韓國數學教育學會誌 <數學教育>
1981. 12. Vol. XX No. 1

Journal of the Korea Society of Mathematical Education
Dec. 1981. Vol. XX No. 1

A Note on Maximizing the Probability of Correctly Ordering Random Variables Using Linear Predictors

By Lee Seung-Ho

Ajou University, Suweon, Korea

1. Introduction

In many situations, an experimenter wishes to predict the value of some random variable or variables that cannot be observed on the basis of a vector of observations on some other random variables.

For example, one may wishes to predict a person's intelligence from his scores on a battery of tests.

Let $\underline{U}$ and $\underline{X}$ be two random vectors having joint distribution.

Then it is well known that the best predictor of $\underline{U}$ based on $\underline{X}$ is

$$\tilde{\underline{U}}=E(\underline{U}|\underline{X})$$

Under normality, i.e.,

$$\begin{pmatrix}\underline{U}\\ \underline{X}\end{pmatrix}\sim \mathrm{MN}\left(\begin{pmatrix}\underline{\mu}\\ \underline{\nu}\end{pmatrix}, \begin{pmatrix}V & \underline{c}'\\ \underline{c} & P\end{pmatrix}\right) \quad (1.1)$$

the best predictor $\tilde{\underline{U}}$ is identical to the best linear predictor $B'X$, i.e.,

$$\tilde{\underline{U}}=E(\underline{U}|\underline{X})=\underline{\mu}+\underline{c}V^{-1}(\underline{X}-\underline{\nu})=A+B'X, \text{ (Searle 1974)}$$

In Henderson (1963) and Searle (1974), it is shown that (under appropriate conditions) the probability of correct pairwise odering is maximized by odering a pair (U_i, U_j) according to the best linear predictors based on $\underline{X}$.

And Stephen Portney (1979) considered the problem of ordering U_i's from highest to lowest. and showed that the probability of correctly ordering $\{U_i\}$ by ordering according to the ranking of $\{\underline{b}'X_i\}$ is maximized when $\underline{b}$ gives the best linear predictor, $\underline{b}=P_i^{-1}\underline{c}_i$.

To derive the results, Portney used the probability content of convex cones.

But the structure of the Variance-Covariance matrix is assumed, it can be shown directrly from the pattern of the matrix.

In this note, it is shown that the result can be obtained directrly by the simple argument based on the structure of the variance-covarince matrix.

2. Maximizing the Probability of Correct Order.

Model I. (iid case)

Let $(U_1, \underset{\sim}{x}_1)$, $(U_2, \underset{\sim}{x}_2), \ldots, (U_n, \underset{\sim}{x}_n)$ be a random sample from a multivariate normal distribution in $(1+k)$ dimensions with

(2.1) $$E(U_i)=\mu,\ \mathrm{Var}(U_i)=\sigma_u^2$$

and

(2.2) $$E(\underset{\sim}{x}_i)=\underset{\sim}{0},\ \mathrm{Var}(\underset{\sim}{x}_i)=P_{(k\times k)},\ \mathrm{Cov}(U_i, \underset{\sim}{x}_i)=\underset{\sim}{c}_{(k\times 1)}$$

For any $\underset{\sim}{b}\in R^k$

(2.3) $$E(\underset{\sim}{b}'\underset{\sim}{x}_i\,|\,U_i)=\frac{\underset{\sim}{b}'\underset{\sim}{c}}{\sigma_u^2}(U_i-\mu)=\frac{\rho\sqrt{\underset{\sim}{b}'P\underset{\sim}{b}}}{\sigma_u}(U_i-\mu)$$

(2.4) $$\mathrm{Var}(\underset{\sim}{b}'\underset{\sim}{x}_i\,|\,U_i)=(\underset{\sim}{b}'P\underset{\sim}{b})(1-\rho^2)$$

where

(2.5) $$\rho^2=\frac{(\underset{\sim}{b}'\underset{\sim}{c})^2}{\sigma_u^2(\underset{\sim}{b}'P\underset{\sim}{b})}$$

For any $\underset{\sim}{b}\in R^k$ introduce the normalization

(2.6) $$\underset{\sim}{b}_N=\underset{\sim}{b}/\sqrt{\underset{\sim}{b}'P\underset{\sim}{b}}$$

Theorem 1. *Consider the sampling situation given above.*

Let $\underset{\sim}{U}'=(U_1, \ldots, U_n)\in R^n$ *and* $\underset{\sim}{X}'=(\underset{\sim}{x}_1, \ldots, \underset{\sim}{x}_n)$.

For $\underset{\sim}{b}\in R^k$, *let* $S(\underset{\sim}{b})$ *denote the event that the n coordinates of* $\underset{\sim}{U}$ *have the same order as the n coordinates of* $\underset{\sim}{X}\underset{\sim}{b}$.

Then

$$Pr(S(\underset{\sim}{b}^*))\geq Pr(S(\underset{\sim}{b})),$$

where

$$\underset{\sim}{b}^*=P^{-1}\underset{\sim}{c},$$

i.e., the probability of correct order is maximized by choosing $\underset{\sim}{b}=\underset{\sim}{b}^*$

Furthermore,

$$Pr(S(\underset{\sim}{b}^*)\,|\,\underset{\sim}{U})\geq Pr(S(\underset{\sim}{b})\,|\,\underset{\sim}{U}),\ \forall b\in R^k.$$

Proof. For fixed $\underset{\sim}{b}\in R^k$, define

(2.7) $$W_i=\underset{\sim}{b}_N'\underset{\sim}{x}_i,\ i=1, \ldots, n.$$

Then

(2.8) $$E(\underset{\sim}{X}\underset{\sim}{b}_N\,|\,\underset{\sim}{U})=\frac{\underset{\sim}{b}_N'\underset{\sim}{c}}{\sigma_u^2}(\underset{\sim}{U}-\mu\underset{\sim}{e})=\frac{\rho}{\sigma_u}(\underset{\sim}{U}-\underset{\sim}{e}),\ \underset{\sim}{e}'=(1, 1, \ldots, 1)$$

(2.9) $$\mathrm{Cov}(\underset{\sim}{X}\underset{\sim}{b}_N\,|\,\underset{\sim}{U})=\tau^2 I,\ \tau^2=1-(\underset{\sim}{b}_N'\underset{\sim}{c})^2/\sigma_u^2=1-\rho^2.$$

that is, $W_1, W_2, \ldots, W_n$, given $\underset{\sim}{U}=(U_1, \ldots, U_n)'$, are independently distributed normal variables with

(2.10) $$E(W_i\,|\,\underset{\sim}{U})=\frac{\rho}{\sigma_u}(U_i-\mu)$$

and

(2.11) $$\mathrm{Var}(W_i\,|\,\underset{\sim}{U})=1-\rho^2,\ i=1, \ldots, n$$

Furthermore, the distribution of W_i, given $\underset{\sim}{U}$, depends on $\underset{\sim}{U}$ only through U_i.

Let $U_{(1)} < U_{(2)} < \cdots < U_{(n)}$ denote the ordered $U_1, U_2, \ldots, U_n$, and let $W_{(i)}$ denote the W associated with $U_{(i)}$.

Then in order to prove the Theorem 1, it suffices to show that

$$Pr_{\underline{b}}(W_{(1)} < W_{(2)} < \cdots < W_{(n)} \mid \underline{U}) \text{ is maximized when } \underline{b} = \underline{b}^*.$$

This follows from the following.

For any given $\underline{b} \in R^k$, $W_{(i)}$ $(i=1, \ldots, n)$, for given $\underline{U}$, can be represented as follows;

$$W_{(i)} = \sqrt{1-\rho^2}\, Z_i + \frac{\rho}{\sigma_u}(U_{(i)} - \mu), \quad i=1, \ldots, n \tag{2.12}$$

where $Z_1, \ldots, Z_n$ are iid $N(0, 1)$ random variables.

Therefore,

$$\begin{aligned} & Pr_{\underline{b}}(W_{(1)} < W_{(2)} < \cdots < W_{(n)} \mid \underline{U}) \\ &= Pr_{\underline{b}}(W_{(i)} < W_{(i+1)},\ i=1, \ldots, n-1 \mid \underline{U}) \\ &= Pr_{\underline{b}}(\sqrt{1-\rho^2}(Z_i - Z_{i+1}) < \frac{\rho}{\sigma_u}(U_{(i+1)} - U_{(i)}),\ i=1, \ldots, n-1 \mid \underline{U}) \\ &= Pr_{\underline{b}}\left(\frac{\sigma_u(Z_i - Z_{i+1})}{U_{(i+1)} - U_{(i)}} < \frac{\rho}{\sqrt{1-\rho^2}},\ i=1, \ldots, n-1 \mid \underline{U} \right) \end{aligned} \tag{2.13}$$

which is clearly non-decreasing in $\frac{\rho}{\sqrt{1-\rho^2}}$, hence in ρ.

Hence

$$Pr_{\underline{b}}(W_{(1)} < W_{(2)} < \cdots < W_{(n)} \mid \underline{U}) \leq Pr_{\underline{b}^*}(W_{(1)} < W_{(2)} < \cdots < W_{(n)} \mid \underline{U}) \tag{2.14}$$

$(\rho \leq \rho^*,\ \rho^*$; associated with $b^*)$

Model Ⅱ. (intraclass correlation case)

Consider the case where the $(U_i, \underline{x}_i)$'s have a components of variance-covariance structure. To specify the model, let $\underline{U}$ be a vector in R^n and $\underline{X}$ a vector in $R^{n'}$ consisting of k block seach of size n. Then (using Kronecker Product notation) the model is specified by

$$\begin{pmatrix} \underline{U} \\ \underline{X} \end{pmatrix} \sim \mathrm{MN}\left(\begin{pmatrix} \underline{e} \\ \underline{0} \end{pmatrix},\ \Sigma_1 \otimes I_n + \Sigma_2 \otimes \underline{e}\underline{e}' \right),\ \underline{e} : (n \times 1) \tag{2.15}$$

where

$$\Sigma_j = \begin{pmatrix} v_j, & \underline{c}_j' \\ \underline{c}_j, & P_j \end{pmatrix} : (k+1) \times (k+1),\ j=1, 2 \tag{2.16}$$

$$\underline{c}_j \in R^k,\ P_j : k \times k,\ v_j > 0$$

To predict the coordinates of $\underline{U}$, we restrict consideration to "stationary" linear predictors $\underline{B}\underline{X}$ with $\underline{B} : n \times (nk)$, of the following form;

$$\underline{B} = (\underline{b}_1' \otimes I_n) + (\underline{b}_2' \otimes \underline{e}\underline{e}'),\ \text{where } \underline{b}_1, \underline{b}_2 \in R^k \tag{2.17}$$

Theorem 2. *Under model (2.15), the probability that the order of the coordinates of $\underline{U}$ is the same as the order of the coordinates of $\underline{B}\underline{X}$ is maximized over $\underline{B}$ satisfying (2.17) by $\underline{B} = \underline{B}^*$, the coefficients of the best linear predictors.*

In this case, $\underline{B}^*$ may be given by

$$\underline{b}_1^* = P_1^{-1}\underline{c}_1 \text{ and } \underline{b}_2^* = \underline{0} \tag{2.18}$$

Proof. Let $\underline{X}' = (\underline{x}^1, \underline{x}^2, \ldots, \underline{x}^k)$

where

$$\underset{\sim}{x}^j=(x_1^j, \ldots, x_n^j)' : (n\times 1) \text{ for } j=1, 2, \ldots, k.$$

Then, since

$$\begin{pmatrix} \underset{\sim}{U} \\ \underset{\sim}{X} \end{pmatrix} \sim \mathrm{MN}\left(\begin{pmatrix} \underset{\sim}{e} \\ \underset{\sim}{0} \end{pmatrix}, \begin{pmatrix} v_1 I_n, & \underset{\sim}{c}_1'\otimes I_n \\ \underset{\sim}{c}_1\otimes I_n. & P_1\otimes I_n \end{pmatrix} + \begin{pmatrix} v_2 J_n, & \underset{\sim}{c}_2'\otimes J_n \\ \underset{\sim}{c}_2\otimes J_n, & P_2\otimes J_n \end{pmatrix}\right)$$

where

$$J_n=ee' : (n\times n)$$

$\underset{\sim}{X}$, given $\underset{\sim}{U}$, has normal distribution with

$$E(\underset{\sim}{X}\,|\,\underset{\sim}{U})=(\underset{\sim}{c}_1\otimes I_n+\underset{\sim}{c}_2\otimes J_n)\frac{1}{v_1}\left(I_n-\frac{\nu_2}{v_1+nv_2}J_n\right)(\underset{\sim}{U}-\mu\underset{\sim}{e})$$

and

$$\begin{aligned}\mathrm{Var}(\underset{\sim}{X}\,|\,\underset{\sim}{U})&=(P_1\otimes I_n+P_2\otimes J_n)\\ &\quad-(\underset{\sim}{c}_1\otimes I_n+\underset{\sim}{c}_2\otimes J_n)\frac{1}{v_1}\left(I_n-\frac{v_2}{v_1+nv_2}J_n\right)\\ &\quad\cdot(\underset{\sim}{c}_1'\times I_n+\underset{\sim}{c}_2'\otimes J_n)\end{aligned}$$

For any fixed $\underset{\sim}{B}$ of the form (2.17), it follows from the above that $\underset{\sim}{B}\underset{\sim}{X}$, given $\underset{\sim}{U}$, has normal distribution with

$$\begin{aligned}E(\underset{\sim}{B}\underset{\sim}{X}\,|\,\underset{\sim}{U})&=\underset{\sim}{B}E(\underset{\sim}{X}\,|\,\underset{\sim}{U})\\ &=(\underset{\sim}{b}_1'\otimes I_n+\underset{\sim}{b}_2'\otimes J_n)(\underset{\sim}{c}_1\otimes I_n+\underset{\sim}{c}_2\otimes J_n)\frac{1}{v_1}\left(I_n-\frac{v_2}{v_1+nv_2}J_n\right)(\underset{\sim}{U}-\mu\underset{\sim}{e})\\ &=\frac{1}{v_1}\left\{\underset{\sim}{b}_1\underset{\sim}{c}_1 I_n+\left[-\underset{\sim}{b}_1'\underset{\sim}{c}_1\frac{v_2}{v_1+nv_2}+\frac{v_1}{v_1+nv_2}(\underset{\sim}{b}_1'\underset{\sim}{c}_2+\underset{\sim}{b}_2'\underset{\sim}{c}_1+n\underset{\sim}{b}_2'\underset{\sim}{c}_2)J_n\right]\right\}\cdot(\underset{\sim}{U}-\mu\underset{\sim}{e})\\ &=\frac{1}{v_1}(\alpha I_n+\beta J_n)(\underset{\sim}{U}-\mu\underset{\sim}{e})\\ &=\frac{1}{v_1}\left\{\alpha\underset{\sim}{U}-\alpha\mu\underset{\sim}{e}+\beta\underset{\sim}{e}\left(\sum_{i=1}^n U_i\right)-n\beta\mu\underset{\sim}{e}\right\}\\ &=\frac{1}{v_1}\{\alpha\underset{\sim}{U}-(\alpha\mu-\beta\sum_{i=1}^n U_i+\beta n\mu)\underset{\sim}{e}\}\\ &=\frac{\underset{\sim}{b}_1'\underset{\sim}{c}_1}{v_1}\underset{\sim}{U}-\eta\underset{\sim}{e}\end{aligned}$$

where η is some scalar depending on $\underset{\sim}{U}$, $\underset{\sim}{b}_1$, and $\underset{\sim}{b}_2$.
and

$$\begin{aligned}\mathrm{Var}(\underset{\sim}{B}\underset{\sim}{X}\,|\,\underset{\sim}{U})&=\underset{\sim}{B}\mathrm{Var}(\underset{\sim}{X}\,|\,\underset{\sim}{U})\underset{\sim}{B}'\\ &=\left(\underset{\sim}{b}_1'P_1\underset{\sim}{b}_1-\frac{1}{v_1}(\underset{\sim}{b}_1'\underset{\sim}{c}_1)^2\right)I_n+\xi J_n\end{aligned}$$

where ξ is some scalar depending on $\underset{\sim}{b}_1, \underset{\sim}{b}_2$.

Note that η depends on $\underset{\sim}{U}$ only through $\sum_{i=1}^n U_i$

Let $U_{(1)}<U_{(2)}<\cdots<U_{(n)}$ denote the ordered $U_1, U_2, \ldots, U_n$ and let $W_{(i)}$ denote the coordinate of $\underset{\sim}{B}\underset{\sim}{X}$ associated with $U_{(i)}$ for $i=1, \ldots, n$.
Then, for given $\underset{\sim}{U}$, $W_{(i)}$ can be represented as

$$W_{(i)}=\frac{\underset{\sim}{b}_1'\underset{\sim}{c}_1}{v_1}U_{(i)}+\eta+\sqrt{\nu}Z_i-(\sqrt{\xi^-}+\sqrt{\xi^+})Z_0;\ i=1, \ldots, n$$

where

$$\nu=\left(\underline{b}_1'P_1\underline{b}_1-\frac{1}{v_1}(\underline{b}_1'\underline{c}_1)^2\right)$$

ξ^-, ξ^+; negative and positive part of ξ

and $Z_0, Z_1, \ldots, Z_n$; $N(0, 1)$ random variables with

$$\mathrm{Cov}(Z_0, Z_i)=\sqrt{\xi^-}/\sqrt{\nu},\quad i=1,\ldots,n$$

$$\mathrm{Cov}(Z_i, Z_j)=0 \text{ for } 1\leq i<j\leq n$$

Therefore,

$$\begin{aligned}
&Pr_{\underline{B}}(W_{(1)}<W_{(2)}<\cdots<W_{(n)}\mid U)\\
&=Pr_{\underline{B}}\left(\sqrt{\nu}(Z_i-Z_{i+1})<\frac{\underline{b}_1'\underline{c}_1}{v_1}(U_{(i+1)}-U_{(i)}),\ i=1,\ldots,n-1\mid U\right)\\
&=Pr_{\underline{B}}\left(\frac{v_1(Z_i-Z_{i+1})}{U_{(i+1)}-U_{(i)}}<\frac{\underline{b}_1'\underline{c}_1}{\sqrt{\nu}},\ i=1,\ldots,n-1\mid U\right)\\
&=Pr_{\underline{B}}\left(\frac{\sqrt{v_1}(Z_i-Z_{i+1})}{U_{(i+1)}-U_{(i)}}<\frac{\rho}{\sqrt{1-\rho^2}},\ i=1,\ldots,n-1\mid U\right)
\end{aligned}$$

where

$$\rho=\frac{\underline{b}_1'\underline{c}_1}{\sqrt{v_1}\sqrt{\underline{b}_1'P_1\underline{b}_1}}$$

Note that the above expression depends on $\underline{B}$(*i.e.*, $\underline{b}_1$ and $\underline{b}_2$), only through $\frac{\rho}{\sqrt{1-\rho^2}}$, since $(Z_1, \ldots, Z_n)\sim \mathrm{MN}(\underline{0}, I_n)$

Hence as in Model I, the probability is maximized when $\rho=\rho^*$,

$$i.e.,\quad \underline{b}_1=\underline{b}_1^*$$

Since $\underline{b}_2$ is arbitrary, we may choose $\underline{b}_2^*=\underline{0}$.

References

(1) Rao, C.R. (1973). *Linear Statistical Inference and Its Applications*, 2nd ed., John Wiley & Sons, New York.

(2) Searle, S.R. (1974). *Prediction, Mixed Models, and Variance Components. Proc, Conf, Reliability and Biometry*. Ed. Proschen and Serfling, S.I.A.M., Philadelphia.

(3) Stephen Portnoy (1979). Maximizing the Probability of Correctly Ordering Random Variables Using Linear Predictors. *Technical Report, Dept. of Mathematics, Univ. of Illinois.*