

\[
E[\text{SSE}] = \sum_{i=1}^{l} (1+F_i) - \sigma^2_A \sum_{i=1}^{l} \left( \sigma^2_A + (n-tb)\sigma^2_e \right)
\]

Under this model, the only part of the relationship matrix
which is used in the expectations comprises the block diagonals of relationships among observations within a generation. The procedure yields unbiased and translation invariant estimators of $\sigma^2$, the additive genetic variance in the base population and of $\sigma^2$. Although in this example we deal only with the balanced case, extension of the procedure to unbalanced data is straightforward.

As an illustration of the above procedure, we show the results of Monte Carlo simulations. We sample from a non-inbred base population which is in equilibrium within and between chromosomes. Thus initially, each gamete has a value for the trait which is normally distributed with variance $\sigma^2/2$. After a first generation, the gametes produced by an individual are sampled from a normal distribution with expected value equal to the mean value of its maternal and paternal chromosomes and variance equal to $(\sigma^2/2)(1-F_i^2)$, where $F_i$ is the individual's inbreeding coefficient.

Table 1. Inbreeding coefficient ($F$), additive genetic variance (from distribution of pairs of chromosomes) within generations $\sigma^A(t)$, and estimates of initial additive genetic variance estimated within generations, $\sigma^A(1)$ and using all the data $\sigma^A(1)$. Twelve cycles of random mating with 5 males and 50 females each generation.

<table>
<thead>
<tr>
<th>Generation</th>
<th>1</th>
<th>4</th>
<th>7</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>$F$</td>
<td>0.00</td>
<td>0.08</td>
<td>0.14</td>
<td>0.24</td>
</tr>
<tr>
<td>$\sigma^2(t)$</td>
<td>9.47</td>
<td>8.88</td>
<td>8.22</td>
<td>7.24</td>
</tr>
<tr>
<td>$\sigma^A(1)$</td>
<td>9.84</td>
<td>10.44</td>
<td>9.20</td>
<td>9.65</td>
</tr>
</tbody>
</table>

Average S.D. = 12.55

Table 1 shows average values of the inbreeding coefficient, the variance among genotypic values within generations and the estimates of the base population variance obtained from (5) together with the standard deviation of 200 replicates. As inbreeding builds up the additive variance declines but the proposed estimator yields estimates in good agreement with the additive variance at generation 1. As expected, the use of all the data leads to a considerably more accurate estimate.

**SELECTION**

In a large population, with the genetic model we have assumed in this work, selection causes a change in the additive variance due to generation of linkage disequilibrium (Bulmer, 1971). In the absence of linkage, after $t$ cycles of selection,
\[ \sigma_A^2(t) = \sigma_A^2(1) + CL(t) \]  
(6) and 

\[ CL(t+1) = 0.5 \Delta r_{IA}^2(t) \sigma_A^2(t) + 0.5 CL(t) \]  
(7) where

\( \sigma_A^2(1) \) is the additive genetic variance before selection, \( CL(t) \) is the effect of disequilibrium, \( r_{IA} \) is the correlation between selection criterion and breeding value and \( \Delta \) is a constant that depends on the type of selection applied. Since selection generates disequilibrium and recombination breaks it down, an equilibrium value is rapidly reached and the additive variance stabilizes.

The problem that we address in this section is how to obtain estimates of the additive variance of the base population before selection started operating, using data that have arisen by some type of selection.

1. Statistical Model and Methodology.

We use the same model we discussed in the section on random mating. The methodology has to be modified since, as a result of selection, the variance of the vector of all observations, \( y_s \), is

\[ V(y_s) = ZGZ'\sigma^2 + I\sigma^2 \]

where \( G \) is the additive relationship matrix appropriately modified to account for selection. In order to do this, we need to know the type of selection applied and the criterion of selection. If selection is by truncation, on the individual's own performance, \( \Delta \) and \( r_{IA} \) in (7) are replaced by \( -i(i-x_T) \) and \( h^2(T) \) respectively, where \( x_T \) is the truncation point and \( h^2(T) \) is the heritability at generation \( T \).

As an illustration, assume that selection operates on the male's own performance. At generation 2, the diagonals of \( G \) must be multiplied by \( (1-0.25 i(i-x_T)^2) \) and the off-diagonals by \( (1-i(i-x_T)h^2(T)) \). At generation three these are replaced by the appropriate expressions that can easily be derived from (6) and (7). In late generations, these formulae become more cumbersome to write. One can use the equilibrium value or some approximation on the lines suggested by Thompson (1976).

The results of Monte Carlo simulation of two cycles of selection (discrete generations) using this approach are illustrated in Table 2. The data comprise 300 observations; the first 100 are from unselected sires and the remaining 200 from selected ones. The additive genetic variance declines as selection proceeds, but using the proposed estimator, the estimate of 9.66, over 400 replicates, is 9.67. The use of the relationship matrix without accounting for disequilibrium yields, as expected, an estimate which is biased downwards (Table 2).
Table 2. Estimates of the additive genetic variance before selection \( \sigma_A^2(1) \), using the proposed method and using the relationship matrix not modified to account for selection. Ten males out of 100 are selected for 2 generations and each one is mated to 10 dams. (400 replicates)

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \sigma_A^2(1) )</td>
<td>9.66*</td>
<td>1.65</td>
</tr>
<tr>
<td>Estimate of ( \sigma_A^2(1) ) using proposed method</td>
<td>9.67</td>
<td>5.65</td>
</tr>
<tr>
<td>Estimate of ( \sigma_A^2(1) ) ignoring disequilibrium</td>
<td>7.01</td>
<td>4.07</td>
</tr>
</tbody>
</table>

*At generation 2 and 3, this variance declines to 8.67 and 8.30 respectively. The corresponding predicted values using (6) and (7) are 8.68 and 8.34. All data are used to estimate the initial variance.

DISCUSSION

Probably the most limiting characteristic of the method we have outlined is that it requires knowledge of the parameters we set up to estimate. This requirement also has to be fulfilled in a method proposed by Henderson (1980) under a different selection model and we cannot provide a better solution to the problem than the one suggested by him. Namely, one has to use educated guessed values and as these approach the true values the bias due to selection decreases.

It is also assumed that selection intensities are known. This is not as troublesome as it may appear since the value of \( i(1-x_T) \) is close to 0.8 for a wide range of selection intensities. The method has to be extended to allow for overlapping generations. Comparison with other methods such as MINQUE may be interesting regarding the sampling error of the estimators. The modification of the relationship matrix can probably be incorporated in the mixed model equations for sire evaluation. This will be presented elsewhere.

Finally we want to point out that the model used assumes that gene frequencies do not change as a result of selection. If genes of large effect are present this will not be true and this method will not yield estimates of the base population additive variance.

SUMMARY

In random mating populations with discrete generations, the additive genetic variance declines due to random genetic drift. Using the complete vector of observations, it is shown that inclusion of the additive genetic relationship matrix among the observations in various expectations of quadratic forms, leads to unbiased estimators of the base population additive genetic variance. Under a model of infinitely many additive loci, selection causes negative linkage...
disequilibrium with associated reduction in the additive genetic variance. It is shown that appropriate modification of the relationship matrix leads to estimators of the base population variance prior to selection. The results are illustrated with Monte Carlo simulations.

REFERENCES


