# Modeling Multi-Stage Selection

## A.S. Leaflet R3331

Hailin Su, Postdoctoral Research Associate;
Jack Dekkers, Distinguished Professor, Department of Animal Science, Iowa State University

**Summary and Implications**

In this study, we introduced and applied Truncated Multivariate Normal distributions, on top of standard quantitative genetics theory and selection index theory, to reliably model the multi-stage selection breeding programs. Compared to the previously available SelAction software (Rutten et al. 2002), this new approach properly models the change in genetic parameters in multi-stage selection, and is capable of evaluating selective breeding programs with more than three selection stages.

**Introduction**

The SelAction software has been successful as a tool to predict selection response in traditional livestock breeding programs. It uses standard quantitative genetics theory and selection index theory to develop deterministic recursive equations, which model changes in trait means and variance-covariance structures due to selection, and then predicts asymptotic responses to multiple trait and multiple stage selection. However, in multi-stage selection, SelAction obtains equilibrium genetic parameters based on the so-called Bulmer effect by iterating on the index from the last stage, as if it concerned single-stage selection, which results in an under-prediction of selection response. Multiple-stage selection programs are increasingly used in animal breeding, as initial selection decisions can now be made at an early age using genomic information, followed by recording phenotypes on the selected individuals prior to final selection of animals for breeding. The objective of this study was to introduce and apply truncated multivariate normal distribution theory to properly model the change of the variance-covariance structure with multiple-stage selection breeding programs.

**Materials and Methods**

From a statistical point of view, the multiple-trait selection index ($\boldsymbol{I}$) that is used in each selection stage and the true breeding values ($\boldsymbol{BV}$) of traits that are under selection follow a multivariate normal (MVN) distribution before selection:

$$\begin{bmatrix} I \\ BV \end{bmatrix} \sim MVN\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} Var(I) & Cov(I,BV) \\ Cov(BV,I) & Var(BV) \end{bmatrix}\right)$$

Elements in the variance-covariance structure of this distribution could be obtained using standard quantitative genetics theory and selection index theory. After the last stage of selection, the genetic selection differential $\boldsymbol{R_g}$ is then equal to the expectation of the true $\boldsymbol{BV}$ given that each index in $\boldsymbol{I}$ is greater than its respective truncation point in $\boldsymbol{t}$, i.e. $\boldsymbol{R_g} = \mathrm{E}(\boldsymbol{BV} \,|\boldsymbol{I} > \boldsymbol{t})$, where $\boldsymbol{I}$ and $\boldsymbol{t}$ are vectors of selection indices and their respective truncation points for the different stages of selection. Similarly, the genetic variance structure after the last stage of selection is $\boldsymbol{G_g} = \mathrm{Var}(\boldsymbol{BV} \,|\boldsymbol{I} > \boldsymbol{t})$. The index selection differentials and their variance structure can be expressed similarly, i.e. $\boldsymbol{R_I} = \mathrm{E}(\boldsymbol{I} \,|\boldsymbol{I} > \boldsymbol{t})$ and $\boldsymbol{G_I} = \mathrm{Var}(\boldsymbol{I} \,|\boldsymbol{I} > \boldsymbol{t})$.

In the SelAction software, equilibrium genetic parameters were obtained by iterating on the index selected on in the last selection stage, as if it concerned single-stage selection, i.e. $\boldsymbol{R_I} = \mathrm{E}(\boldsymbol{I} \mid \mathrm{I_n} > \mathrm{t_n})$ and $\boldsymbol{G_I} = \mathrm{Var}(\boldsymbol{I} \mid \mathrm{I_n} > \mathrm{t_n})$, where n is number of stages, and $\mathrm{I_n}$ and $\mathrm{t_n}$ are the selection index and its corresponding truncation point for stage n, i.e. the last stage.

One way to correctly model response in multi-stage selection is to reconsider $\boldsymbol{R_g}$, $\boldsymbol{R_I}$ and $\boldsymbol{G_g}$, $\boldsymbol{G_I}$ as the expectations and variances of multivariate normal integrals. For $\boldsymbol{R_g}$ and $\boldsymbol{R_I}$,

$$R_{I_i} = \frac{1}{p}\int_{t_1}^{\infty}\int_{t_2}^{\infty}\cdots\int_{t_n}^{\infty} I_i \cdot f(I,BV)\,dI_1\,dI_2\ldots dI_n$$

where $\mathrm{I_i}$ is the selection index used at the *i*-th stage, p is the overall proportion selected, and $\mathrm{t_i}$ is the truncation point corresponding to stage *i*. Variance structures $\boldsymbol{G_g}$ and $\boldsymbol{G_I}$ of such integrals cannot be calculated analytically. However, both the expectations and variance structures of the integrals are computationally feasible based on the moment generating functions for the truncated multivariate normal distribution, using the numerical integration algorithms for multivariate normal distributions that were derived by Tallis (1961).

**Results**

The approach implemented in the SelAction software results in an under-prediction of selection response: if the indices in the different stages show a high correlation, intensive selection in earlier stages can be applied with minor loss of selection response; in contrast, if the indices have low correlations, pre-selection in earlier stages may result in removal of individuals that would have been selected in the later stages. Thus, if the indices have low correlation, one needs to be cautious about modeling multi-stage selection using SelAction.

Moment generating functions for truncated multivariate normal distributions were implemented and applied to correctly model multi-stage selection. The restriction of a maximum of 3 stages of selection in SelAction was also removed. Ongoing work focuses on incorporating genomic information when modeling selection programs.

## Conclusion

Equilibrium genetic parameters for prediction of response in multiple-stage selection programs can be obtained properly by iterating on all the indices used in multi-stage selection. The resulting software that incorporates these developments will enable prediction of response to complex breeding programs with multiple stages and genomic information, which is necessary to properly design animal breeding programs.

## Acknowledgements

This project was supported by USDA-NIFA Grant no. 2017-67015-26299 from the USDA National Institute of Food and Agriculture.