

ANIMAL BREEDING NOTES

CHAPTER 8

BEST PREDICTION

Derivation of the Best Predictor (BP)

Let $g = [g_1 \ g_2 \ \dots \ g_p]'$ be a vector of unobservable random variables **jointly distributed** with an observable random vector $y = [y_1 \ y_2 \ \dots \ y_n]'$.

We want to predict g using y . Let $h(y)$ denote the predictor, i.e., if y is observed to be equal to \dot{y} , then $h(\dot{y})$ will be the prediction for g . Also, we want to **choose the function h** such that **$h(y)$ tends to be close to g** . **One possible criterion for closeness is to choose $h(y)$ such that it minimizes the mean square error of prediction (MSEP)**, i.e., we want

$$E[(h(y) - g)' G (h(y) - g)] \rightarrow 0$$

where G is **any** symmetric positive definite matrix, e.g., $G = \text{cov}(g, g')$. Under this criterion the **best predictor (BP) of g is:**

$$h(y) = E[g | y] \equiv \hat{g}$$

the conditional mean of g given y .

Proof:

$$\begin{aligned} E[(h(y) - g)' G (h(y) - g)] &= \int_y \int_g (h(y) - g)' G (h(y) - g) f(y, g) \, dg \, dy \\ &= \int_y \left[\int_g (h(y) - g)' G (h(y) - g) f(g | y) \, dg \right] f(y) \, dy \end{aligned}$$

\Rightarrow to minimize the $E[\]$ with respect to $h(y)$ requires to minimize only the integral over g (in brackets), because minimizing for each y implies minimizing over all y 's, so:

$$\begin{aligned}
& \frac{\partial}{\partial \mathbf{h}(\mathbf{y})} \left[\int_{\mathbf{g}} (\mathbf{h}(\mathbf{y}) - \mathbf{g})' \mathbf{G} (\mathbf{h}(\mathbf{y}) - \mathbf{g}) f(\mathbf{g} | \mathbf{y}) d\mathbf{g} \right] \\
&= \int_{\mathbf{g}} \frac{\partial}{\partial \mathbf{h}(\mathbf{y})} \left[(\mathbf{h}(\mathbf{y}) - \mathbf{g})' \mathbf{G} (\mathbf{h}(\mathbf{y}) - \mathbf{g}) f(\mathbf{g} | \mathbf{y}) \right] d\mathbf{g} \\
&= \int_{\mathbf{g}} \frac{\partial}{\partial \mathbf{h}(\mathbf{y})} \left[\mathbf{h}(\mathbf{y})' \mathbf{G} \mathbf{h}(\mathbf{y}) - 2\mathbf{h}(\mathbf{y})' \mathbf{G} \mathbf{g} + \mathbf{g}' \mathbf{G} \mathbf{g} \right] f(\mathbf{g} | \mathbf{y}) d\mathbf{g} \\
&= \int_{\mathbf{g}} (2\mathbf{G} \mathbf{h}(\mathbf{y}) - 2\mathbf{G} \mathbf{g}) f(\mathbf{g} | \mathbf{y}) d\mathbf{g} = 0
\end{aligned}$$

$$\Rightarrow 2\mathbf{G} \mathbf{h}(\mathbf{y}) \int_{\mathbf{g}} f(\mathbf{g} | \mathbf{y}) d\mathbf{g} = 2\mathbf{G} \int_{\mathbf{g}} \mathbf{g} f(\mathbf{g} | \mathbf{y}) d\mathbf{g}$$

But $\int_{\mathbf{g}} f(\mathbf{g} | \mathbf{y}) d\mathbf{g} = 1$,

$$\Rightarrow \mathbf{h}(\mathbf{y}) = \int_{\mathbf{g}} \mathbf{g} f(\mathbf{g} | \mathbf{y}) d\mathbf{g} = E[\mathbf{g} | \mathbf{y}] = \hat{\mathbf{g}}$$

\Rightarrow the BP of \mathbf{g} , $\hat{\mathbf{g}} = E[\mathbf{g} | \mathbf{y}]$, the conditional mean of \mathbf{g} given \mathbf{y} .

Remarks:

- (1) $\hat{\mathbf{g}} = E[\mathbf{g} | \mathbf{y}]$ hold for **all** density functions, and
- (2) $\hat{\mathbf{g}} = E[\mathbf{g} | \mathbf{y}]$ does **not** depend on \mathbf{G} .

Two useful results and an alternative proof for BP of $\mathbf{g} = E[\mathbf{g} | \mathbf{y}]$.

(a) Expectation of a quadratic form

Let $\mathbf{x}_{p \times 1}$ be a random vector and $\mathbf{b}_{p \times 1}$ be one of its realized vectors, where \mathbf{b} is a vector other than the mean vector $\boldsymbol{\mu}_{p \times 1}$. Then, the expected value of the ellipsoid centered at \mathbf{b} can be written as:

$$E_{\mathbf{x}}[(\mathbf{x} - \mathbf{b})' \mathbf{A} (\mathbf{x} - \mathbf{b})] = \text{tr}(\mathbf{A} \text{var}(\mathbf{x})) + (\boldsymbol{\mu} - \mathbf{b})' \mathbf{A} (\boldsymbol{\mu} - \mathbf{b})$$

where \mathbf{A} is any s.p.d. matrix.

Proof:

$$\begin{aligned}
E_x[(x - b)'A(x - b)] &= E_x[(x - \mu + \mu - b)'A(x - \mu + \mu - b)] \\
&= E_x[(x - \mu)'A(x - \mu) + (\mu - b)'A(\mu - b)]
\end{aligned}$$

Since a quadratic form is a scalar, it equals its own trace, thus,

$$\begin{aligned}
E_x[(x - b)'A(x - b)] &= E_x \operatorname{tr}[A(x - \mu)(x - \mu)'] + E_x[(\mu - b)'A(\mu - b)] \\
&= \operatorname{tr}[A E_x[(x - \mu)(x - \mu)']] + (\mu - b)'A(\mu - b) \\
&= \operatorname{tr}(A \operatorname{var}(x)) + (\mu - b)'A(\mu - b)
\end{aligned}$$

(b) Computation of variances by conditioning

Let $x_{p \times 1}$ be a random vector and $\mu_{p \times 1}$ its mean vector. If the vector x is conditioned on another random vector y , then, the variance of x can be written as the sum of the expected variance of x given y plus the variance of the expected value of x given y , i.e.,

$$\mathbf{var}(x) = E_y[\mathbf{var}_x(x | y)] + \mathbf{var}_y(E_x[x | y])$$

Proof:

$$\begin{aligned}
\operatorname{var}(x) &= E[(x - \mu)(x - \mu)'] \\
&= E_y[E[(x - \mu)(x - \mu)' | y]] \\
&= E_y[E[(x - E[x | y] + E[x | y] - \mu)(x - E[x | y] + E[x | y] - \mu)' | y]] \\
&= E_y[E[(x - E[x | y])(x - E[x | y])' | y] \\
&\quad + 2E[(x - E[x | y])(E[x | y] - \mu)' | y] \\
&\quad + E[(E[x | y] - \mu)(E[x | y] - \mu)' | y]]
\end{aligned}$$

The first term of $\operatorname{var}(x)$ is the expected variance of x given y :

$$\begin{aligned}
&= E_y[E[xx' | y] - 2(E[x | y])^2 + (E[x | y])^2] \\
&= E_y[\mathbf{var}_x(x | y)]
\end{aligned}$$

The second term of $\operatorname{var}(x)$ is equal to zero:

$$\begin{aligned}
&= 2E_y[(E[x | y])^2 - E[x | y]\mu' - (E[x | y])^2 + E[x | y]\mu'] \\
&= 0
\end{aligned}$$

The third term of $\text{var}(x)$ is the variance of the expected value of x given y :

$$\begin{aligned}
&= E_y[(E[x | y])^2 - E[x | y]\mu' - \mu(E[x | y])' + \mu\mu'] \\
&= E_y[(E[x | y] - \mu)(E[x | y] - \mu)'] \\
&= E_y[\text{var}_y(E_x[x | y])] \\
&= \text{var}_y(E_x[x | y])
\end{aligned}$$

(c) Alternative proof for BP of $\mathbf{g} = \mathbf{E}[\mathbf{g} | \mathbf{y}]$

The MSEP of $h(y)$ can be written, using result (a), as follows:

$$\begin{aligned}
E_y[(h(y) - g)'G(h(y) - g)] &= \text{tr } G \{E_y[(h(y) - g)(h(y) - g)']\} \\
&= \text{tr } G \{\text{var}(h(y) - g) \\
&\quad + (E_y[h(y) - g](E_y[h(y) - g])'\}
\end{aligned}$$

By result (b)

$$\text{var}(h(y) - g) = E_y[\text{var}_g((h(y) - g) | y)] + \text{var}_y(E_g[(h(y) - g) | y])$$

where

$$E_y[\text{var}_g((h(y) - g) | y)] = E_y[\text{var}_g(g | y)], \text{ and}$$

$$\text{var}_y(E_g[(h(y) - g) | y]) = \text{var}_y(h(y) - E_g[g | y])$$

because $h(y) | y$ is a constant.

Thus,

$$\text{var}(h(y) - g) = E_y[\text{var}(g | y)] + \text{var}(h(y) - E_g[g | y]),$$

and the MSEP of $h(y)$ is:

$$E_y[(h(y) - g)'G(h(y) - g)] = \text{tr } G \{E_y[\text{var}_g(g | y)] + \text{var}_g(h(y) - E_g[g | y])\}$$

$$+ (E_y[h(y)] - g)(E_y[h(y)] - g)'\}$$

Because the first and the third terms of the MSEP of $h(y)$ are constants, to minimize the MSEP of $h(y)$ we only need to minimize:

$$\text{var}(h(y) - E_g[g | y])$$

i.e., we want this term to go to zero.

Clearly,

$$\text{var}(h(y) - E_g[g | y]) = 0 \quad \text{if } h(y) = E_g[g | y]$$

⇒ the MSEP of $h(y)$ is minimized if $h(y)$ is equal to the conditional mean of g given y , and

⇒ the BP of g is $E[g | y] \equiv \hat{g}$.

As a consequence of $h(y)$ being equal to $E[g | y]$ we have that:

$$\begin{aligned} \text{(i)} \quad (E_y[h(y) - g])(E_y[h(y) - g])' &= E_y[(E_g[h(y) - g] E_g[h(y) - g]) | y] \\ &= E_y[(h(y) - E_g[g | y]) (h(y) - E_g[g | y])] \\ &= 0 \quad \text{when } h(y) = E_g[g | y] \end{aligned}$$

⇒ **the BP of g is unbiased.**

$$\begin{aligned} \text{(ii)} \quad E_y[(h(y) - g)'(h(y) - g)] &= \text{tr}\{\text{var}(h(y) - g)\} \\ &= \text{tr}\{E_y[\text{var}(g | y)]\} \quad \text{when } h(y) = E_g[g | y] \end{aligned}$$

⇒ **the MSEP of $h(y)$ = the error variance of prediction (EVP) of $h(y)$.**

Properties of the best predictor

$$[1] \quad E_y[\hat{g}] = E_y[E_g[g | y]] = E[g]$$

⇒ the BP is unbiased although it was **not** a condition in its development

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

$$[2] \quad \text{var}(\hat{g} - g) = E_y[\text{var}(g | y)]$$

Proof:

$$\begin{aligned}
 \text{var}(\hat{g} - g) &= E_y[(\hat{g} - g)(\hat{g} - g)'] \\
 &= E_y[\hat{g}\hat{g}' - \hat{g}g' - g\hat{g}' + gg'] \\
 &= E_y[E_g|y][(\hat{g}\hat{g}' - \hat{g}g' - g\hat{g}' + gg') | y]] \\
 &= E_y[E_g[g | y] E_g[g | y]' - E_g[g | y] E_g[g | y]' - E_g[g | y] E_g[g | y]' + E_g[gg' | y]] \\
 &= E_y[E_g[gg' | y] - E_g[g | y] E_g[g | y]'] \\
 &= E_y[\text{var}_g(g | y)]
 \end{aligned}$$

⇒ the EVP of \hat{g} is the weighted average of the variances of the elements of random vector g over all possible realizations of random vector y .

$$[3] \quad \text{var}(\hat{g}) = \text{var}_y(E_g[g | y])$$

Proof:

$$\begin{aligned}
 \text{var}(\hat{g}) &= E_y[\text{var}_g(\hat{g} | y)] + \text{var}_y(E[\hat{g} | y]) \\
 &= E_y[\text{var}_g(E_g[g | y])] + \text{var}_y(E_g[g | y]) \\
 &= E_y[0] + \text{var}_y(E_g[g | y]) \\
 &= \text{var}_y(E_g[g | y])
 \end{aligned}$$

⇒ the $\text{var}(\hat{g})$ is equal to the variance of the expected value of g given y .

$$[4] \quad \text{var}(g) = E_y[\text{var}_g(g | y)] + \text{var}_y(E_g[g | y])$$

$$\text{var}(g) = \text{var}(\hat{g} - g) + \text{var}(\hat{g}) \quad \text{by [2] and [3]}$$

$$\Rightarrow \quad \mathbf{var}(\hat{g} - g) = \mathbf{var}(g) - \mathbf{var}(\hat{g})$$

$$[5] \quad \text{cov}(\hat{g}, g') = \text{var}(\hat{g})$$

Proof:

Version 1:

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

But,

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

Thus,

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

$$\Rightarrow \text{cov}(\hat{g}, g') = \text{var}(\hat{g})$$

Version 2:

$$\begin{aligned} \text{cov}(\hat{g}, g') &= E[\hat{g} g'] - E[\hat{g}]E[g] \\ &= E[\hat{g} g'] - (E[\hat{g}])^2 \quad \text{by [1]} \end{aligned}$$

and

$$\begin{aligned} E[\hat{g} g'] &= \int_y \int_g \hat{g} g' f_{yg}(y, g) dg dy \\ &= \int_y \hat{g} \left[\int_g g f_{g|y}(g | y) dg \right] f_y(y) dy \\ &= \int_y \hat{g} E[g | y] f_y(y) dy \\ &= \int_y \hat{g} \hat{g}' f_y(y) dy \\ &= E[\hat{g} \hat{g}'] \end{aligned}$$

$$\Rightarrow \text{cov}(\hat{g}, g') = E[\hat{g} \hat{g}'] - (E[\hat{g}])^2$$

$$\Rightarrow \text{cov}(\hat{g}, g') = \text{var}(\hat{g})$$

$$[6] \quad r(\hat{g}, g') = [\text{cov}(\hat{g}, g')][\text{var}(\hat{g}) \text{var}(g)]^{-1/2}$$

Note: if $\text{var}(g)$ and $\text{var}(\hat{g})$ are positive definite (Property (6), Chapter 4), there are orthogonal matrices L and M , such that

$$[\text{var}(g)]^{-1/2} = \text{diag}\{(\lambda_{g_i})^{-1/2}\}L'$$

and

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

where the λ_{g_i} and the $\lambda_{\hat{g}_i}$ are the eigenvalues of the matrices $\text{var}(g)$ and $\text{var}(\hat{g})$ and L and M are the corresponding matrices of eigenvectors. But,

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

thus,

$$\begin{aligned} r(\hat{g}, g') &= [\text{var}(\hat{g})]^{1/2} [\text{var}(g)]^{-1/2} \\ &= \text{diag}\left\{\frac{\lambda_{\hat{g}_i}}{\lambda_{g_i}}\right\}L'M \end{aligned}$$

- \Rightarrow if the eigenvalues of $\text{var}(\hat{g})$ and $\text{var}(g)$ are the same, M , the inverse of the orthogonal matrix M' , will also be the inverse of L' , i.e., $M = L$,
- \Rightarrow $r(\hat{g}, g')$ will be an identity matrix, and
- \Rightarrow $r(\hat{g}, g')$ is maximized if the sets of eigenvalues of the $\text{var}(\hat{g})$ and $\text{var}(g)$ matrices are identical.

Also, recall that

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

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

Squaring both sides yields

[8-9]

$$\begin{aligned} r(\hat{g}, g') r(\hat{g}, g')' &= \text{var}(g) / \text{var}(g) - \text{var}(\hat{g} - g) / \text{var}(g) \\ &= I - \text{var}(\hat{g} - g) / \text{var}(g) \end{aligned}$$

and taking square roots of both sides gives

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

$$= I, \text{ if } \text{var}(\hat{g} - g) = 0$$

⇒ because the **BP** minimizes $\text{var}(\hat{g} - g)$, it also maximizes $r(\hat{g}, g')$.

$$[7] \quad \text{var}(\hat{g} - g) = [I - r(\hat{g}, g') r(\hat{g}, g')'] \text{var}(g)$$

Proof:

$$\begin{aligned} \text{var}(\hat{g} - g) &= \text{var}(g) - \text{var}(\hat{g}) \\ &= [[\text{var}(g)[\text{var}(g)]^{-1} - \text{var}(\hat{g})[\text{var}(g)]^{-1}] \text{var}(g) \end{aligned}$$

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

$$\text{var}(\hat{g} - g) = [I - r(\hat{g}, g') r(\hat{g}, g')'] \text{var}(g)$$

[8] **Selection rules:**

[8.1] **Cochran's rule:** select all animals in a population whose

$$\hat{g}_i = E[g_i | y_i] \geq t$$

where t is a truncation point chosen such that

$$P\{g_i = E[g_i | y_i] \geq t\} = s$$

where s is the selected fraction of the population, and $y_i = [y_{1i} y_{2i} \dots y_{ni}]'$.

Remarks:

(1) Cochran's rule requires the assumption that the cumulative distribution function of \hat{g}_i (i.e.,

$E[g_i | y_i]$ is continuous and monotone such that for any selected fraction s , $0 < s < 1$, there is only one t that satisfies $P\{\hat{g}_i \geq t\} = s$.

- (2) If the (g_i, y_i) sampled are IID, then selection of animals based on the $E[g_i | y_i] = \hat{g}_i$ maximizes $E_s(g)$, the expected genetic value of the animals in the selected fraction s .

Note that IID means that:

- (a) animals must have the same amount and type of information, and
- (b) animals must be unrelated.

[8.2] **Fernando's rule:** Select s individuals out of the n animals in the population using

$$\hat{g} = E[g | y], \text{ where } g = [g_1 \ g_2 \ \dots \ g_n]', \text{ and } y = [y_{11} \ y_{12} \ \dots \ y_{nq_n}].$$

Remarks:

- (1) Fernando's rule makes **no** assumptions on:
 - (a) the distribution of $(g | y)$, i.e., it holds for **any** distribution, and
 - (b) the quality and quantity of information for individual, i.e., animals may have unequal information and they can be related.
- (2) Selection of s out of n animals using $\hat{g} = E[g | y]$ maximizes $E_s(g)$, the expected genetic value of the s selected individuals.

[9] **Drawback of BP:** How to compute it?

Must know:

- (a) The conditional distribution of $(g | y)$, and
- (b) The parameters of the distribution.

However, when the joint distribution of (g, y) is multivariate normal, the form of the BP simplifies greatly. In such case,

- (a) The conditional mean of u is linear in y ,
- (b) The only parameters needed are the first and second moments, and
- (c) BP is identical computationally to the best linear prediction (BLP).

Thus, assume

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

Thus,

$$(g|y) \sim MVN \{ \mu_g + C'V^{-1}(y - \mu_y), G - C'V^{-1}C \}$$

$$\Rightarrow \quad \hat{g} = \mu_g + C'V^{-1}(y - \mu_y) \quad \text{under normality.}$$

Properties of the best predictor under normality

$$\begin{aligned} [1] \quad E_y[\hat{g}] &= E_y[\mu_g + C'V^{-1}(y - \mu_y)] \\ &= \mu_g + C'V^{-1}(\mu_y - \mu_y) \\ &= \mu_g \\ &= E[g] \quad (\text{"weak property of BP"}). \end{aligned}$$

$$[2] \quad E[g|\hat{g}] = E_g[g|y] = \hat{g}$$

Proof:

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

$$\begin{aligned} \text{cov}(\hat{g}, g') &= \text{cov}(C'V^{-1}y, g') \\ &= C'V^{-1}C \end{aligned}$$

$$\begin{aligned}
\Rightarrow E_y[g | \hat{g}] &= \mu_g + CV^{-1}C'(C'V^{-1}C)^{-1}(\hat{g} - E_y[\hat{g}]) \\
&= \mu_g + (\mu_g + C'V^{-1}(y - \mu_y) - \mu_g) \\
&= \mu_g + C'V^{-1}(y - \mu_y) \\
&= E_g[g | y] \\
&= \hat{g} \quad (\text{"strong property of BP"}).
\end{aligned}$$

$$\begin{aligned}
[3] \quad \text{var}(\hat{g} - g) &= \text{var}(g) - \text{var}(\hat{g}) \\
&= G - C'V^{-1}C
\end{aligned}$$

$$\begin{aligned}
[4] \quad \text{var}(\hat{g}) &= \text{var}(\mu_g + C'V^{-1}(y - \mu_y)) \\
&= C'V^{-1}VV^{-1}C \\
&= C'V^{-1}C
\end{aligned}$$

$$\begin{aligned}
[5] \quad \text{cov}(\hat{g}, g') &= \text{cov}(C'V^{-1}y, g') \\
&= C'V^{-1}C \\
&= \text{var}(\hat{g})
\end{aligned}$$

$$\begin{aligned}
[6] \quad r(\hat{g}, g') &= [\text{var}(\hat{g})]^{1/2} [\text{var}(g')]^{-1/2} \\
&= [C'V^{-1}C]^{1/2} [G]^{-1/2}
\end{aligned}$$

$$[7] \quad \text{var}(\hat{g} - g') = [I - C'V^{-1}CG^{-1}]G$$

References

Cochran, W. G. 1951. Improvement by means of selection. Proc. Second Berkeley Symp. Math. Stat. and Prob., pp 449-470.

Fernando, R. L. and D. Gianola. 1986. Optimal properties of the conditional mean as a selection

criterion. Theor. Appl. Genet. 72:822-825.

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

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).