ANIMAL BREEDING NOTES

CHAPTER 9

BEST LINEAR PREDICTION

Derivation of the Best Linear Predictor (BLP)

Consider:

$y = [y_1\ y_2\ ...\ y_n]$, an observable random vector, and

$g = [g_1\ g_2\ ...\ g_p]$, an unobservable random vector.

The vectors y and g are **jointly** distributed. If the joint distribution of y and g is **unknown** or mathematically intractable, but the means and variances of y and g and the covariance between y and g are **known**, g can be predicted using the best linear predictor (BLP) of g with respect to y. **Thus, g is predicted using**

$\hat{g}$ = **a + By,**

where the vector a and the matrix B are chosen such that they minimize the mean square error of prediction (MSEP), i.e., they minimize

E[(a + By − g)′ A (a + By − g)],

where A is any s.p.d. matrix.

Let

$$\begin{bmatrix} y \\ g \end{bmatrix} \sim \left\{ \begin{bmatrix} \mu_y \\ \mu_g \end{bmatrix} \begin{bmatrix} V & C \\ C' & G \end{bmatrix} \right\}.$$

Then, minimizing the MSEP with respect to the vector a and the matrix B yields $\hat{g}$**, the BLP of g, where**

$$\hat{g} = \boldsymbol{\mu}_g + \mathbf{C}'\mathbf{V}^{-1}(\mathbf{y} - \boldsymbol{\mu}_y).$$

Proof: First, the following theorem (based on theorem 1, Searle, 1971, pg. 55) is needed.

Theorem:

Let y and g be two random vectors, where

$$\begin{bmatrix} y \\ g \end{bmatrix} \sim \left\{ \begin{bmatrix} \mu_y \\ \mu_g \end{bmatrix}, \begin{bmatrix} V & C \\ C' & G \end{bmatrix} \right\}.$$

Then,

(a) $E[y'Ag] = tr(AC') + \mu_y'A\mu_g$,

(b) $E[y'Ay] = tr(AV) + \mu_y'A\mu_y$, and

(c) $E[g'Ag] = tr(AG) + \mu_g'A\mu_g$.

Proof of Theorem:

(a) $cov(g,y') = C'$

$= E[gy'] - \mu_g\mu_y'$

$\Rightarrow E[gy'] = C' + \mu_g\mu_y$

Because y′Ag is a scalar, it equals its own trace, thus

$$\begin{aligned} E[y'Ag] &= E[tr(y'Ag)] \\ &= E[tr(Agy')] \\ &= tr(E[Agy']) \\ &= tr(AE[gy']) \\ &= tr(A[C' + \mu_g\mu_y]) \\ &= tr(AC') + tr(A\mu_g\mu_y') \\ &= tr(AC') + \mu_y'A\mu_g \end{aligned}$$

(b) $V = E[yy'] - \mu_y\mu_y'$

$\Rightarrow \quad E[yy'] = V + \mu_y\mu_y'$

$\Rightarrow \quad E[y'Ay] = tr(AE[yy']]$

$= tr(AV) + \mu_y'A\mu_y$

(c) $G = E[gg'] - \mu_g\mu_g'$

$\Rightarrow \quad E[gg'] = G + \mu_g\mu_g'$

$\Rightarrow \quad E[g'Ag] = tr(AE[gg'])$

$= tr(AG) + \mu_gA\mu_g'$

Proof of BLP of g:

$E[(a + By - g)'A(a + By - g)]$

$= E[a'Aa + a'ABy - a'Ag + y'B'Aa + y'B'ABy - y'B'Ag - g'Aa - g'ABy + g'Ag]$

$= E[a'Aa + 2a'ABy - 2a'Ag + y'B'ABy - 2y'B'Ag + g'Ag]$, because quadratic forms are scalars,

$= a'Aa + 2a'AB\mu_y - 2a'A\mu_g + tr(B'ABV) + \mu_y'B'AB\mu_y - 2tr(B'AC') - 2\mu_y'B'A\mu_g + tr(AG) + \mu_g'A\mu_g$

$\equiv L$

$$\frac{\partial L}{\partial a} = 2Aa + 2AB\mu_y - 2A\mu_g = 0$$

$$a + B\mu_y = \mu_g$$

$$\Rightarrow \quad a = \mu_g - B\mu_y$$

$$\frac{\partial L}{\partial B} = 2A'a\mu_y' + ABV + A'BV' + AB\mu_y\mu_y' + A'B\mu_y\mu_y' - 2AC - A\mu_y\mu_y' = 0$$

because:

$$\frac{\partial}{\partial B}(2a'AB\mu_y) = \frac{\partial}{\partial B}\mathrm{tr}(2a'AB\mu_y)$$

$$= \frac{\partial}{\partial B}\mathrm{tr}(2B\mu_y a'A)$$

$$= 2A'a\mu_y'$$

$$\frac{\partial}{\partial B}\mathrm{tr}(B'ABV) = ABV + A'BV'$$

$$\frac{\partial}{\partial B}(\mu_y'B'AB\mu_y) = \frac{\partial}{\partial B}\mathrm{tr}(\mu_y'B'AB\mu_y)$$

$$= \frac{\partial}{\partial B}\mathrm{tr}(B'AB\mu_y\mu_y')$$

$$= AB\mu_y\mu_y' + A'B\mu_y\mu_y'$$

$$\frac{\partial}{\partial B}\mathrm{tr}(B'AC') = AC'$$

$$\frac{\partial}{\partial B}(2\mu_y'B'A\mu_g) = \frac{\partial}{\partial B}\mathrm{tr}(2B'A\mu_g\mu_y')$$

$$= 2A\mu_g\mu_y'$$

But $A' = A$ and $V' = V$. Thus,

$$\frac{\partial L}{\partial B} = 2Aa\mu_y' + 2ABV + 2AB\mu_y\mu_y' - 2AC' - 2A\mu_g\mu_y' = 0$$

$$a\mu_y' + BV + B\mu_y\mu_y' - C' - \mu_g\mu_y' = 0$$

$$(a + B\mu_y - \mu_g)\mu_y' + BV = C'$$

Also, because $a = \mu_g - B\mu_y$,

$$(\mu_g - B\mu_y + B\mu_y - \mu_g)\mu_y' + BV = C'$$

$$BV = C'$$

$\Rightarrow$ $B = C'V^{-1}$

$\Rightarrow$ $a = \mu_g - C'V^{-1}\mu_y$

Substituting these expressions for a and B in $\hat{g} = a + By$ yields

$$\hat{g} = \mu_g - C'V^{-1}\mu_y + C'V^{-1}y$$

$\hat{g} = \mu_g + C'V^{-1}(y - \mu_y)$, the **BLP of g.**

Properties of the Best Linear Predictor

[1] $E[\hat{g}] = E[\mu_g + C'V^{-1}(y - \mu_y)]$

$= \mu_g + C'V^{-1}(\mu_y - \mu_y)$

$= \mu_g$

$= E[g]$

$\Rightarrow$ the BLP is unbiased even though unbiasedness was **not** required in its derivation, and

$\Rightarrow$ the BLP minimizes the error variance of prediction (EVP) of $\hat{g}$, because $E[\hat{g} - g] = 0$.

[2] $\text{var}(\hat{g}) = \text{cov}(\hat{g}, \hat{g}')$

$= \text{cov}(C'V^{-1}y, y'V^{-1}C)$

$= C'V^{-1}VV^{-1}C$

$= C'V^{-1}C$

[3] $\text{cov}(\hat{g}, g') = \text{cov}(C'V^{-1}y, g')$

$= C'V^{-1}C$

$= \text{var}(\hat{g})$

[4] $\text{var}(\hat{g} - g) = \text{var}(\hat{g}) - 2\text{cov}(\hat{g}, g') + \text{var}(g)$

$$= \text{var}(\hat{g}) - 2\text{var}(\hat{g}) + \text{var}(g)$$

$$= \text{var}(g) - \text{var}(\hat{g})$$

$$= G - C'V^{-1}C$$

[5] Let $\tilde{g}$ = **BP of g** and $\hat{g}$ = **BLP of g**.

Then,

$$\text{var}(\hat{g} - g) = \text{var}(\tilde{g} - g) + \text{var}(\hat{g} - \tilde{g})$$

Proof:

$$\text{var}(\hat{g} - g) = E_y[\text{var}((\hat{g} - g) \mid y)] + \text{var}(E[\hat{g} - g) \mid y])$$

$$= E_y[\text{var}(\hat{g} \mid y) + \text{var}(g \mid y) - 2\text{cov}((\hat{g} \mid y),(g \mid y))]$$

$$+ \text{var}(\hat{g} - E[g \mid y])$$

But $(\hat{g} \mid y) = (\mu_g + C'V^{-1}(y - \mu_y) \mid y)$ is a constant, thus

$$\text{var}(\hat{g} - g) = E_y[\text{var}(g \mid y)] + \text{var}(\hat{g} - \tilde{g})$$

$$= \text{var}(\tilde{g} - g) \quad \} \text{ EVP of BP}$$

$$+ \text{var}(\hat{g} - \tilde{g}) \quad \} \text{ variance due to the nonlinearity of the BP of } \tilde{g}$$

The $\text{var}(\hat{g} - \tilde{g})$ is the price paid for limiting the BP to linear functions only.

[6] **The BLP maximizes $r(\hat{g}, g)$ in the class of linear predictors (a + By)**.

Proof:

$$r(\hat{g}, g') = [\text{var}(\hat{g})]^{1/2} [\text{var}(g)]^{-1/2}$$

$$= [\text{var}(g) - \text{var}(\hat{g} - g)]^{1/2} [\text{var}(g)]^{-1/2}$$

But $\hat{g}$, the BLP of g, minimizes the EVP, i.e., it minimizes $\text{var}(\hat{g} - g)$, thus, as $\text{var}(\hat{g} - g) \rightarrow 0$

$$r(\hat{g}, g') \rightarrow [\text{var}(g)]^{½}[\text{var}(g)]^{-½}$$

$$\rightarrow I$$

⇒ BLP maximizes $r(\hat{g}, g')$ in the class of $(a + By)$ predictors.

[7] Under **multivariate normality**,

$$E_y[g|\hat{g}] = E_g[g|y] = \hat{g}$$

Proof:

$$\begin{bmatrix} \hat{g} \\ g \end{bmatrix} \sim MVN\left\{\begin{bmatrix} \mu_g \\ \mu_g \end{bmatrix}, \begin{bmatrix} C'V^{-1}C & C'V^{-1}C \\ C'V^{-1}C & G \end{bmatrix}\right\}$$

$$\begin{aligned} E_y[g|\hat{g}] &= \mu_g + C'V^{-1}C(C'V^{-1}C)^{-1}(\hat{g} - E_y[\hat{g}]) \\ &= \mu_g + (\mu_g + C'V^{-1}(y - \mu_y) - \mu_g) \\ &= \mu_g + C'V^{-1}C(y - \mu_y) \\ &= \hat{g} \quad \text{the BLP of } g \\ &= E_g[g|y] \quad \text{the BP of g under normality.} \end{aligned}$$

This is called the "strong property of the BP and the BLP of g under normality", because it has direct bearing upon their property of correct pairwise ranking.

[8] **Under normality, the ranking on $\hat{g}$, the BLP (and the BP) of g, maximizes the probability of correct pairwise ranking**.

Proof:

Let $m'g$ be a contrast of two g_i's, i.e., $g_i - g_{i'}$.

Then,

Probability {correct ranking} $= P\{m'g > 0 | m'\hat{g} > 0\} + P\{m'g < 0 | m'\hat{g} < 0\}$

But

$$\begin{bmatrix} y \\ g \end{bmatrix} \sim MVN \left\{ \begin{bmatrix} \mu_y \\ \mu_g \end{bmatrix}, \begin{bmatrix} V & C \\ C' & G \end{bmatrix} \right\}$$

and

$$\begin{bmatrix} \hat{g} \\ g \end{bmatrix} \sim MVN \left\{ \begin{bmatrix} \mu_g \\ \mu_g \end{bmatrix}, \begin{bmatrix} C'V^{-1}C & C'V^{-1}C \\ C'V^{-1}C & G \end{bmatrix} \right\}$$

Thus,

$$\begin{bmatrix} m'\hat{g} \\ m'g \end{bmatrix} \sim BVN \left\{ \begin{bmatrix} m'\mu_g \\ m'\mu_g \end{bmatrix}, \begin{bmatrix} m'C'V^{-1}Cm & m'C'V^{-1}Cm \\ m'C'V^{-1}Cm & m'Gm \end{bmatrix} \right\}$$

and

$$\begin{aligned} E_y[m'g | m'\hat{g}] &= m'\mu_g + m'C'V^{-1}Cm\,(m'C'V^{-1}Cm)^{-1}\,(m'\hat{g} - E_y[m'\hat{g}]) \\ &= m'\mu_g + (m'\mu_g + m'C'V^{-1}(y - \mu_y) - m'\mu_g) \\ &= m'\hat{g} \quad \text{the BLP of } m'g \text{ under normality}, \\ &= m'E_g[g|y] \quad \text{the BP of } m'g \text{ under normality}. \end{aligned}$$

Thus, the probability of correct pairwise ranking depends on two factors:

(a) $m'\mu_g$ and

(b) $m'C'V^{-1}(y - \mu_y)$.

To maximize the probability of correct pairwise ranking is equivalent to maximizing the correlation between $m'g$ and $m'\hat{g}$ **with the condition** that $m'\mu_g = 0$. This implies that all μ_g are equal, e.g., all animals come from the same population, hence μ_g is the same for all of them.

But the BLP (and the BP) of g maximizes r(g, $\hat{g}$), and r(m′g, $\hat{g}'$ m) = m′ r(g, $\hat{g}'$) m. Consequently, the BLP (and the BP) of g also maximizes r(m′g, $\hat{g}'$ m).

Thus, the BLP (and the BP) of g **under normality maximizes the probability of correct pairwise ranking for all pairs** (g_i,$g_{i'}$) when the **means** of the {g_i} are the **same**.

Disadvantages of the Best Linear Predictor

(1) It requires E[g] and E[y]. In Animal Breeding it has been assumed that all animals to be evaluated belong to the same population. Consequently, E[g] has been **assumed** to be equal to a **zero** (any constant would be appropriate because its value does not affect the ranking of the BLP), and E[y] = Xβ, where X is a known incidence matrix and β is a fixed known vector. However, β is usually **unknown**, thus, the **usual** procedure has been to **estimate** Xβ **from the data** and then **compute the BLP** of g **as if** $X\hat{\beta} = X\beta$.

(2) It requires the variances and covariances, which in many instances are unknown. The usual strategy has been to compute these covariances from the data or to take them from the literature and then compute the BLP as if $\hat{V} = V$ and $\hat{C} = C$.

(3) Computational problems arise in cases of multiple cross-classified data with large number of unbalanced and(or) missing subclasses. The BLP cannot be used with large unbalanced data sets. Examples of problems are the computation of a large nondiagonal V^{-1} and of vector β.

References

Henderson, C. R. 1963. Selection index and expected genetic advance. In: Statistical Genetics and Plant Breeding, NAS-NRC 982, pp 141-163.

Henderson, C. R. 1973. Sire evaluation and genetic trends. In: Proc. Anim. Breed. and Genet. Symp. in Honor of J.L. Lush, ASAS-ADSA, pp 10-41.

Quaas, R. L. 1986. Personal Communication. Animal Science 720, Cornell University, Ithaca, NY.

Searle, S. R. 1971. Linear Models. John Wiley and Sons, Inc., New York.

Searle, S. R. 1973. Derivation of prediction formulae. Mimeo BU-482-M, Biometrics Unit, Cornell University, Ithaca, NY.

Searle, S. R. 1974. Prediction, mixed models and variance components. In: Proc. Conf. Reliability and Biometry, Tallahassee, Florida. F. Proschan and R.J. Serfling (Ed.), SIAM, Philadelphia, Pennsylvania. (Mimeo BU-468-M, Biometrics Unit, Cornell Univ, Ithaca, NY., 1973).