A comparison of exact and sequential methods in multi-stage index selection*

A. M. Saxton
Department of Animal Science, P.O. Box 5127, North Carolina State University, Raleigh, NC 27650, USA

Received February 5, 1982; Accepted February 15, 1983
Communicated by D. Van Vleck

Summary. The theory of sequential multi-stage index selection makes an implicit assumption that the correlation between indices at different stages is zero. This assumption was shown to result in errors in the estimation of genetic gain and in the proportion of the population selected by truncating the joint distribution of the indices. Knowledge of the means and volumes of truncated multivariate normal distributions was used to correct these estimates. Effects of selection intensity and the correlation between the first and second stage indices (\(\rho\)) on the accuracy of the approximate sequential method were examined. Computational constraints limited this analysis to two-stage index selection procedures. The sequential method performed well for \(\rho\) less than 0.6 but accuracy deteriorated rapidly as \(\rho\) increased beyond this value. The effect of selection intensity on accuracy was smaller than \(\rho\). On a percentage basis, errors in actual percent selected and underestimation of genetic gain increased with selection intensity while overestimation decreased. The types of errors which occur and their magnitude depend on the intensity of first stage selection.

Key words: Independent culling – Multi-stage selection – Sequential selection

Introduction

Multiple stage selection procedures are an important technique in animal breeding, and are often used in practice without regard for the theoretical implications. For example, an initial culling often occurs before putting animals on a more complex test procedure, due to limited facilities. The effect of this initial culling on the variances and on the prediction of gains could be significant. The initial culling should then be considered as part of the selection program, and a two-stage selection program results. While the theoretical foundations of multi-stage selection are still being developed, many situations can be analyzed with present knowledge.

The idea of several stages of selection appears to have begun with the work of Dickerson and Hazel (1944), who explored the use of individual or family performance information in the first stage and progeny test information in the second stage for improving a single trait. Jain and Amble (1962) extended single-trait selection to an arbitrary number of stages with no restriction on the type of information used at each stage, using the theory developed by Cochran (1951). Namkoong (1970) considered the problem of optimum allocation of selection intensity in two stages of selection for a single trait, with the addition of a cost function to account for the costs of obtaining the second stage information.

When two traits are being selected, the theory developed for independent culling by Young and Weiler (1960) can be used. Predicted gains are calculated by assuming a simultaneous truncation selection on each of the two traits, but in practice the selections are often performed at different times, particularly if the traits are measured at different ages. Young (1964) extended the theory to permit selection of more than two traits and to allow, within each stage, selection of more than one trait. Young (1964) also introduced the idea of using part and whole indices for the two traits to be independently culled, and stated that expected gain would be greatest when all intensity
was on the second stage, where the whole index was being used as the criterion for selection. The use of such part and whole indices is expected to be more efficient than the usual independent culling method because the indices use all available information.

In contrast to the ‘exact’ solution given by independent culling theory, several authors have developed an approximate sequential method for selection in several stages. Cotterill and James (1981) put independent culling selection in a sequential two-stage form and considered multiple sources of information at the second stage. Cunningham (1975) used the part and whole index concept of Young (1964) to develop a method for sequential multi-stage index selection. Using Cunningham’s algorithm, Bartels et al. (1980) found that the relative efficiency of multi-stage index selection could be greater than the classical single-stage selection index. This result is contrary to the intuitive expectation of Young (1964).

The idea of sequential selection was developed to ease the computational burden created by multivariate distributions, but results in estimates of genetic gain which are known to be approximations. The present paper will examine how close these approximations are by comparing the algorithm of Cunningham (1975) with exact results obtained from a modified computer program for independent culling selection (Saxton 1982).

### Multi-stage index selection

For convenience, the theory of sequential and exact selection methods will be briefly reviewed. Following the development of Cunningham (1975) let

- \( P \) be the \( n \) by \( n \) phenotypic variance-covariance matrix,
- \( G \) be the \( n \) by \( m \) covariance matrix between the \( n \) traits in the index and the \( m \) traits in the aggregate genotype (\( H \)), and let
- \( C \) be the \( m \) by \( m \) genotypic variance-covariance matrix.

Selection procedure 4 suggested by Cunningham (1975) will be used throughout this paper. This procedure selects in the first stage upon an index \( (I_1) \) incorporating a subset of the \( n \) traits, without loss of generality assumed to be the first \( r \) traits. In the second stage selection is on an index \( (I_2) \) incorporating all \( n \) traits.

In sequential selection, a proportion of the population, \( p_1 \), is selected by truncation selection on \( I_1 \), using the usual univariate selection theory. Variances and covariances are adjusted for this first stage selection (Cochran 1951). Truncation selection on \( I_2 \) is then performed, selecting a fraction \( p_2 \) of the remaining population, such that \( p_1 p_2 = S \), where \( S \) is the desired fraction of the population to be selected. Univariate theory is used for this second stage selection also, and thus the complexities of truncated multivariate normal distributions are avoided. This requires an assumption that the bivariate normality between \( I_1 \) and the aggregate genotype is not seriously affected by selection on \( I_2 \) or equivalently, that the phenotypic correlation between \( I_1 \) and \( I_2 \) is zero. Total gain is calculated by adding the estimates of gain from the two stages.

### Table 1. Variances and covariances among the selection indices \((I_1, I_2)\) and the aggregate genotype \((H)\)

<table>
<thead>
<tr>
<th>Definition</th>
<th>Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variance of ( H )</td>
<td>( v' C v )</td>
</tr>
<tr>
<td>Variance of ( I_1 )</td>
<td>( b_1 P_1 b_1 )</td>
</tr>
<tr>
<td>Variance of ( I_2 )</td>
<td>( b_2 P_2 b_2 )</td>
</tr>
<tr>
<td>Covariance between ( I_1 ) and ( I_2 )</td>
<td>( b_1 P_2 b_2 )</td>
</tr>
<tr>
<td>Covariance between ( I_1 ) and ( H )</td>
<td>( b_1 G_1 v )</td>
</tr>
<tr>
<td>Covariance between ( I_2 ) and ( H )</td>
<td>( b_2 G_2 v )</td>
</tr>
</tbody>
</table>

\( P^* \) is the \( n \) by \( r \) submatrix of \( P \)

Exact results are obtained by using independent culling selection on the two indices. For the first stage, index weights are given by

\[
b_1 = P_1^{-1} G_1 v,
\]

where \( P_1 \) is the \( r \) by \( r \) matrix consisting of the first \( r \) rows and columns of \( P \), \( G_1 \) is the \( r \) by \( m \) matrix containing the first \( r \) rows and all \( m \) columns of \( G \) and \( v \) is the vector of economic weights. The second stage index weights are of course given by

\[
b_2 = P_2^{-1} G_2 v.
\]

Parameters needed for independent culling can now be calculated according to the formulae in Table 1. Finally, define the fraction selected as

\[
S = \lambda (2\Pi)^{-\frac{r}{2}} \int \int \exp \left[-\frac{1}{2} \left[x^2 + \frac{1}{2} (xy + y^2)\right]\right] dx dy
\]

where

\[
\lambda = (1-q^2)^{-\frac{1}{2}},
\]

\( q \) = correlation between \( I_1 \) and \( I_2 \),

\( t_1 \) = truncation point for \( I_1 \), and

\( t_2 \) = truncation point for \( I_2 \).

A computer program was written which sequentially took values of \( t_2 \) over its possible range, calculated the \( t_2 \) which results in selecting a fraction \( S \) of the population (Saxton 1982) and then computed the gain in the aggregate genotype resulting from truncating \( I_1 \) and \( I_2 \) at \( t_1 \) and \( t_2 \), respectively (Tallis 1961). These results are exact, given the common assumptions of initial multivariate normality and an infinite population.

The exact and sequential methods differ only in the assumption of normality in the second stage population required by the sequential method. Therefore, it is expected that only parameters which influence this normality will affect the accuracy of the sequential method. These parameters are the correlation between \( I_1 \) and \( I_2 \) and the intensity of selection.

### Examples

The first example illustrates the types of errors which can occur in using the sequential method. Data given in Cunningham (1975) are used and it is assumed that 6% of the population is to be selected. There are five traits available and the first stage index is based on the first three traits. Table 2 gives two points of comparison between the exact algorithm and the sequential method of Cunningham (1975). Method 1 is the usual single-