Expected responses to selection on indices which include an all-or-none trait

By J. P. Mueller and J. W. James

School of Wool and Pastoral Sciences, The University of New South Wales, Kensington, Australia.

Introduction

A widespread technique in animal breeding is to combine estimators of an individual’s breeding value into a single score or index. Cochran (1951) has shown that the best selection index is given by the (possibly non-linear) regression of breeding value on the available measurements, so long as the cumulative distribution function of index values is continuous and strongly monotone. Selection decisions and response prediction under these conditions require knowledge of the form of the regression function and the distribution of index values. When all variables follow a multivariate normal distribution, the optimum index is a linear function of measurements with coefficients given by the partial regressions of breeding value on the measurements, and the distribution of index values is normal. The index coefficients can be calculated from population parameter estimates, and selection response can be predicted using normal distribution theory.

When the assumption of multivariate normality does not apply, a linear index is not necessarily the best selection criterion and prediction of response using normal distribution theory is likely to be in error. For many quantitative traits, normality is a reasonable assumption and in most applications of selection index theory multivariate normality is assumed.

However, in at least one important case appreciable non-normality must be suspected. This is when the index includes a threshold character such as fertile or infertile, resistant or susceptible to a disease. The extent of error in prediction of response to selection using ordinary index theory in such a case has been examined in a detailed simulation study by Danell and Rönningen (1981). These authors did not find a general pattern of circumstances in which unacceptable errors occurred. However, larger differences between observed and predicted gains tended to occur when the all-or-none trait received heavier emphasis because of high economic value or high heritability, especially when the correlation between the traits was negative. James (1982) suggested that unreliability of the usual index formulae would be indicated by separation of index value distributions for the two classes. In this paper, James’ suggestion is developed further. The subpopulations should follow nearly normal distributions, and by considering selection between and within subpopulations accurate predictions should be obtainable. In addition, this approach provides a framework within which improvements on the usual selection index may be sought.

Theory

To state the problem explicitly we shall first recall the index equations for an all-or-none trait following closely the presentation by Danell and Rönningen (1981).
Let \( X \) be an continuous normally distributed trait which for convenience is measured as deviation from the mean with unit standard deviation. \( Z \) is a binomial trait which is the phenotypic expression of a normally distributed liability \( Y \), such that all individuals with a liability above the threshold show the trait \((Z = 1)\) and those below the threshold do not \((Z = 0)\). If \( k \) is the population incidence of the trait the variance of \( Z \) is \( \sigma_Z^2 = k(1-k) \). The heritability of \( Z \) is related to that of \( Y \) by the equation

\[
h_Z^2 = h_Y^2 u^2 / \sigma_Z^2
\]

where \( u \) is the ordinate of the normal curve at the threshold, while the phenotypic correlation between \( X \) and \( Z \) is given by

\[r_{PXZ} = r_{PXY} u / \sigma_Z\]

as shown by Dempster and Lerner (1950). Furthermore, the genetic correlations on both scales are equal \((r_{GZX} = r_{GXY})\) since the breeding value of \( Z \) is defined by the liability.

True breeding value \( T \) is defined on the liability scale with economic value for \( Y \) determined by the economic gain in \( Z \) achieved with a marginal change in the threshold. This gives

\[T = a_X g_X + u a_Z g_Y\]

where \( g_X \) and \( g_Y \) are breeding values for \( X \) and \( Y \), and \( a_X \) and \( a_Z \) are economic returns from unit increases in \( X \) and \( Z \) respectively.

Consider selection on an index

\[I = b_X X + b_Z Z\]

Making use of results above for the phenotypic variances and covariances of \( X \) and \( Z \), the equations for \( b_X \) and \( b_Z \) are

\[
(1) \ b_X + (r_{PXZ} \sigma_Z) b_Z = \text{cov}(XT) \\
(\sigma_Z^2) b_X + (\sigma_Z^2) b_Z = \text{cov}(ZT)
\]

From the definition of \( T \) it is not hard to see that

\[
\text{cov}(XT) = h_Y^2 a_X + r_{GXY} h_X h_Y u a_Z \\
\text{cov}(ZT) = r_{GXY} h_X h_Y u a_X + h_Y^2 u^2 a_Z
\]

Since \( g_Z = u g_Y \) it is not surprising that

\[\text{cov}(ZT) = u \text{cov}(YT)\]

The index equations can readily be solved, and the regression of \( T \) on \( I \) is one. Writing \( s(p) \) as the standardised selection differential when a proportion \( p \) is selected from a normal distribution, the usual prediction of response in breeding value is

\[\Delta T = s(p) \sigma_I\]

where

\[
\sigma_I^2 = b_X \text{cov}(XT) + b_Z \text{cov}(ZT).
\]

\( \Delta T \) is often expressed as a fraction of the maximum possible gain for a given selection intensity \( s(p) \sigma_I \). The ratio \( \sigma_I / \sigma_T \) equals the correlation between index and breeding value.

**Prediction of response from subpopulations**

Let the subscript \( i \) refer to subpopulations for which \( Z = i \), and let \( T_i \) denote the mean breeding value in each subpopulation. Since the initial means of \( X \) and \( Y \) are taken as zero, overall mean breeding value is zero and we thus have
\begin{align*}
\bar{T}_1 &= s(k) \text{cov}(YT) \\
\bar{T}_o &= -s(1-k) \text{cov}(YT)
\end{align*}

If proportions \( p_i \) are chosen within the subpopulations on the basis of index value \( I_i \), the mean breeding values of those selected in subpopulations \( i \) will be

\[ \bar{T}_i + s(p_i)\beta_{TI_i} \sigma_i \]

where \( \sigma_i \) is the standard deviation of index values in the \( i \)th group and \( \beta_{TI_i} \) is the regression of breeding value on index within the group. In this equation it is assumed that the distribution of \( T \) and \( I \) will be close enough to bivariate normal for linearity of regression and normal selection differentials to apply. Provided correlations with \( Y \) are not very high, this approximation is known to be good (Cochran 1951). The total proportion selected is

\[ p = p_k k + p_o (1-k). \]

The overall selection response will then consist of two components: one is due to selection between subpopulations and is given by

\[ p \Delta T_B = p_k k \bar{T}_1 + p_o (1-k) \bar{T}_o \]

while the other is due to selection within subpopulations, and is given by

\[ p \Delta T_W = p_k k s(p_i)\beta_{TI_i} \sigma_i + p_o (1-k)s(p_o)\beta_{TI_o} \sigma_o. \]

Then, the overall response is

\[ \Delta T = \Delta T_B + \Delta T_W. \]

Selection within subpopulations clearly is on \( X \), since within a group index values differ by \( b_X \) times the difference in \( X \). Now

\[ \beta_{TI_i} = \text{cov}(TI_i)/\sigma_i^2 = b_X \text{cov}(TX_i)/b_X \sigma_X^2. \]

Now if \( t(q) \) represents the standardised deviation from the mean of a threshold corresponding to a fraction \( q \), it is well known that the variance within the fraction above the threshold is given by \( \sigma_X^2 (1-r_{PX_i}^2 F_i) \) where \( F_i = s(q) (s(q) - t(q)) \). Letting

\[ F_1 = s(k) (s(k) - t(k)) \]
\[ F_0 = s(1-k) (s(1-k) + t(k)) \]

we have

\[ \sigma_X^2 = \sigma_X^2 (1-r_{PX_i}^2 F_i). \]

In the same way

\[ \text{cov}(TX_i) = \text{cov}(XT) - r_{PXY} \text{cov}(YT) F_i \]

since \( \sigma_X = \sigma_Y = 1 \).

We find the response within a subpopulation as

\[ \Delta T_{W_i} = s(p_i) \text{cov}(TX_i)/\sigma_X \]

and therefore we can calculate the response over both subpopulations from

\[ p \Delta T_W = p_1 k s(p_i) \text{cov}(TX_i)/\sigma_X + p_o (1-k)s(p_o)\text{cov}(TX_o)/\sigma_X. \]

It is thus possible to predict selection response for any method which uses \( Z \) and \( X \) as a basis for selection decisions.
If selection is on \( I \), then truncation of the combined distribution will take place at a point where

\[ I_1 = t(p_1)\sigma_{I_1} = I_o + t(p_o)\sigma_{I_o}, \]

with \( p_o \) and \( p_1 \) satisfying the condition that a fraction \( p \) be selected in total. The mean index values are

\[ I_1 = b_X s(k)r_{PXY} + b_Z \]

\[ I_o = -b_X s(1-k)r_{PXY}. \]

In predicting response to index selection it would normally be necessary to locate the appropriate truncation point by trial and error. While it may not be immediately obvious, it can be shown that

\[ I_1 - I_o = \bar{T}_1 - \bar{T}_o, \]

so that we could as well use \( \bar{T}_1 \) for \( I_1 \) to find \( p_1 \).

We are now in a position to ask, and to answer the question, whether it is possible to improve on index selection. We find values of \( p_1 \) and \( p_o \) subject to the restriction that a total fraction \( p \) be selected, which maximise \( \Delta T \). This is done by equating to zero the derivative with respect to \( p_1 \) and \( p_o \) of

\[ \Delta T = \lambda(p_1k + p_o(1-k) - p) \]

where \( \lambda \) is a Lagrange multiplier. The solution of these equations is given by

\[ \bar{T}_1 = t(p_1)\beta_{TX}\sigma_{X_1} = \bar{T}_o + t(p_o)\beta_{TX}\sigma_{X_o}. \]

It can be seen that this differs from the usual index by the fact that separate \( \beta_{TX} \) regressions are used in the optimum procedure, while a single \( \beta_{TX} \) is used in the index, giving between populations a nonlinear response on the \( X \) scale, though assuming linear relations within classes.

**Results**

The suitability of our prediction method, using responses to selection between and within subpopulations, for index selection has been evaluated using Table 9 of Danell and Rönningen (1981). These workers listed realised deviations from response expected on normal index theory obtained in their simulations which included a number of economic weights, heritabilities and correlations, and for which incidences of the all-or-none trait and proportion selected also varied. The indexes they used were based on simulated parameters rather than the true values used in our analyses. Nevertheless the agreement of their simulations and our predictions is quite good (Table 1). In some cases, especially where agreement between our predictions and their simulations was not so good, further simulations, using theoretical parameter values, were run. These simulations showed closer agreement with our predictions than did those of Danell and Rönningen, possibly because of our use of theoretical parameters, but perhaps purely through sampling fluctuations. The prediction method is not exact, since it assumes \( X \) and \( T \) remain jointly normally distributed in the subpopulations, whereas depending on the correlations between \( Y, X, \) and \( T \), a more or less skewed distribution is expected. However, the results in Table 1 confirm the expectation that this source of error should generally be negligible.

Response to optimal selection has been calculated for the same set of conditions as in Table 1. The maximum improvement on the usual selection index is 4%. This small difference was expected from the following considerations. Examining the equations for the truncation points it will be seen that the proportions \( p \); in the optimum procedure will differ from the \( p \) resulting from selection on \( I \) only when \( \beta_{TI} \) is different from \( \beta_{TL} \). When \( r_{PXY} = 0 \) or \( k = 0.5 \) it is obvious that \( \beta_{TI} = \beta_{TL} = b_X \) and index and optimum selection procedure
Predicted (P) and simulated (S) deviations of response to selection on indices with different parameter combinations for a continuous trait X and an all-or-none trait Z (depending on a continuous liability trait Y) expressed as a percentage of response predicted by ordinary index theory. Predicted results are based on the theory described in the text and simulation results are taken from Danell and Rönningen (1981).

<table>
<thead>
<tr>
<th>( r_{PXY} )</th>
<th>( r_{GXY} )</th>
<th>( h_X^2 )</th>
<th>( h_Y^2 )</th>
<th>( \text{relative economic weights} (a_{x,y})_{1:5} )</th>
<th>( \text{relative economic weights} (a_{x,y})_{1:12} )</th>
<th>( \text{relative economic weights} (a_{x,y})_{1:15} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( F_{PXY} )</td>
<td>( F_{GXY} )</td>
<td>0.8</td>
<td>0.1</td>
<td>0.1</td>
<td>0.5</td>
<td>0.1</td>
</tr>
<tr>
<td>( F_{PXY} )</td>
<td>( F_{GXY} )</td>
<td>0.1</td>
<td>0.5</td>
<td>0.5</td>
<td>0.1</td>
<td>0.5</td>
</tr>
<tr>
<td>( F_{PXY} )</td>
<td>( F_{GXY} )</td>
<td>0.4</td>
<td>0.1</td>
<td>0.1</td>
<td>0.4</td>
<td>0.1</td>
</tr>
<tr>
<td>( F_{PXY} )</td>
<td>( F_{GXY} )</td>
<td>0.0</td>
<td>0.1</td>
<td>0.1</td>
<td>0.2</td>
<td>0.1</td>
</tr>
<tr>
<td>( F_{PXY} )</td>
<td>( F_{GXY} )</td>
<td>-0.4</td>
<td>0.1</td>
<td>0.1</td>
<td>0.3</td>
<td>0.1</td>
</tr>
<tr>
<td>( F_{PXY} )</td>
<td>( F_{GXY} )</td>
<td>-0.8</td>
<td>0.1</td>
<td>0.1</td>
<td>0.6</td>
<td>0.1</td>
</tr>
</tbody>
</table>

Coincide. It can also be seen that for the β’s being different the ratio \( \text{cov}(YT)/\text{cov}(XT) \) must be different from \( r_{PXY} \) together with extreme incidence levels to get appreciable difference in the F’s. Such conditions would also be responsible for the separation of the subpopulations, increasing the contribution of the between group (ΔTₖ) component of selection response. Since the optimum procedure corrects for the regression within groups, the circumstances which favour the use of the optimum procedure are also those which restrict the expression of its results.
Discussion

Prediction of response to selection using our method requires more complex calculations than the use of the standard selection index equations, particularly since the proportions $p_i$ and $p_o$ must be located iteratively, except in special cases. Although these calculations are not very onerous when done by computer, it is of interest to know the circumstances in which the simpler prediction method is not adequate. The difference in mean index values, $I_1 - I_o$, is in fact a good indicator of possible errors, as suggested by James (1982). This difference depends largely on $\text{cov}(YT)$. For instance, in the cases considered in Table 1, all prediction errors of more than 10% occurred when $(I_1 - I_o)/\sigma_1$ was greater than 1.85. Not all cases where $(I_1 - I_o)/\sigma_1$ was large produced important prediction errors, but it does seem safe to conclude that if this ratio is small, there is little risk that important prediction errors will occur if the usual selection index theory is applied.

The error in the usual prediction depends on the relationship between $p$ and $k$. Figure 1 illustrates the use of the prediction method in a special case, and this can be used to show

![Graph](image)

Example of the distribution of index values in a simulated population (a) and the normal distribution of index values assumed in ordinary index theory (b). One thousand index values were drawn from a population in which $h^2 = 0.1$, $h^2 = 0.5$, $r_{XY} = -0.8$, $a_1 = 1$, $a_2 = 15$ and $k = 0.5$.

some features of prediction errors. There is a general tendency for gain to be underestimated when $p = k$ and overestimated when $p \neq k$. In Figure 1, we see that as $p$ departs from zero, the tail of the single index distribution first predicts the existence of candidates where there are none, overestimating gain. At $p = k = 0.5$, gain is underestimated because the two distributions are appreciably separated. As $p$ increases further, gain tends to be overestimated since many candidates are expected where there are in fact only the two tails of the subgroups.
The detailed analysis here has been of a single continuous trait and an all-or-none trait. The results would apply, however, to the case of a single all-or-none trait and a number of continuous traits combined in an index, which would play the role of X in our discussion. The case of two or more threshold traits may benefit from further analysis, though it seems highly likely that the general conclusions would be similar to those in our analyses.

It is sometimes of interest to predict the response in each trait in the economic objective, and even of other traits. For example, one may wish to know how much is gained in wool traits and how much in lambing rate when selection is on an index combining these traits. The gains would usually be predicted from $s(p)b'G/\sigma_2$ where $b'$ is the row vector of index coefficients and $G$, is the genetic covariance matrix. The use of our prediction method is straightforward. Instead of covariances involving T in predicting response, it is only necessary to use genetic covariances involving X or Y as required.

Acknowledgement
J.P.M. wishes to acknowledge the tenure of a studentship by the Instituto Nacional de Tecnología Agropecuaria, INTA (Argentina).

Summary
A procedure to allow prediction of response to selection on an index involving an all-or-none trait is developed. Comparison with the results of simulation experiments shows that such predictions are accurate. The usual selection index prediction formula may become inaccurate when the two subpopulations of animals with and without the trait have distributions which have little overlap. An improved selection procedure allowing for the non-linearity introduced by the all-or-none trait is derived, but it is shown that the usual index is never much less efficient.

Résumé
Sucess de sélection esperé d'indices avec caractères à seuil
Un procédé est développé permettant la prédiction du succès d'un index de sélection avec caractères à seuil. La comparaison avec les résultats des expériences de simulation montre que de telles prédictions sont précises. La formule généralement utilisée pour l’estimation du succès d’un index de sélection peut devenir inexacte, lorsque deux sous-populations d’animaux avec ou sans le caractère en question, ont des distributions s’entrecoupant légèrement. Une méthode de sélection améliorée permet la prise en considération de la non-linéarité, causée par les caractères à seuil, cependant il se montre, que l’index usuel n’est jamais beaucoup plus moins précis.

Resumen
Respuesta a la selección por índices que incluyen una característica de distribución binomial
Se presentan fórmulas que permiten estimar la respuesta a la selección cuando ésta se basa en índices que incluyen una característica de distribución binomial. Comparaciones con resultados de ensayos de simulación demuestran que las predicciones basadas en el método propuesto son exactas aun cuando los índices de las dos subpoblaciones tienen distribuciones marcadamente separadas, en cambio la metodología habitual, que asume normalidad, puede arrojar resultados erróneos. Por otro lado se desarrolla un método de selección que toma en cuenta la no-linealidad introducida por la característica en cuestión, sin embargo, la eficiencia de este método no es muy superior a la del índice habitual.

Zusammenfassung
Erwartete Selektionsfortschritte bei Verwendung von Indizes mit Schwelleneigenschaften
Es wird eine Methode entwickelt, die Schätzung des Selektionserfolges aufgrund eines Index mit Schwelleneigenschaften erlaubt. Der Vergleich mit Resultaten eines Simulationsversuches erweist die Schätzungen als genau. Die gewöhnliche Formel zur Schätzung des Erfolges von Indexselektion kann ungenau werden, wenn die zwei Subpopulationen mit und ohne der betreffenden Eigenschaft Verteilungen haben, die wenig Überlappung zeigen. Es wird eine verbesserte Selektionsmethode abgeleitet, die die Nichlinearität, welche durch Schwelleneigenschaften verursacht wird, berücksichtigt. Es zeigt sich aber, daß der gewöhnlich verwendete Index nie sehr viel weniger effizient ist.
Expected responses to selection on indices which include an all-or-none trait

Literature


Authors’ addresses: J. W. JAMES; School of Wool and Pastoral Sciences, The University of New South Wales, P. O. Box 1, Kensington, Australia, 2033. J. P. MUELLER: Instituto Nacional de Tecnologia Agropecuaria, Casilla de Correo 277, San Carlos de Bariloche 8400, Argentina