NOTE ON REPEATED SELECTION I: THE NORMAL CASE

Technical Report No. 19

Department of Navy
Office of Naval Research

Contract No. Monr-409(39)
Project No. (NR 042-212)

BIOMETRICS UNIT
DEPARTMENT OF PLANT BREEDING

NEW YORK STATE COLLEGE OF AGRICULTURE

CORNELL UNIVERSITY
ITHACA, NEW YORK
NOTE ON REPEATED SELECTION IN THE NORMAL CASE

Technical Report No. 19

Department of Navy
Office of Naval Research

Contract No. Nonr-409(39)
Project No. (NR 042-212)

D. J. Robson
Biometrics Unit
New York State College of Agriculture
Cornell University
Ithaca, New York

This work was supported in part by the Office of Naval Research. Reproduction in whole or in part is permitted for any purpose of the United States Government.
NOTE ON REPEATED SELECTION IN THE NORMAL CASE

BU-166-M

D. S. Robson

April, 1964

ABSTRACT

A $k$-cycle selection model is specified by a $(k+1)$-variate normal distribution of the variables $X, Y_1 = X + \epsilon_1, \ldots, Y_k = X + \epsilon_k$ with selection at the $i^{th}$ stage removing a fraction

$$P_i = P(Y_i > y_i \mid Y_1 > y_1, \ldots, Y_{i-1} > y_{i-1})$$

The distribution of $X$ in this selected fraction is then convolved with the $N(0, \epsilon_{i+1}^2)$ distribution of $\epsilon_{i+1}$ to form the distribution of $Y_{i+1}$. An expression is given for the characteristic function of $X$ in the $k^{th}$ selected fraction.
Selection for a quantitative trait often continues for several cycles, as in the successive annual screening of a plant population in the process of developing new varieties. With plant selection, as with most other selection problems, the trait \( x \) being selected for cannot be measured without error, and actual selections are based on the observation \( y_i = x + e_i \) in the \( i^{th} \) cycle of the process. We shall assume here that the error chance variable \( e_i \) is \( N(0, \sigma^2) \) (normally distributed with mean 0 and variance \( \sigma^2 \)) and that the error \( e_i \) attaching to \( x \) in the \( i^{th} \) stage is independent of the error \( e_j \) attaching to that same \( x \) (or any other \( x \)) in the \( j^{th} \) stage. Further, we suppose that in the unselected population the chance variable \( x \) is \( N(\xi, \sigma^2) \), so that \( y_1 = x + e_1 \) is \( N(\xi, \omega^2 = \sigma^2 + \sigma^2_1) \).

The population available at the \( k^{th} \) stage is assumed to be of infinite size, and selection consists of removing the upper fraction \( P_k \) of the available \( y \)-population for further selection at stage \( k+1 \). The fraction of the original population available for selection at stage \( k+1 \) is therefore \( P_1 P_2 \cdots P_k \), and our concern here shall lie with the distribution of \( x \) in this remaining fraction. These fractions are defined by

\[
P_1 = P(y_1 > y_1)
\]

\[
P_1 P_2 = P_1 P(y_2 > y_2 \mid y_1 > y_1)
\]

\[
P_1 P_2 \cdots P_k = P_1 P_2 \cdots P_{k-1} P(y_k > Y_k \mid y_1 > y_1, y_2 > y_2, \ldots, y_{k-1} > y_{k-1})
\]

and our results are based upon the observation that this remaining fraction is
simply the tail probability in a $k$-variate normal distribution,

$$P_{1}P_{2} \cdots P_{k} = P(Y_1 > y_1, Y_2 > y_2, \ldots, Y_k > y_k)$$

Since the joint distribution of $X, Y_1, Y_2, \ldots, Y_k$ is the $(k+1)$-variate normal distribution with mean $\xi$ and covariance matrix

$$\Lambda = \begin{bmatrix} \sigma^2 & \sigma^2 & \ldots & \sigma^2 \\ \sigma^2 & \omega^2 & \ldots & \sigma^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma^2 & \sigma^2 & \ldots & \omega^2 \end{bmatrix} = [\sigma_{ij}]$$

then the distribution of $x$ for fixed values of $Y_1, \ldots, Y_k$ is normal with mean

$$E(X|Y_1, \ldots, Y_k) = \xi - \frac{\Lambda_{11}}{\Lambda_{oo}} (y_1 - \xi) - \cdots - \frac{\Lambda_{kk}}{\Lambda_{oo}} (y_k - \xi)$$

and

$$\text{var}(X|Y_1, \ldots, Y_k) = \frac{\Lambda}{\Lambda_{oo}}$$

where $\Lambda$ is the determinant of $\Lambda$ and $\Lambda_{ij}$ is the cofactor of the $ij$th element of $\Lambda$.

The joint distribution of $Y_1, \ldots, Y_k$ is normal with mean $\xi = (\xi, \ldots, \xi)$ and covariance matrix $\Lambda_{oo}$. Using the expansion

$$\Lambda = \sigma_{oo} \Lambda_{oo} - \sum_{i,j=1}^{k} \sigma_{io} \sigma_{oj} \Lambda_{oo} \cdot ij$$

where $\Lambda_{oo} \cdot ij$ is the cofactor of $\sigma_{ij}$ in $\Lambda_{oo}$, we may then express the conditional
moment generating function of $X$ as

$$E(e^{tX} | y_1, \ldots, y_k)$$

$$= e^{t\xi + \frac{t^2}{2}(\sigma_0^2 - \frac{1}{\omega_{oo}}) \sum_{i,j=1}^{k} \sigma_{io} \sigma_{oj} \lambda_{oo,ij} + \frac{t}{\lambda_{oo}} \sum_{i,j=1}^{k} \sigma_{io} (y_j - \xi) \lambda_{oo,ij}}$$

and then

$$E(e^{tX} | y_1 > y_1, \ldots, y_k > y_k) = e^{t\xi + \frac{t^2}{2} \sigma_0^2 (P_1 P_2 \cdots P_k)^{-1} \frac{1}{(2\pi)^{\frac{k}{2}} \sqrt{\lambda_{oo}}}} \cdot$$

$$\int e^{\frac{1}{2\lambda_{oo}} \sum_{i,j=1}^{k} [(u_i - \xi)(u_j - \xi) - 2t\sigma_{io}(u_j - \xi) + t^2\sigma_{io} \sigma_{oj}] \lambda_{oo,ij}} \prod_{u_i > y_i} du_1 \cdots du_k$$

The exponent in the integral reduces to

$$\sum_{i,j=1}^{k} (u_i - \xi - \sigma_{io} t)(u_j - \xi - \sigma_{oj} t) \lambda_{oo,ij}$$

hence, transforming to the standard normal $z_i = (y_i - \xi)/\sqrt{\sigma_{ii}}$, we obtain

$$E(e^{tX} \left| \frac{y_1 - \xi}{\omega_1} > z_1, \ldots, \frac{y_k - \xi}{\omega_k} > z_k \right.)$$

$$= e^{t\xi + \frac{t^2}{2} \sigma^2} \prod_{i,j=1}^{k} P_{ij} \frac{1}{(2\pi)^{\frac{k}{2}} \sqrt{\lambda_{oo}}} \int e^{\frac{1}{2\lambda_{oo}} \sum_{i,j=1}^{k} R_{oo,ij} v_i v_j} \prod_{v_i > z_i = \frac{v_i}{\omega_i}} dv_1 \cdots dv_k$$
where

\[
R_{00} = \begin{pmatrix}
1 & \frac{\sigma^2}{\omega_1 \omega_2} & \cdots & \frac{\sigma^2}{\omega_1 \omega_k} \\
\frac{\sigma^2}{\omega_1 \omega_2} & 1 & \cdots & \frac{\sigma^2}{\omega_2 \omega_k} \\
\vdots & \vdots & \ddots & \vdots \\
\frac{\sigma^2}{\omega_1 \omega_k} & \frac{\sigma^2}{\omega_2 \omega_k} & \cdots & 1
\end{pmatrix}
\]

The mean value of X in this selected fraction of the population is obtained by differentiating once with respect to t, first writing

\[
E(e^{tX} | \frac{Y_1 - \xi}{\omega_1} > z_1, \ldots, \frac{Y_k - \xi}{\omega_k} > z_k)
\]

\[
= \phi_X(t) P_{R_{00}}(v_1 > z_1 - \frac{\sigma^2}{\omega_1} t, \ldots, v_k > z_k - \frac{\sigma^2}{\omega_k} t) / P_1 P_2 \cdots P_k
\]

so that the derivative becomes

\[
\frac{1}{P_1 P_2 \cdots P_k} \left\{ \phi_X'(t) P_{R_{00}}(v_1 > z_1 - \frac{\sigma^2}{\omega_1} t, \ldots, v_k > z_k - \frac{\sigma^2}{\omega_k} t) \right\}
\]

\[
+ \frac{\sigma^2}{\sqrt{2\pi}} \sum_{j=1}^{k} \frac{1}{\omega_j} e^{-\frac{1}{2}(\frac{z_j - \frac{\sigma^2}{\omega_j} t)^2}{\omega_j}}
\]

\[
P_{R_{00}}(v_1 > z_1 - \frac{\sigma^2}{\omega_1} t, \ldots, v_{j-1} > z_{j-1} - \frac{\sigma^2}{\omega_{j-1}} t, v_{j+1} > z_{j+1}
\]
Setting $t=0$, we obtain the mean value

$$\xi + \frac{\sigma^2}{\prod \omega_j} \sum_{j=1}^{k} \frac{1}{\omega_j \sqrt{2\pi}} e^{-\frac{z_j^2}{2}} P_{R_{\infty}}(v_1 > z_1, \ldots, v_{j-1} > z_{j-1}, v_{j+1} > z_{j+1}, \ldots, v_k > z_k | v_j = z_j)$$

or

$$\xi + \frac{\sigma^2}{\prod \omega_j} \sum_{j=1}^{k} \frac{1}{\omega_j \sqrt{2\pi}} e^{-\frac{z_j^2}{2}} P_{R_{\infty}}(u_1 > z_1 - \frac{\sigma^2}{\omega_j \omega_j} z_j, \ldots, u_{j-1} > z_{j-1}, u_j > z_j, u_{j+1} > z_{j+1}, \ldots, u_k > z_k - \frac{\sigma^2}{\omega_j \omega_k} z_j)$$

where
\[ R(\omega) = \begin{pmatrix}
1 - \frac{\sigma^2}{\omega_1^2} & \cdots & \frac{\sigma^2}{\omega_1 \omega_j} & \left(1 - \frac{\sigma^2}{\omega_j^2}\right) & \frac{\sigma^2}{\omega_1 \omega_{j+1}} & \left(1 - \frac{\sigma^2}{\omega_{j+1}^2}\right) & \cdots & \frac{\sigma^2}{\omega_1 \omega_k} & \left(1 - \frac{\sigma^2}{\omega_k^2}\right) \\
\vdots & & \vdots & \ddots & \vdots & \vdots & & \vdots & \vdots \\
\frac{\sigma^2}{\omega_1 \omega_{j-1}} & \left(1 - \frac{\sigma^2}{\omega_j^2}\right) & \cdots & 1 - \frac{\sigma^2}{\omega_j \omega_{j+1}} & \left(1 - \frac{\sigma^2}{\omega_{j+1}^2}\right) & \frac{\sigma^2}{\omega_j \omega_{j+2}} & \left(1 - \frac{\sigma^2}{\omega_{j+2}^2}\right) & \cdots & \frac{\sigma^2}{\omega_j \omega_k} & \left(1 - \frac{\sigma^2}{\omega_k^2}\right) \\
\vdots & & \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\
\frac{\sigma^2}{\omega_1 \omega_k} & \left(1 - \frac{\sigma^2}{\omega_k^2}\right) & \cdots & \frac{\sigma^2}{\omega_{j-1} \omega_k} & \left(1 - \frac{\sigma^2}{\omega_{j+1}^2}\right) & \frac{\sigma^2}{\omega_{j+1} \omega_{j+2}} & \left(1 - \frac{\sigma^2}{\omega_{j+2}^2}\right) & \cdots & 1 - \frac{\sigma^4}{\omega_{j+1} \omega_k} 
\end{pmatrix} \]
Head, Logistics and Mathematical Statistics Branch 3 copies
Office of Naval Research
Washington, D. C. 20360

Commanding Officer 2 copies
Office of Naval Research Branch Office
Navy 100 Fleet Post Office
New York, New York

Defense Documentation Center 20 copies
Cameron Station
Alexandria, Virginia 22314

Defense Logistics Studies 1 copy
Information Exchange
Army Logistics Management Center
Fort Lee, Virginia
Attn: William B. Whichard

Technical Information Officer 6 copies
Naval Research Laboratory
Washington, D. C. 20390

Commanding Officer 1 copy
Office of Naval Research Branch Office
207 West 24th Street
New York, New York 10011
Attn: J. Laderman

Commanding Officer 1 copy
Office of Naval Research Branch Office
1030 East Green Street
Pasadena 1, California
Attn: Dr. A. R. Laufer

Bureau of Supplies and Accounts 1 copy
Code OW
Department of the Navy
Washington 25, D. C.

Institute for Defense Analyses 1 copy
Communications Research Division
von Neumann Hall
Princeton, New Jersey