Multiple Stage Selection
Jack Dekkers

Multi-trait breeding goal: \( H = v_1g_1 + v_2g_2 + v_3g_3 + \ldots + v_ng_n = v'g \)

Information sources: \( X_1 , X_2 , X_3 , X_4 , \ldots , X_m \)

Multi-trait selection index: \( I = b_1X_1 + b_2X_2 + b_3X_3 + \ldots + b_mX_m \)

Optimal index weights: \( b = P'^{-1}G'v \)

Selection on \( I \) maximizes response to selection in \( H \), but requires all animals to be measured for all traits.

Multiple-stage selection:

Stage 1: select on \( I_1 = b_1X_1 + b_2X_2 + \ldots + b_kX_k = b_1'X_1 \)

Stage 2: select on \( I_2 = b_1X_1 + b_2X_2 + \ldots + b_kX_k + b_{k+1}X_{k+1} + \ldots + b_mX_m = b_2'X \)

Only animals that are selected in stage 1 have to be evaluated for \( X_{k+1} , \ldots , X_m \)

→ Cost savings
→ Opportunities to increase population size for early stages

Optimal index weights:

Stage 1: \( I_1 : b_1 = P_{11}^{-1}G_1v \) \[ P_{11} = \text{Var}(X_1) \] \[ G_1 = \text{Cov}(X_1, g) \]

Stage 2: \( I_2 : b_2 = P^{-1}Gv \) \[ P = \begin{pmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{pmatrix} \quad G = \begin{pmatrix} G_1 \\ G_2 \end{pmatrix} \]

Optimal weights for index \( I_2 \) are not affected by selection on \( I_1 \) in stage 1, provided all data included in \( I_1 \) is also included in \( I_2 \) (Cunningham 1975 Theor. Appl. Genet. 46:55)

But accuracy and response to selection on \( I_2 \) are affected by selection on \( I_1 \):

Stage 1: accuracy of \( I_1 \): \( r_1 = \sqrt{\frac{b_1'G_1v}{v'Cv}} \) \[ \text{Trait response vector: } S_{g,1} = i_1 \frac{b_1'G_1}{\sqrt{b_1'P_{11}b_1}} \]

Stage 2: accuracy of \( I_2 \): \( r_2 = \sqrt{\frac{b_2'G^*v}{v'C^*v}} \) \[ \text{Trait response vector: } S_{g,2} = i_2 \frac{b_2'G^*}{\sqrt{b_2'P^*b_2}} \]

This assumes multi-variate normality of variables at stage 2 (despite stage 1 selection).

Total response vector across both stages: \( S_g = S_{g,1} + S_{g,2} \)

Matrices \( P^*, G^*, \) and \( C^* \) are \( P, G, \) and \( C \) matrices adjusted for selection on \( I_1 \)
Matrix equivalent of adjustment of (co-)variance for selection on variable w (used for Bulmer effect):

\[ \sigma_{xy}^* = \sigma_{xy} - k \frac{\sigma_{wx} \sigma_{wy}}{\sigma_w^2} \]

\[ k = i(i-t) \quad t = \text{truncation point} \]

Consider vectors \( w, x, y \) Select on index \( b'w \)

\[ \text{Cov}(x,y)^* = \text{Cov}(x,y) - k \frac{\text{Cov}(x,b'w) \text{Cov}(b'w,y)}{\text{Var}(b'w)} \]

\[ = \text{Cov}(x,y) - k \frac{\text{Cov}(x,w)bb'\text{Cov}(w,y)}{b'\text{Var}(w)b} \]

With stage 1 selection on \( b'w = b_1'X_1 \) \( \rightarrow \) Matrices to use in Stage 2:

\[ P^* = \text{Var}(X)^* = \text{Cov}(X,X)^* = P - k \frac{\text{Cov}(X,X_1)b_1 b_1' \text{Cov}(X_1,X)}{b_1' \text{Var}(X_1)b_1} \]

\[ = P - k \frac{\begin{bmatrix} P_{11} \\ P_{21} \end{bmatrix} b_1 b_1' \begin{bmatrix} P_{11} & P_{21} \end{bmatrix}}{b_1' P_{11} b_1} \]

\[ G^* = \text{Cov}(X,g)^* = G - k \frac{\text{Cov}(X,X_1)b_1 b_1' \text{Cov}(X_1,g)}{b_1' \text{Var}(X_1)b_1} \]

\[ = G - k \frac{\begin{bmatrix} P_{11} \\ P_{21} \end{bmatrix} b_1 b_1' G_1}{b_1' P_{11} b_1} \]

\[ C^* = \text{Var}(g)^* = \text{Cov}(g,g)^* = C - k \frac{\text{Cov}(g,X_1)b_1 b_1' \text{Cov}(X_1,g)}{b_1' \text{Var}(X_1)b_1} \]

\[ = C - k \frac{G_1'b_1 b_1' G_1}{b_1' P_{11} b_1} \]

See 2-stage selection example.xls
Multi-stage selection with availability of multi-trait EBV:

EBV for all $m$ traits available at every stage (but with different accuracies)
  - select on complete index at every stage with weights = economic values
    \[ I = v_1 \hat{g}_1 + v_2 \hat{g}_2 + \ldots + v_n \hat{g}_n \]

Optimization of proportions selected at each stage

Total proportion selected over $s$ stages = $P$
Proportion selected at stage $i = p_i$

\[
P = \prod_{i=1}^{s} p_i
\]

\[ a_i = \text{cost of traits measured at stage } i \]

Total cost = $TC = a_1 + \sum_{i=2}^{s} a_i \prod_{j=1}^{i-1} p_j$

Maximize gain in breeding goal per unit of cost: $Q = \Delta H/TC$

Or Maximize Profit from sale of breeding stock: Profit = $NPk b \Delta H - TC$

\[ N = \text{total # animals in breeding program} \]
\[ p = \text{percent selected} \]
\[ k = \text{# times breed is multiplied before distribution} \]
\[ b = \text{slope of supply-demand curve} \]
\[ = \text{extra returns from sale of animal per extra unit genetic worth} \]

Proportions selected and measured at each stage can then be optimized based on these objective functions and associated responses to selection.

References:
Xu and Muir (1991) Genetics 129:963