Apart from designing a breeding programme the estimation of breeding values is one of the most central tasks in any breeding scheme, and it is a task which comes up repeatedly. There are of course many problems and only a few can be dealt with in this paper. The reader interested in more general reviews may consult the papers by HENDERSON (1972, 1975) and THOMPSON (1979).

To indicate the scope of this paper I will narrow my discussion to one trait only, discussion usually being centered around BLUP (Best Linear Unbiased Prediction) and the discussion of most points being motivated by problems arising in dairy cattle breeding. Usually I tried to ask simple questions which often arise in a disguised manner in practical application. By asking simple questions the problem can often be reduced to the most simple and fundamental form, but there is the danger that looking at one factor in isolation will not give the correct picture due to interaction.

In dealing with statistical problems of estimating breeding values one can accept the given design and try to find the appropriate statistical technique no matter how complicated it may be or else one can try to work in both directions, appropriately modifying the breeding design and then finding the appropriate statistical technique. In using the second approach quite often simpler methods will be adequate for the purpose.

For a practical breeder quite a large fraction of the current literature on estimation of breeding values might seem too sophisticated taking into account the rough nature of the data and of the operation in the field. But here I contend that an exact analysis can indicate far better where there are critical and sensitive areas in the operation and where on the other hand simplifications will be possible with little or no costs.

General considerations

Discussing problems in estimation a short look at the field of statistics may be worthwhile. As I see it there are three non independent mainstreams which are relevant to the present discussion.

The by now classical statistics as e.g. reviewed in books like MOOD & GRAYBILL (1963) or KENDALL & STUART (1973) with the search for general estimation procedures (Method of Moments, Maximum Likelihood, Least Squares) and the use of various optimality criteria (Unbiasedness, efficiency, consistency, minimum mean square error, unbiased minimum variance etc.)

*Lehrstuhl für Tierzucht der TU-München
D-8050 Freising-Weihenstephan

104
The Bayesian approach to statistics as for the special case is laid out e.g. by Lindley & Smith (1972).

The decision theoretic framework of statistics as it was started mainly by Wald (1950) and presented in books like Ferguson (1967).

In the decision theoretic approach great care is taken to pick a procedure (in our case estimation procedure) which minimizes the average loss. The loss is defined in such a way that if we knew the items of interest there would not be any loss. Loosely speaking in the estimation of breeding values there are two kinds of losses to be considered. Firstly the reduction of the achievable genetic progress due to the fact that we have to estimate the breeding values (better or worse depending on the procedure) and secondly the cost of carrying out the estimation. The first kind of loss will be dealt with later; the second one is easily illustrated. Dairy sire evaluation by the contemporary comparison method is a very cheap procedure and therefore can even in very large populations be carried out every month. Using an elaborated BLUP procedure we can considerably gain in precision (Demmlfe a. Hagger, 1982) but the computing costs are also far higher, and it is also of questionable value if as a consequence estimation is carried out fewer times, since then waiting time increases (lengthened generation interval) with consequences for the genetic progress per year.

The Bayesian nature of our procedures will be illustrated with the following simplified examples.

Example 1: A simple random sample of animals is drawn and the animals are randomized over 5 groups. Each group is getting a different ration (e.g. different amount of energy) and by chance (or bad design) the number of animals in each group (n_i) may be different.

Example 2: Random samples (of unequal size) are drawn from 5 populations and are tested at a testing station.

Example 3: From a random mating population we draw at random 5 sires, mate them at random and take n_i daughters from each sire on a testing station.

Now the conventional linear model for all three examples can be written as

\[ y_{ij} = \mu_i + e_{ij} \quad \text{or} \quad y_{ij} = \mu + a_i + e_{ij} \]

where for convenience we assume that the e is independently normally distributed. In the first example a good (by many criteria) estimator of \( \mu_i \) would be the least squares (LS) estimator. The LS-estimator utilizes all the available information in the data (it is sufficient) and is simply given by

\[ \hat{\mu}_i = \frac{1}{n_i} \sum_j y_{ij} \]

The same estimator could also be used in the second and third example. However, in the third example we can do more. There \( a_i \) is half the breeding value of sire i and the whole population of sires
we can view as a collection of breeding values. By the very definition (FALCONER 1960, p.120) of breeding values (in simple cases), we know that the mean of the population of breeding values is zero and the variance is the additive genetic variance. Therefore we have \( a_i \sim (0, \sigma_A^2/4) \). If we draw a sire at random out of that population, (with no records on relatives), our best estimator (by a number of criteria) is the mean of the population. That estimator is unbiased and the error variance [average over samples of \((\bar{a}_i - a_i)^2\)] is equal to \( \text{Var}(\bar{a}_i - a_i) = \text{Var}(0 - a_i) = \sigma_A^2/4 \). Thus apart from the least squares estimator of \( \mu + a_i \) we have an estimator of \( a_i \), the a priori estimator. If we know \( \hat{\mu} \), then our least squares estimator of \( a_i \) would be \( \hat{a}_\text{LS} = \bar{y}_i - \hat{\mu} \) with error variance of \( \sigma_A^2/n_i \). Thus we have two estimates and combine them linearly, weighted by the reciprocal of the error variances.

\[
\hat{\sigma}_c = \frac{1}{\sigma_A^2/4 + \sigma_e^2/n_i} \left[ \frac{1}{\sigma_A^2/4} \hat{a}_\text{ap} + \frac{1}{\sigma_e^2/n_i} \hat{a}_\text{LS} \right],
\]

\[
= \frac{1}{\sigma_A^2/4 + \sigma_e^2/n_i} \left[ \frac{1}{\sigma_A^2/4} \bar{y}_i - \bar{y}_i \right]
\]

\[
= \frac{\sigma_e^2/4}{\sigma_A^2/4 + \sigma_e^2/n_i} (\bar{y}_i - \mu)
\]

Since usually we do not know \( \mu \) we replace it by the best estimator which, if we accept the random nature of \( a_i \) is equal to

\[
\hat{\mu} = \sum_i \frac{1}{\sigma_A^2/4 + \sigma_e^2/n_i} \bar{y}_i.
\]

Thus having no prior estimate of \( \mu \), we estimate

\[
\mu + a_i \text{ by } \hat{\mu} + \frac{\sigma_e^2/4}{\sigma_A^2/4 + \sigma_e^2/n_i} (\bar{y}_i - \hat{\mu})
\]

The last estimator is the BLUP estimator in this case. For known \( \mu \) ROBERTSON (1955) gave the same estimator by a slightly different derivation.

In example 3 this approach is certainly justified if we take a random sample of sires out of a random mating population, whose additive genetic variance is known, at least approximately. In sire evaluation I would not even hesitate to regard the male offsprings from carefully planned matings as a random sample. After all the estimated breeding values of the sires used and likewise of the dams are very similar and one could produce (by modern techniques) very many offsprings. The bulls tested are just a random sample out of that hypothetical population, whose variance, however, would be smaller than \( \sigma_A^2/4 \).
In example 1 it stretches the imagination too much to consider the different rations as a random sample out of a well defined population. In addition what parameters (variances) would be appropriate to that population. Thus only the information provided by the data is utilized by most statisticians.

In example 2 opinions differ. It is hard to imagine a population of different breeds out of which the tested breeds are a random sample. Also what are the parameters of such a population? There I would restrict myself to the information provided by the data but clearly in some random sample tests an essentially bayesian approach is used.

Some statisticians are very critical of bayesian methods but to paraphrase KEMPTHORNE (1972) there is no controversy about one type of use of Bayes's theorem, if this is merely a statement of conditional probability and the probabilities are frequency probabilities. It is therefore important that the prior distribution of \( a_i \) and the distribution of \( e_{ij} \) are at least partially known and not just picked for mathematical (or other) convenience.

As already stated from a decision theoretic point of view the main criterium to judge the methods should be the achievable genetic progress and, usually of lesser importance, the cost of the estimation procedure. Though there are considerations on that point (COCHRAN, 1951, HENDERSON, 1963) it is still true that due to the complexity of the situation simpler criteria like Mean Square Error (MSE), or Unbiased Minimum Variance are used. As the following extremely simple example shows, the values of these criteria are not necessarily proportional to the genetic progress.

Example 4: From a well defined population (with known \( \mu, \sigma^2, \sigma^2 \)) we draw a random sample of unrelated animals, test them and select the better fraction of animals. Since we have the same amount of information on each animal there are two obvious methods of estimating the breeding values or more generally of calculating the selection criterion

\[
\hat{A}_i = Y_i - \mu
\]

\[
\hat{A}^* = h^2 (Y_i - \mu)
\]

It is easy to show that the prediction error variance (variance of difference between estimated and true breeding value) of the first estimator is \((1-h^2) \sigma^2 \) and of the second \( h^2 (1-h^2) \sigma^2 \). However the same animals will be selected and thus the same genetic progress is realized. It should be noted that the correlation between true and estimated breeding value is the same in both cases.

In the next example (with known \( \mu, \sigma^2 \) and \( \sigma^2 \)) the advantage of using the a priori knowledge and the relationship of various criteria to the genetic progress is demonstrated.

Example 5: There are two (in some cases completely identical) populations available, which have mean \( \mu_1 \) and \( \mu_2 \) and the same \( \sigma^2 \) and \( \sigma^2 \). Out of each population a large sample of sires is randomly drawn and from each sire of population 1, \( n_1 \) daughters and from each sire of population 2, \( n_2 \) daughters are tested under identical conditions. Of all sires tested a fraction of \( p_1 \) comes from population 1. The
Table 1 Characteristics of the four procedures

(see text) $h^2=0.25$, $\mu_2=-\mu_1$, $\sigma^2/4=1$; fr. = fraction selected

<table>
<thead>
<tr>
<th>$\mu_1$</th>
<th>$P_1$</th>
<th>$n_1$</th>
<th>$n_2$</th>
<th>fr.</th>
<th>Proced.</th>
<th>$\mathcal{P}_{\mathcal{Q}}$</th>
<th>$\mathcal{P}_{\mathcal{Q}_0}$</th>
<th>MSE</th>
<th>$\Delta G$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.5</td>
<td>20</td>
<td>20</td>
<td>0.2</td>
<td>1,3,4</td>
<td>0.756</td>
<td>0.756</td>
<td>0.43</td>
<td>1.058</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>0.756</td>
<td>0.756</td>
<td>0.75</td>
<td>1.058</td>
</tr>
<tr>
<td>-0.5</td>
<td>0.5</td>
<td>20</td>
<td>20</td>
<td>0.2</td>
<td>1</td>
<td>0.756</td>
<td>0.811</td>
<td>0.43</td>
<td>1.270</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>0.756</td>
<td>0.791</td>
<td>0.75</td>
<td>1.239</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3,4</td>
<td>0.756</td>
<td>0.791</td>
<td>0.47</td>
<td>1.239</td>
</tr>
<tr>
<td>-1.0</td>
<td>0.5</td>
<td>20</td>
<td>20</td>
<td>0.2</td>
<td>1</td>
<td>0.756</td>
<td>0.886</td>
<td>0.43</td>
<td>1.730</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>0.756</td>
<td>0.853</td>
<td>0.75</td>
<td>1.688</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td>0.756</td>
<td>0.853</td>
<td>0.61</td>
<td>1.688</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4</td>
<td>0.756</td>
<td>0.853</td>
<td>0.55</td>
<td>1.688</td>
</tr>
<tr>
<td>-1.5</td>
<td>0.5</td>
<td>20</td>
<td>20</td>
<td>0.2</td>
<td>1</td>
<td>0.756</td>
<td>0.932</td>
<td>0.43</td>
<td>2.230</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>0.756</td>
<td>0.901</td>
<td>0.75</td>
<td>2.230</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td>0.756</td>
<td>0.901</td>
<td>0.84</td>
<td>2.230</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4</td>
<td>0.756</td>
<td>0.901</td>
<td>0.61</td>
<td>2.230</td>
</tr>
<tr>
<td>0</td>
<td>0.8</td>
<td>10</td>
<td>30</td>
<td>0.2</td>
<td>1</td>
<td>0.730</td>
<td>0.730</td>
<td>0.47</td>
<td>1.020</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>0.707</td>
<td>0.707</td>
<td>1.00</td>
<td>0.997</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3,4</td>
<td>0.730</td>
<td>0.730</td>
<td>0.47</td>
<td>1.020</td>
</tr>
<tr>
<td>0</td>
<td>0.8</td>
<td>10</td>
<td>30</td>
<td>0.2</td>
<td>1</td>
<td>0.673</td>
<td>0.673</td>
<td>0.55</td>
<td>0.940</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>0.659</td>
<td>0.659</td>
<td>1.30</td>
<td>0.925</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3,4</td>
<td>0.673</td>
<td>0.673</td>
<td>0.55</td>
<td>0.940</td>
</tr>
<tr>
<td>-0.5</td>
<td>0.5</td>
<td>10</td>
<td>30</td>
<td>0.2</td>
<td>1</td>
<td>0.730</td>
<td>0.792</td>
<td>0.47</td>
<td>1.304</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>0.707</td>
<td>0.745</td>
<td>1.00</td>
<td>1.178</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td>0.730</td>
<td>0.767</td>
<td>0.53</td>
<td>1.270</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4</td>
<td>0.730</td>
<td>0.767</td>
<td>0.52</td>
<td>1.263</td>
</tr>
<tr>
<td>-0.5</td>
<td>0.2</td>
<td>10</td>
<td>30</td>
<td>0.2</td>
<td>1</td>
<td>0.783</td>
<td>0.816</td>
<td>0.39</td>
<td>1.238</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>0.767</td>
<td>0.790</td>
<td>0.70</td>
<td>1.182</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td>0.783</td>
<td>0.792</td>
<td>0.44</td>
<td>1.225</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4</td>
<td>0.783</td>
<td>0.793</td>
<td>0.43</td>
<td>1.222</td>
</tr>
<tr>
<td>-0.5</td>
<td>0.2</td>
<td>30</td>
<td>10</td>
<td>0.2</td>
<td>1</td>
<td>0.673</td>
<td>0.727</td>
<td>0.55</td>
<td>1.017</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>0.659</td>
<td>0.687</td>
<td>1.30</td>
<td>1.017</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td>0.673</td>
<td>0.715</td>
<td>0.57</td>
<td>1.005</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4</td>
<td>0.673</td>
<td>0.713</td>
<td>0.57</td>
<td>1.007</td>
</tr>
<tr>
<td>-0.5</td>
<td>0.8</td>
<td>10</td>
<td>30</td>
<td>0.2</td>
<td>1</td>
<td>0.673</td>
<td>0.727</td>
<td>0.55</td>
<td>1.159</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>0.659</td>
<td>0.687</td>
<td>1.30</td>
<td>1.062</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td>0.673</td>
<td>0.715</td>
<td>0.57</td>
<td>1.133</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4</td>
<td>0.673</td>
<td>0.713</td>
<td>0.57</td>
<td>1.129</td>
</tr>
<tr>
<td>-0.5</td>
<td>0.8</td>
<td>30</td>
<td>10</td>
<td>0.2</td>
<td>1</td>
<td>0.783</td>
<td>0.816</td>
<td>0.39</td>
<td>1.225</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>0.767</td>
<td>0.790</td>
<td>0.70</td>
<td>1.213</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td>0.783</td>
<td>0.792</td>
<td>0.44</td>
<td>1.167</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4</td>
<td>0.783</td>
<td>0.793</td>
<td>0.43</td>
<td>1.172</td>
</tr>
<tr>
<td>-0.5</td>
<td>0.8</td>
<td>10</td>
<td>30</td>
<td>0.05</td>
<td>1</td>
<td>0.673</td>
<td>0.727</td>
<td>0.55</td>
<td>1.858</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>0.659</td>
<td>0.687</td>
<td>1.30</td>
<td>1.521</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td>0.673</td>
<td>0.715</td>
<td>0.57</td>
<td>1.821</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4</td>
<td>0.673</td>
<td>0.713</td>
<td>0.57</td>
<td>1.807</td>
</tr>
</tbody>
</table>
genetic progress is measured as the difference of the expected mean (of transmitting ability) of the selected animals and of all tested animals (given $p_1$). For this situation $a_1$ can be regarded as transmitting ability and we have $\mu_i+a_{ij}\sim N(\mu_i,\sigma^2_A/4)$. Now everything is scaled in such a way that $\sigma^2_A/4$ is equal to one. Compared are the following four estimation procedures

Procedure 1
$$\hat{\gamma}_{ij} = \mu_i + a_{ij} = \mu_i + \frac{\sigma^2_A/4}{\sigma^2_A/4+\sigma^2_e/n_i} (\bar{y}_{ij} - \bar{\mu})$$

Procedure 2
$$\hat{\gamma}_{ij} = \bar{y}_{ij}.$$  

Procedure 3
$$\hat{\gamma}_{ij} = \bar{\mu} + \frac{\sigma^2_A/4}{\sigma^2_A/4+\sigma^2_e/n_i} (\bar{y}_{ij} - \bar{\mu})$$

Procedure 4
$$\hat{\gamma}_{ij} = \bar{\mu} + \frac{\sigma_g^2 + \sigma^2_A/4}{\sigma_g^2 + \sigma^2_A/4+\sigma^2_e/n_i} (\bar{y}_{ij} - \bar{\mu})$$

$$\bar{\mu} = p_1\mu_1 + p_2\mu_2$$

Procedure 1 and 2 have already been discussed, whereas with procedure 3 we regard all tested sires as a random sample from one population (which however is a mixture of two normal distributions) but we still use the within subpopulation variances. Procedure 4 is similar to procedure 3 but here we use the "correct" variances, since the genetic variance within the composite population is $\sigma_g^2+\sigma^2_A/4$. Procedures 3 and 4 do not take into account which subpopulation a bull is coming from. Thus the a priori information is not used in the best way. Procedures 3 and 4 are very similar to the situation where we have sires from several years or where crossing is being carried out and we use the contemporary comparison method or a BLUP procedure without genetic groups and/or relationship matrix.

In order to characterize the four procedures a number of different criteria were calculated: genetic progress, MSE, and the correlation between estimate and true breedings value calculated within and over the subpopulations.

In table 1 these characteristics of the four procedures are given for a number of different situations. It should be remembered that in these very simple situations the animals are all in one testing environment (no systematic environmental effect) and that $\mu_i, \sigma^2_A$ and $\sigma^2_e$ are known.

Some of the more important observations from table 1 are:

- If there is an equal amount of information on each animal available and $\mu_1=\mu_2$ then the same genetic gain is achieved with all four procedures.
- If there is an equal amount of information available on each animal and $\mu_1\neq\mu_2$ then the superiority of procedure 1 over the others depends on the difference $\mu_2-\mu_1$. As mentioned they are equal at $\mu_2-\mu_1$, and the superiority of procedure 1 increases as $\mu_2-\mu_1$ gets
larger up to a certain point (which depends on selection intensity) then they become more equal again (by testing Charolais and Holstein Friesian bulls for milk yield under the same conditions any of the four procedures would most likely pick only bulls of the HF-subpopulation).

With unequal information procedure 1 is always as good as or better than the others, but among the remaining procedures neither one is uniformly better than the others.

Under not quite unrealistic conditions the difference in the genetic gains are substantial (up to 15 %).

None of the usual optimality criteria has a simple straightforward relationship to genetic progress. This is quite obvious since the relative efficiency is even dependent on the selection intensity.

The greatest difference between procedure 1 and 2 occurs if from the better population a few bulls are tested with many daughters (imported bulls!) and the many bulls from the inferior population are tested with only few daughters.

From the form of the procedures it is obvious that the more information available (in the data) the more similar the procedures.

When is BLUP optimal?

The last example showed that the use of a priori information can be really worthwhile. As already indicated when the mean of the observation (Ey) and/or the mean of the vector to be predicted (Ev) are not known, the BLUP procedure replaces those unknown parameters by their best linear estimate.

The question of optimality will again be discussed with the help of an example.

Example 6: In order to form a new synthetic breed large random samples from several distinct local breeds were combined. Before combining them these breeds were isolated and no comparison with regard to the mean was possible, so nothing is known about the relation among the means. The within breed variance, however, should be known at least approximately. If BLUP is applied the fixed effects are estimated by the generalized least squares method, which is also the best linear unbiased estimator, and assuming normality this is also the maximum likelihood estimator which in turn is a function of a set of sufficient statistics. Thus all the available information contained in the data is utilized. For the random part (again assuming normality) the bayes estimator is used which utilizes all available information in the data and combines it in the best way with the a priori information. For the example given (with no a priori knowledge on the mean) it is therefore hard to see how the BLUP estimator could be improved (though with respect to Mean Square Error the use of some biased estimator might be even better (EFRON & MORRIS, 1975)).

We can contrast the given example with the quite common situation where we have a composite population of Simmental and Red Holstein or European Braunvieh and American Brown Swiss or sample of bulls born in different years. If we have a composite population of Simmental and Red Holstein and draw a bull at random out of each

110
subpopulation most experts would be willing to bet that with re-
spect to genetic merit for milk yield the Red Holstein is most
likely to be superior. With BLUP this knowledge is not utilized
since the estimate for the fixed effect is based only on the data
at hand. If some information on the breed difference together with
some notion of the precision (error variance) is available this
prior knowledge could be incorporated e.g. along the path used by
DEMPLFE (1977) in an alternative derivation of the BLUP estimator.
The possible improvement would quite heavily depend on the amount
of information available in the data to estimate the fixed effects.
This question is not quite academic since in situations where
crossing is going on we usually have many pure bred and F₁ bulls
but then also a few backcrosses. In such a case the data may not
provide much information on the backcross and the likely ordering
(local breed < F₂ < F₁ < imported breed) is not used. In practice
those genetic groups where there are few animals are quite often
combined using essentially an estimator similar to procedure 3 or
4 in example 5. In the extreme case where some genetic groups are
represented by one sire only, the BLUP estimate of those sires would
be a pure least squares estimate. Since the variability of these
least squares estimates is larger than that of the BLUP estimates
we would have a situation similar to example 5.
Example 7: We have a large random sample of bulls out of one large
population and in addition some bulls out of various populations
but only one bull from each. An appropriate model for the daughter
average would be \( \bar{y}_{ij} = g_i + \alpha_{ij} + e_{ij} \). Each bull is tested with
15 daughters and if there is no difference between the various sub-
populations the variance of the estimates would be slightly greater
than 0.5 (\( \sigma_e^2 = 1, \lambda = 15 \)) for the bulls belonging to the large popu-
lation and 2 for the others. Both distributions would have the same
mean. Now if 90 % of the tested bulls come from the large population
and the remaining 10 % from various populations and if 5 % of all
tested bulls are selected, then 3.5 % of the tested bulls from the
large population are selected as opposed to 18.3 % from the other
bulls.
In this case a much higher fraction of bulls is selected from
those where there is less information (all have the same number of
daughters but only for the large group of bulls is there a priori
information used). In the above situation genetic progress will be
higher if we restrict ourselves to the bulls from the large popula-
tion and select 5.6 % out of them. In practice where \( g_i \) is not known,
that conclusion cannot be drawn. Also it is difficult to see how
the estimation of the group effect (if there is only one sire in it)
could be improved except by some bayesian argument with respect to
\( g \). The obvious lesson from this example is to design the breeding
programme in such a way that this situation is avoided.

Relationship of BLUP to some other procedures

In dairy sire evaluation the daughters of the testbulls are usually
not on one station but on very many farms and the design is far from
being balanced. Assuming for simplicity that the sires are a ran-
dom sample from one population and that the daughter records are
precorrected for systematic effects (e.g. age, service period etc.)
we would have the following linear model \( y = Xh + Zu + e \); \( h \) herd
effect, \( u \) sire effect; \( \text{Var}(y) = ZG\sigma_h^2 + R\sigma_u^2 \). If we assume further
that \( G=I; R=I \) and \( \sigma_h^2/\sigma_u^2 = \lambda \) then the Mixed Model Equation (to calcu­
late BLUP estimates) simplify to

\[
\begin{bmatrix}
X'X & X'Z \\
Z'X & Z'Z + I \lambda
\end{bmatrix}
\begin{bmatrix}
\hat{h} \\
\hat{u}
\end{bmatrix}
= 
\begin{bmatrix}
X'y \\
Z'y
\end{bmatrix}
\]

Eliminating the herd effects we get

\[
\begin{bmatrix}
Z'PZ + I \lambda
\end{bmatrix}
\begin{bmatrix}
\hat{u}_{BL}
\end{bmatrix}
= Z'Py
\]

Historically, other methods were used earlier.

Contemporary comparison (CC) (ROBERTSON u. RENDEL, 1954)
This is essentially a two stage procedure where \( h \) and \( u \) are re­
garded as fixed. The reduced LS-equation for \( u \) is

\[
Z'PZ\hat{u}_{LS} = Z'Py
\]

Before the availability of modern computers this equation was diffi­
cult to solve and as a consequence an approximation was used.

Diag \( \{Z'PZ\} \hat{u} = Z'Py \)
(Diag \( \{Z'PZ\} \) means that all off-diagonal elements of \( Z'PZ \) are set to
zero).

In the second step the random nature of \( u \) was taken into account
(the a priori information) and it was calculated

\[
\hat{u}_{CC} = \frac{\text{Cov}(u,\hat{u})}{\text{Var}(\hat{u})} \hat{u}
\]

As was pointed out by POWELL a. FREEMAN (1974) this is identical to

Diag \( \{Z'PZ+I \lambda\} \hat{u}_{CC} = Z'Py \)

Regressed Least Squares (RLS)
One improvement of the CC method would be to use not an approxi­
mation to the least squares equation but to find an exact solution.
Taking into account the a priori information we calculate

\[
\hat{u}_{RLS_i} = \frac{\text{Cov}(\hat{u}_i,\hat{u}_{LS_i})}{\text{Var}(\hat{u}_{LS_i})} \cdot (\hat{u}_{LS_i} - E \hat{u}_{LS_i})
\]

In this procedure not all information is used, because for estima­
ting \( u_i \) only the LS estimate of \( u_i \) is used. But it can be shown
that there is additional information about \( u_i \) in \( \hat{u}_{LS_i} \) which is not
contained in \( \hat{u}_{LS_i} \). Therefore using all the information (by multiple
regression) we get

\[
\hat{u} = \text{Cov}(\hat{u}_i,\hat{u}_{LS}) \text{Var}(\hat{u}_{LS})^{-1} (\hat{u}_{LS} - E \hat{u}_{LS})
\]

As pointed out by DEMPFLE (1976) this estimate is identical to the
BLUP estimate. Thus by making two improvements to the CC methods

112
(using an exact LS solution and using not the simple but the multiple regression) we arrive at a BLUP solution.

Three remarks may be in order

i) It is hard to see under what circumstances regressed LS would be better than BLUP, since it does not make full use of all the information. But see also HARVEY (1979)

ii) With the CC method we implicitly assume that all sires are a random sample out of one population — with BLUP we are more flexible.

iii) With the CC method by using the approximation to the LS solution we still eliminate the effects of the herds but we pretend that each sire has the same competition from other sires (that the weighted average of sires being represented in the same herds as e.g. sire A is the same as e.g. for sire B). That would be true in a balanced complete block design but not in the usual sire testing schemes.

Further problems arising in using BLUP

All the optimal properties of BLUP are true only if the variances and covariances are known. Sometimes we have good estimates, thus we can expect that these optimal properties hold approximately. If the variances are not known then we have to distinguish several cases.

The observation vector is a random vector (e.g. no selection going on) and

i) the estimates of the variances are unrelated to the data used to estimate the breeding values.

In this case it is known that using any positive definite matrix for V we get unbiased estimates for the fixed effects and the error variance is smallest if the matrix used is equal to the true V. With regard to the random part the weights which we put on the different pieces of information are optimal only if we use V.

ii) The estimates of the variances are from the same data set. In this case the estimate of \( \hat{\beta} \) is given by

\[
\hat{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}y = (X'V(y)^{-1}X)^{-1}X'V(y)^{-1}y
\]

Since in this case V is a function of y, \( \hat{\beta} \) is not a linear function of y but a far more complicated function. KACKAR & HARVILLE (1980) point out that in realistic situations the estimate of \( k_1\hat{\beta}+k_2u \) is still unbiased.

However, in most realistic situations the observation vector is not quite a random vector since in the population selection is going on. Not using the correct variances causes not only a greater error variance but also a systematic change. This will be illustrated with a simple example.

Example 8 The environment is assumed to be constant (or else we could have a control population and use the deviation from it) and the parameters are unknown. In generation 0 we have a random sample
of individuals which are tested under standard conditions and selection is going on according to a given scheme.

The linear model for the observation in generation 0 is

\[ y_0 = X\beta + Z\mu + \varepsilon \equiv 1_0\mu + I_{A_0} + \varepsilon_0 \]

\[ A_0 \sim N(0, I_{A_0}^2); \quad \mu_0 \sim N(0, I_{\mu_0}^2); \quad y_0 \sim N(1_0\mu, \sigma_y^2) \]

\[ V_0 = I_{A_0}^2 + I_{\mu_0}^2 \quad \beta = \alpha_0^2/\sigma_{\mu_0}^2 \quad 0 < \lambda < \infty \]

In this formulation \( A \) represents the breeding values and \( \mu \) represents all other genetic and environmental factors. Since in a progeny testing situation sire usually vary in number of daughters we could simulate this by assuming that we have a trait which can be measured repeatedly.

When the records from the animals of generation 0 are available we have (though we cannot calculate it)

\[ E(A_j \mid y_0) = GV^{-1}(y_0 - 1_0\mu) \]

If we mate animals \( \ell \) and \( m \) then the expected value of the breeding value of an offspring given \( y_0 \) is \( (A_\ell + A_m)/2 \) whereas the actual breeding value is \( (A_\ell + A_m)/2 + \Delta \) where \( \Delta \) is the deviation of the breeding value from the true full sib mean. More generally we can connect the breeding values of the animals of generation \( i \) and \( j \) as

\[ A_{ij} = M_{ij}A_i + \Delta_{ij} \]

with \( E(\Delta) = 0 \) \( \quad \text{Var}(\Delta) = \sigma_{\Delta}^2/2 \) (without inbreeding)

\( M \) is a design matrix the structure of which is such that if animal \( k \) is an offspring of animal \( \ell \) and \( m \) then the row \( k \) of \( M \) has 0.5 in column \( \ell \) and \( m \) and zero everywhere else. The linear model for the observation in generation 1 is

\[ y_1 = 1_1\mu + A_1 + e_1 \]

which can be written as

\[ y_1 = 1_1\mu + M_{01}A_{01} + \Delta_{11} + e_1 \]

The vectors \( \Delta_{11} \) and \( e_1 \) are random vectors with \( E(\Delta_{11}) = 0 \). Thus for given \( A_0 \) we have \( E(A_{11} \mid A_0) = M_{01}A_{01} \) and \( E(y_1 \mid A_0) = 1_1\mu + M_{01}A_{01} \) and for given \( y_0 \) \( E(A_{11} \mid y_0) = M_{01}E(A_{01} \mid y_0) = M_{01}GV^{-1}(y_0 - 1_0\mu) \) and \( E(y_1 \mid y_0) = 1_1\mu + M_{01}GV^{-1}(y_0 - 1_0\mu) \).

It is clear that the expected values of \( A_1 \) and \( y_1 \) depend only on the values of \( \sigma_{\mu_0}^2, \sigma_{A_0}^2, \sigma_y^2 \) and on our selection strategy which is described by \( M_{01} \). We can go one step further and ask what will happen on average given a well described selection strategy since \( y_0 \) is also a random vector.

Given that we have the following basic situation, 6 \( \sigma_{\mu_0}^2 \) and 6 \( \sigma_{A_0}^2 \) are tested in generation 0; they are measured 3 times; the true parameters are \( \mu = 20; \quad \sigma_{\mu_0}^2 = 1; \quad \sigma_{A_0}^2 = 9 \) thus \( \lambda = 9 \). The two highest ranking males and likewise females are used to produce the four individuals in generation 1. Given \( y_0 \) we can write down what values we would have on average for \( A_0, A_1 \) and \( y_1 \). Instead of using any odd \( y_0 \), the \( y_0 \) which we get on average is used. If we order the individuals according to yield we would expect the following values for the animals in generation 0

114
Animal 1 2 3 4 5 6 | 7 8 9 10 11 12
Ey\(\circ\) 22.53 21.28 20.40 19.60 18.72 17.47 | 22.53 21.28 20.40 19.60 18.72 17.47
E(A_0|y_0) 0.634 .321 .101 -.101 -.321 -.634 | .634 .321 .101 -.101 -.321 -.634

and for the individuals in generation 1
Animal 13(1x7) 14(1x7) 15(2x8) 16(2x8)
E(A_i|y_0) .634 .634 .321 .321
E(y_1|y_0) 20.634 20.634 20.321 20.321

Up to this point it has nothing to do with any estimation procedure, as it is just a description of what we would have on average given \(y_0\), where for \(y_0\) we use those values which we would get again on average. With the data \(y_0\) and \(y_1\) several procedures can be applied and the estimates should agree with \(E(A_0|y_0)\) and \(E(A_1|y_0)\) since in this case these are identical to \(E(A_0|y_0, y_1)\) and \(E(A_1|y_0, y_1)\). The general estimation procedure used is

\[
\hat{\lambda} = \hat{\beta} + GV^{-1}(y-\hat{x}_E)
\]

where \(\hat{\beta}\) depends on the model and is either \(\mu\) (then \(P=0\)) or \(\delta^i=(\mu, g)\) where \(g\) is a parameter measuring the difference between animals of generation 1 and 0 and \(V^* = G + \lambda*R\) depends on the \(\lambda^*\) used in the estimation procedure. The results are given in table 2.

Comments to table 2

In rows 1 to 7 the expected results are given if we are using the "correct" model with respect to the fixed effects (only \(\mu\) in the model) and by using various \(\lambda^*\) in the computation. If the correct value of \(\lambda\) is used, all quantities of interest (\(\mu\), all \(A_i\) and average of breeding values in generation 0 and 1) come out in the correct way. If \(\lambda^*\) is not the correct value everything changes and there are changes especially in the ranking of breeding values and the most interesting aspect is the behaviour of the estimate of genetic progress. That estimate is very low at a high value of \(\lambda^*\), it is the correct value if \(\lambda^*=\lambda\) and increases in the example till \(\lambda^*\sim1.2\) then goes again to the correct value if \(\lambda^*=0\). Obviously with little information in the data and a guessed prior value of \(\lambda^*\) this estimate of the genetic progress is not very trustworthy (For a similar observation in a different set-up see THOMPSON, 1979).

In rows 8 to 11 the only change which occurred was that there was far more information on the animals of generation 1 (30 measurements).

In rows 12 to 15 it is the same set-up as 8 to 11 with the exception that after selection we get more information on animal 1 (similar to sire evaluation where the selected sires are then used quite heavily).

In rows 16 to 19 there are three measurements on each individual and the model was enlarged with another fixed effect \(g\). As can be seen this parameter is superfluous if the correct \(\lambda^*\) is used. The estimate of genetic progress is now remarkably stable but the ranking of breeding values is still changing with \(\lambda^*\).

In rows 20 to 23 the situation is somewhat different. There the records of the dams were not used in the estimation procedure. This
Table 2 Comparison of true expected values and the expected values from various procedures ($\hat{A}_0$ is always 0)

<table>
<thead>
<tr>
<th>Ref.</th>
<th>$\hat{A}^*$</th>
<th>$\mu$</th>
<th>$g$</th>
<th>$A_1$</th>
<th>$A_2$</th>
<th>$A_5$</th>
<th>$A_6$</th>
<th>$A_{13}$</th>
<th>$A_{15}$</th>
<th>$\hat{A}_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>true values</td>
<td>20.0</td>
<td>-</td>
<td>.634</td>
<td>.321</td>
<td>-.321</td>
<td>-.634</td>
<td>.634</td>
<td>.321</td>
<td>.477</td>
<td></td>
</tr>
<tr>
<td>1 100</td>
<td>20.103</td>
<td>-</td>
<td>.084</td>
<td>.040</td>
<td>-.040</td>
<td>-.077</td>
<td>.090</td>
<td>.042</td>
<td>.066</td>
<td></td>
</tr>
<tr>
<td>2 15</td>
<td>20.034</td>
<td>-</td>
<td>.441</td>
<td>.219</td>
<td>-.220</td>
<td>-.428</td>
<td>.455</td>
<td>.225</td>
<td>.340</td>
<td></td>
</tr>
<tr>
<td>3 9</td>
<td>20.000</td>
<td>-</td>
<td>.634</td>
<td>.321</td>
<td>-.321</td>
<td>-.534</td>
<td>.634</td>
<td>.321</td>
<td>.477</td>
<td></td>
</tr>
<tr>
<td>4 3</td>
<td>19.927</td>
<td>-</td>
<td>1.155</td>
<td>.608</td>
<td>-.605</td>
<td>-1.231</td>
<td>1.006</td>
<td>.536</td>
<td>.771</td>
<td></td>
</tr>
<tr>
<td>5 1.2</td>
<td>19.903</td>
<td>-</td>
<td>1.602</td>
<td>.849</td>
<td>-.848</td>
<td>-1.741</td>
<td>1.118</td>
<td>.609</td>
<td>.864</td>
<td></td>
</tr>
<tr>
<td>6 0.5</td>
<td>19.921</td>
<td>-</td>
<td>1.970</td>
<td>1.032</td>
<td>-.605</td>
<td>-2.105</td>
<td>1.027</td>
<td>.558</td>
<td>.792</td>
<td></td>
</tr>
<tr>
<td>7 0.0</td>
<td>20.000</td>
<td>-</td>
<td>2.534</td>
<td>1.284</td>
<td>-1.284</td>
<td>-2.534</td>
<td>0.634</td>
<td>.321</td>
<td>.477</td>
<td></td>
</tr>
<tr>
<td>8 15</td>
<td>20.082</td>
<td>-</td>
<td>.474</td>
<td>.218</td>
<td>-.228</td>
<td>-.436</td>
<td>.512</td>
<td>.228</td>
<td>.370</td>
<td></td>
</tr>
<tr>
<td>9 9</td>
<td>20.000</td>
<td>-</td>
<td>.634</td>
<td>.321</td>
<td>-.321</td>
<td>-.634</td>
<td>.634</td>
<td>.321</td>
<td>.477</td>
<td></td>
</tr>
<tr>
<td>10 1.2</td>
<td>19.863</td>
<td>-</td>
<td>1.514</td>
<td>.822</td>
<td>-.819</td>
<td>-1.713</td>
<td>.826</td>
<td>.485</td>
<td>.655</td>
<td></td>
</tr>
<tr>
<td>11 0</td>
<td>20.000</td>
<td>-</td>
<td>2.534</td>
<td>1.284</td>
<td>-1.284</td>
<td>-2.543</td>
<td>.634</td>
<td>.321</td>
<td>.477</td>
<td></td>
</tr>
<tr>
<td>12 15</td>
<td>20.090</td>
<td>-</td>
<td>.509</td>
<td>.213</td>
<td>-.229</td>
<td>-.437</td>
<td>.513</td>
<td>.222</td>
<td>.368</td>
<td></td>
</tr>
<tr>
<td>13 9</td>
<td>20.000</td>
<td>-</td>
<td>.634</td>
<td>.321</td>
<td>-.321</td>
<td>-.634</td>
<td>.634</td>
<td>.321</td>
<td>.477</td>
<td></td>
</tr>
<tr>
<td>14 1.2</td>
<td>19.809</td>
<td>-</td>
<td>.941</td>
<td>.866</td>
<td>-.780</td>
<td>-1.674</td>
<td>.961</td>
<td>.538</td>
<td>.700</td>
<td></td>
</tr>
<tr>
<td>15 0</td>
<td>19.857</td>
<td>-</td>
<td>.966</td>
<td>1.426</td>
<td>-1.141</td>
<td>-2.392</td>
<td>.776</td>
<td>.463</td>
<td>.620</td>
<td></td>
</tr>
<tr>
<td>16 15</td>
<td>20.000</td>
<td>.159</td>
<td>.429</td>
<td>.207</td>
<td>-.214</td>
<td>-.432</td>
<td>.593</td>
<td>.362</td>
<td>.477</td>
<td></td>
</tr>
<tr>
<td>17 9</td>
<td>20.000</td>
<td>.000</td>
<td>.634</td>
<td>.321</td>
<td>-.321</td>
<td>-.634</td>
<td>.634</td>
<td>.321</td>
<td>.477</td>
<td></td>
</tr>
<tr>
<td>18 1.2</td>
<td>20.000</td>
<td>-.886</td>
<td>1.740</td>
<td>.987</td>
<td>-.917</td>
<td>-1.810</td>
<td>.732</td>
<td>.223</td>
<td>.477</td>
<td></td>
</tr>
<tr>
<td>19 0</td>
<td>20.000</td>
<td>-.1432</td>
<td>2.534</td>
<td>1.284</td>
<td>-1.284</td>
<td>-2.543</td>
<td>.634</td>
<td>.321</td>
<td>.477</td>
<td></td>
</tr>
<tr>
<td>20 15</td>
<td>20.000</td>
<td>.318</td>
<td>.436</td>
<td>.200</td>
<td>-.214</td>
<td>-.422</td>
<td>.552</td>
<td>.402</td>
<td>.477</td>
<td></td>
</tr>
<tr>
<td>21 9</td>
<td>20.000</td>
<td>.239</td>
<td>.647</td>
<td>.307</td>
<td>-.321</td>
<td>-.634</td>
<td>.580</td>
<td>.375</td>
<td>.477</td>
<td></td>
</tr>
<tr>
<td>22 1.2</td>
<td>20.000</td>
<td>-.205</td>
<td>1.798</td>
<td>.929</td>
<td>-.917</td>
<td>-1.810</td>
<td>.651</td>
<td>.304</td>
<td>.477</td>
<td></td>
</tr>
<tr>
<td>23 0</td>
<td>20.000</td>
<td>-.477</td>
<td>2.534</td>
<td>1.284</td>
<td>-1.284</td>
<td>-2.534</td>
<td>.634</td>
<td>.321</td>
<td>.477</td>
<td></td>
</tr>
<tr>
<td>24 15</td>
<td>20.000</td>
<td>.477</td>
<td>.422</td>
<td>.214</td>
<td>-.214</td>
<td>-.422</td>
<td>.503</td>
<td>.451</td>
<td>.477</td>
<td></td>
</tr>
<tr>
<td>25 9</td>
<td>20.000</td>
<td>.477</td>
<td>.634</td>
<td>.321</td>
<td>-.321</td>
<td>-.634</td>
<td>.516</td>
<td>.438</td>
<td>.477</td>
<td></td>
</tr>
<tr>
<td>26 1.2</td>
<td>20.000</td>
<td>.477</td>
<td>1.810</td>
<td>.917</td>
<td>-.917</td>
<td>-1.810</td>
<td>.589</td>
<td>.366</td>
<td>.477</td>
<td></td>
</tr>
<tr>
<td>27 0</td>
<td>20.000</td>
<td>.477</td>
<td>2.534</td>
<td>1.284</td>
<td>-1.284</td>
<td>-2.534</td>
<td>.634</td>
<td>.321</td>
<td>.477</td>
<td></td>
</tr>
</tbody>
</table>

is again quite common in dairy sire evaluation when we have only sires in the model or also (to a slightly lesser degree) if we have sire and maternal grandsire in the model. In this situation we need $g$ since even with the correct $\hat{A}^*$ we have

$$E(A_1|y^r) \neq G^r_v^r-1[I-X^r(X^r'v^r-1x^r)^{-1}x^r'v^r-1]y^r$$

($y^r \equiv$ reduced vector of observation)

In this case $g$ represents the genetic progress brought about by selection on females.

In rows 24 to 27 information comes only from one parent and we have three measurements per animal. But in this case the numerator relationship matrix ($G$) is used only within generation (group).

With regard to the numerator relationship matrix and groups there are several points to be noted (see also QUASS & POLLAK (1981)). If the correct $\hat{A}^*$ is used and all information relevant to the selection
is included in y then there seems to be no necessity for groups. If they are included the expected value of g should be zero. But this does not mean that there is no disadvantage if we include it. After all we estimate something which need not be estimated. The related question when to include different group effects if they are not too different was investigated by KENNEDY (1981). A disadvantage of the numerator relationship matrix when it is applied to animals of several generations is that it creates essentially a closed population. If there is heavy selection the population progresses, and if there is an animal e.g. in the third generation with unknown (lost) parentage, then the expected values of the contemporaries are much above zero due to the selection in earlier generations, but for the one animal with unknown parentage the method assumes that it has an expectation of 0, the expectation of the base population. The animal has only a chance of being selected if there is a lot of information in the data.

This extremely simple example should be sufficient to indicate that the application of the BLUP procedure or similar procedures is not without pitfalls and that we are well advised to design our breeding scheme to be as good as possible in order not to rely too heavily on the validity of the estimation procedure.

Summary

In the first part of the paper there is a general discussion about the aim of estimating breeding values (to maximize genetic progress) and it is shown that there is not necessarily a simple relationship between the usual optimality criteria and genetic progress. Consideration is also given to the question of what type of information is available (a priori information and information provided by the data) and there the relationship to Bayes procedures is stressed. With the help of examples the question under which circumstances the BLUP procedure is optimal is discussed and some relationship of the BLUP procedures to older procedures are reviewed. Again with examples the problems of using BLUP are illustrated if the population is undergoing selection and the variances are not well known and/or if not all information utilized in selection is included in the estimation.

Zusammenfassung

References


