REPEATED SELECTION WITH MULTIVARIATE NORMAL DISTRIBUTION

Ying Y. Wang

BU-231-M December, 1966

ABSTRACT

This is a multivariate extension of a k-cycle selection model with normal distribution, as discussed by Robson (BU-171-M). In the selection problem, we assume that the p-dimensional trait vector $\underline{X}$ is observed with errors in the i^{th} cycle as $\underline{Y}_i = \underline{X} + \underline{E}_i$, where $\underline{X}$ and $\underline{E}_i$, $i = 1, 2, \cdots, k$, are all multivariate p-dimensional normal vectors and independent of each other. Selection at i^{th} stage retains the upper P_i fraction for the next stage. We shall consider two kinds of selections, componentwise selection and selection using a linear function on $\underline{Y}_i$.

Department of Plant Breeding and Biometry, Cornell University, Ithaca, New York.

REPEATED SELECTION WITH MULTIVARIATE NORMAL DISTRIBUTION

Ying Y. Wang

BU-231-M December, 1966

This is a multivariate extension of a k-cycle selection model with normal distribution, as discussed by Robson (BU-171-M). In the selection problem, we assume that the p-dimensional trait vector $\underline{X}$ is observed with errors in the i^{th} cycle as $\underline{Y}_i = \underline{X} + \underline{E}_i$, where $\underline{X}$ and $\underline{E}_i$, $i = 1, 2, \cdots, k$, are all multivariate p-dimensional normal vectors and independent of each other. Selection at i^{th} stage retains the upper P_i fraction for the next stage. We shall consider two kinds of selections, componentwise selection and selection using a linear function on $\underline{Y}_i$.

Let

$$\underline{X} \sim N(\boldsymbol{\xi}, \Sigma), \qquad p \times 1, \qquad \Sigma = (\sigma_{ij})$$

$$\underline{Y}_i = \underline{X} + \underline{E}_i, \quad \underline{E}_i \sim N(\underline{0}, \Omega_i), \quad \Omega_i = (\omega_{ijk})$$

$$\underline{Y}_i \sim N(\boldsymbol{\xi}, \Sigma + \Omega_i), \quad i = 1, 2, \cdots, k.$$

$$\mathrm{Cov}(\underline{X}, \underline{E}_i) = 0, \quad \mathrm{Cov}(\underline{E}_i, \underline{E}_j) = 0; \quad i \neq j.$$

The $\underline{Y}_i$'s are selected $\underline{Y}_i \geq \underline{y}_i$, componentwise, with

$$P_1 = \Pr\{\underline{Y}_1 \geq \underline{y}_1\}$$

$$P_1 P_2 = \Pr\{\underline{Y}_1 \geq \underline{y}_1, \underline{Y}_2 \geq \underline{y}_2\}$$

$$\vdots$$

$$P_1 P_2 \cdots P_k = \Pr\{\underline{Y}_1 \geq \underline{y}_1, \cdots \underline{Y}_k \geq \underline{y}_k\}.$$

Let

$$W_i = (X_i - \xi_i)/\sqrt{\sigma_{ii}} \qquad , \quad i = 1, 2, \cdots, p$$

$$Z_j = (Y_{\ell i} - \xi_i)/\sqrt{\sigma_{ii} + \omega_{\ell ii}} \quad , \quad j = \ell p + i, \quad j = 1, 2, \cdots, kp$$

so

$$\begin{array}{l} p \times 1 \\ kp \times 1 \end{array} \begin{bmatrix} \underline{W} \\ \underline{Z} \end{bmatrix} \sim N \begin{bmatrix} \underline{0} & R \end{bmatrix}$$

where

$$R = \begin{bmatrix} R_1 & R_3' \\ R_3 & R_2 \end{bmatrix} = \text{correlation matrix of } \underline{X} \text{ and } \begin{bmatrix} \underline{Y}_1 \\ \vdots \\ \underline{Y}_k \end{bmatrix}.$$

Let

$$\underline{z} = (z_j) \qquad z_j = (y_{\ell i} - \xi_i)/\sqrt{\sigma_{ii} + \omega_{\ell ii}}\,, \qquad j = \ell p + i$$

$$j = 1, 2, \cdots, kp$$

the conditional distribution of $\underline{W}$ given $\underline{Z} = \underline{z}$ is

$$\underline{W}|\underline{z} \sim N(R_3'R_2^{-1}\underline{z}\,,\ \ R_1 - R_3'R_2^{-1}R_3).$$

The moment generating function for $\underline{W}$ given $\underline{Z} \geq \underline{z}$ is

$$M_{\underline{W}|\underline{Z}\geq\underline{z}}(\underline{t}) = E(e^{\underline{t}'\underline{W}}|\underline{Z} \geq \underline{z}) = E_{\underline{U}\geq\underline{z}}\left(E(e^{\underline{t}'\underline{W}}|\underline{Z} = \underline{U})\right)$$

$$= E_{\underline{U}\geq\underline{z}}\left(e^{\underline{t}'R_3'R_2^{-1}\underline{U} + \frac{1}{2}\underline{t}'(R_1 - R_3'R_2^{-1}R_3)\underline{t}}\right)$$

$$= e^{\frac{1}{2}\underline{t}'R_1\underline{t}} \frac{1}{(|R_2|2\pi)^{\frac{kp}{2}} P_1\cdots P_k} \int_{\underline{U}\geq\underline{z}} e^{-\frac{1}{2}\underline{U}'R_2^{-1}\underline{U} - \frac{1}{2}\underline{t}'R_3'R_2^{-1}R_3\underline{t} + \underline{t}'R_3'R_2^{-1}\underline{U}}\, d\underline{U}$$

$$= M_{\underline{W}}(\underline{t}) \frac{1}{(|R_2|2\pi)^{\frac{kp}{2}} P_1\cdots P_k} \int_{\underline{U}\geq\underline{z}-R_3\underline{t}} e^{-\frac{1}{2}\underline{U}'R_2^{-1}\underline{U}}\, d\underline{U}\,.$$

So the conditional mean of $\underline{W}$ has i^{th} component

$$E(W_i|\underline{Z} \geq \underline{z}) = \frac{\partial}{\partial t_i} M_{\underline{W}|\underline{Z}\geq\underline{z}}(\underline{t})\Big|_{\underline{t}=\underline{0}}$$

$$= \sum_{j=1}^{kp} \Pr\{U_\ell \geq z_\ell,\ \forall \ell \neq j\,|\,\underline{U} \geq \underline{z},\ U_j = z_j\}\, R_{3ji}\, e^{-z_j^2/2}\,.$$

Let ${}_{j}\underline{U}$ denote the $\underline{U}$ vector with U_j deleted, then given $U_j = z_j$, ${}_{j}\underline{U}$ is distributed as

$$ {}_{j}\underline{U} \sim N(R_{2j.}z_j \ , \ R_{2..} - R_{2j.}R'_{2j.}) $$

where $R_{2..}$ is the matrix R_2 with j^{th} row and column deleted,

$R_{2j.}$ is the j^{th} column of R_2 without the j^{th} element.

The matrix $\underset{kp \times p}{R_3}$ has ji^{th} element R_{3ji} , $\quad j = np + \ell$

$$ R_{3ji} = E\left(\frac{X_\ell - \xi_\ell}{\sqrt{\sigma_{\ell\ell} + \omega_{n\ell\ell}}}\right)\left(\frac{X_i - \xi_i}{\sqrt{\sigma_{ii}}}\right) $$

$$ = \frac{\sigma_{i\ell}}{\sqrt{\sigma_{ii}}\sqrt{\sigma_{\ell\ell} + \omega_{n\ell\ell}}} $$

Let ${}_{j}\underline{V} = {}_{j}\underline{U} - R_{2j.}z_j \sim N(0, R^*_j)$, $\quad R^*_j = R_{2..} - R_{2j.}R'_{2j.}$ so finally

$$ E(W_i|\underline{Z} \geq \underline{z}) = \sum_{j=1}^{kp} \frac{e^{-z_j^2/2}}{P_1 \cdots P_k \ (2\pi|R^*_j|)^{\frac{1}{2}}} \ \Pr\{{}_{j}\underline{V} \geq {}_{j}\underline{z} - R_{2j.}z_j\} $$

where $\quad j = np + \ell$

$$ |R_2| = |R_{2jj}| \cdot |R_{2..} - R_{2j.}R'_{2j.}| = |R^*_j| \ . $$

Alternatively, if the selection uses linear functions on $\underline{Y}_i$ and chooses

$$ \underline{a}_i'\underline{Y}_i \geq y_i \ , \quad i = 1, 2, \cdots, k, $$

with

$$ P_1 = \Pr\{\underline{a}_1'\underline{Y}_1 \geq y_1\}, \quad P_1 \cdots P_k = \Pr\{\underline{a}_i'\underline{Y}_i \geq y_i, \quad i = 1, 2, \cdots, k\}, $$

then

$$ V_i = \underline{a}_i'\underline{Y}_i = \underline{a}_i'\underline{X} + \underline{a}_i'\underline{E}_i \sim N(\underline{a}_i'\underline{\xi}, \ \underline{a}_i'(\Sigma + \Omega_i)\underline{a}_i). $$

Let $W_i = (X_i - \xi_i)/\sqrt{\sigma_{ii}}$; $i = 1, 2, \cdots, p$

$Z_j = (V_j - \underline{a}_j'\underline{\xi})/\sqrt{\underline{a}_j'(\Sigma + \Omega_j)\underline{a}_j}$; $j = 1, 2, \cdots, k,$

and

$$\begin{array}{c} p \times 1 \\ k \times 1 \end{array} \quad \begin{bmatrix} \underline{W} \\ \underline{Z} \end{bmatrix} \sim N\left[\underline{0},\ R\right]$$

where R is the correlation matrix of $\underline{X}$ and $\underline{V}$.

Similar to before

$$E(e^{\underline{t}'\underline{W}}|\underline{Z} \geq \underline{z}) = M_{\underline{W}}(\underline{t}) \frac{1}{P_1 \cdots P_k\ (|R_2|2\pi)^{\frac{k}{2}}} \int_{\underline{U} \geq \underline{z} - R_3\underline{t}} e^{-\frac{1}{2}\underline{U}'R_2^{-1}\underline{U}}\ d\underline{U}$$

$$E(W_i|\underline{Z} \geq \underline{z}) = \sum_{\ell=1}^{k} e^{-\frac{1}{2}z_\ell^2} \frac{\sigma_{i\ell}}{\sqrt{\sigma_{ii}\ \omega_{\ell\ell}}} \frac{\Pr\{{}_\ell\underline{U} \geq {}_\ell\underline{z} - R_{2\cdot\ell}z_\ell\}}{P_1 \cdots P_k\ (2\pi|R_\ell^*|)^{\frac{1}{2}}}$$

where

$${}_\ell\underline{U} \sim N(\underline{0},\ R_\ell^*)$$

$$R_\ell^* = R_{2\cdot\cdot} - R_{2\cdot\ell}R_{2\ell\cdot}$$

$$\omega_{\ell\ell} = \underline{a}_\ell'\ (\Sigma + \Omega_\ell)\underline{a}_\ell$$

$$z_\ell = (y_\ell - \underline{a}_\ell'\underline{\xi})/\sqrt{\underline{a}_\ell'\ (\Sigma + \Omega_\ell)\underline{a}_\ell} \ .$$

In actual computation for the multidimensional normal integral

$$\Pr\{\underline{V} \geq \underline{c}\}$$

where $\underline{V}$ is a pk-1 or k-1 dimensional vector and

$$\underline{V} \sim N(\underline{0},\ R)$$

most of the work for dimension higher than four have been restricted to special cases of the correlation matrix R. A general discussion of this topic and bibliography given by Gupta (AMS, 34, pp. 792-838) will be very helpful.