

# Appendix 2

## The Multivariate Normal

Draft Version 1 December 2000, ©Dec. 2000, B. Walsh and M. Lynch

Please email any comments/corrections to: [jbwalsh@u.arizona.edu](mailto:jbwalsh@u.arizona.edu)

### THE MULTIVARIATE NORMAL DISTRIBUTION

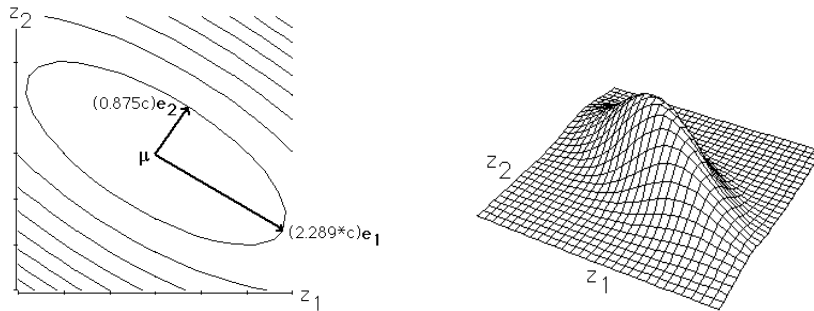
Recall the density of the multivariate normal distribution,

$$\phi(\mathbf{x}) = (2\pi)^{-n/2} |\Sigma_{\mathbf{X}}|^{-1/2} \exp \left[ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^T \Sigma_{\mathbf{X}}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right] \quad (\text{A2.19a})$$

Thus surfaces of equal probability for MVN distributed vectors satisfy

$$(\mathbf{x} - \boldsymbol{\mu})^T \Sigma_{\mathbf{X}}^{-1} (\mathbf{x} - \boldsymbol{\mu}) = c^2 \quad (\text{A2.19b})$$

From the discussion following Equation A2.17b, these surfaces are  $n$ -dimensional ellipsoids centered at  $\boldsymbol{\mu}$  whose axes of symmetry are given by the principal components (the eigenvectors) of  $\Sigma_{\mathbf{X}}$ . The length of the ellipsoid along the  $i$ th axis is  $c\sqrt{\lambda_i}$  where  $\lambda_i$  is the eigenvalue associated with the eigenvector  $\mathbf{e}_i$  (Figure A2.1).



**Figure A2.1.** Left: Surfaces of equal probability assuming that the additive genetic values associated with the characters  $z_1$  and  $z_2$  in Example 1 in Appendix 1 are  $\text{MVN}(\boldsymbol{\mu}, \mathbf{G})$ . These surfaces are ellipses centered at  $\boldsymbol{\mu}$ , with the major axis of the ellipse along  $\mathbf{e}_1$  and the minor axis along  $\mathbf{e}_2$ . Right: A plot of the associated

probability density. Slicing along either the major or minor axis gives a normal curve. Since the variance in the major axis is greater, the curve is much broader along this axis. The covariance between the breeding values of  $z_1$  and  $z_2$  rotates the distribution so that the principal axes do not coincide with the  $(z_1, z_2)$  axes.

Applying the canonical transformation (Equation A2.15a), we can change coordinate systems by a rigid rotation to remove any correlations between the variables in  $\mathbf{x}$ . Taking  $\mathbf{y} = \mathbf{U}^T(\mathbf{x} - \boldsymbol{\mu})$ ,

$$\mathbf{y} \sim \text{MVN}(\mathbf{0}, \boldsymbol{\Lambda}) \quad (\text{A2.20a})$$

where  $\boldsymbol{\Lambda}$  and  $\mathbf{U}$  are the matrices defined by Equations A2.10b/c for the diagonalization of  $\boldsymbol{\Sigma}_{\mathbf{x}}$ . In particular,

$$y_i = \mathbf{e}_i^T(\mathbf{x} - \boldsymbol{\mu}) \quad \text{where} \quad y_i \sim \text{N}(0, \lambda_i) \quad (\text{A2.20b})$$

Note from Equation A2.19 that since the  $y_i$  are uncorrelated, they are independent as the joint probability density is the product of  $n$  individual univariate normal densities. We can further transform the original vector by taking

$$y_i = \frac{\mathbf{e}_i^T(\mathbf{x} - \boldsymbol{\mu})}{\sqrt{\lambda_i}} \quad \text{giving} \quad y_i \sim \text{N}(0, 1) \quad (\text{A2.20c})$$

Thus, the transformation

$$\mathbf{y} = \boldsymbol{\Lambda}^{-1/2}\mathbf{U}^T(\mathbf{x} - \boldsymbol{\mu}) \quad (\text{A2.20d})$$

implies that  $\mathbf{y} \sim \text{MVN}(\mathbf{0}, \mathbf{I})$ , the elements of  $\mathbf{y}$  being  $n$  independent unit normal random variables.

## TESTING FOR MULTIVARIATE NORMALITY

While multivariate normality is often assumed, it is rarely tested. In Chapter 10 we briefly discussed two approaches for testing univariate normality, one graphical and the other based on if the observed skewness and/or kurtosis exceeds that expected for a Gaussian. Both of these can be extended to testing for multivariate normality. Additional methods are reviewed by Gnanadesikan (1977) and Seber (1984).

### Graphical Tests: Chi-square Plots

A fairly simple graphical test can be developed by extending the notion of the normal probability plot used to check univariate normality (Chapter 10). Recall that

in this case the observations are ranked and then plotted against their expected values under normality. Departures from linearity signify departures from normality.

We can apply this same approach to check for multivariate normality. From Equation A2.20d, if  $\mathbf{z} \sim \text{MVN}(\boldsymbol{\mu}, \boldsymbol{\Sigma}_{\mathbf{z}})$ , then each element of the vector

$$\mathbf{y} = \boldsymbol{\Lambda}^{-1/2} \mathbf{U}^T (\mathbf{z} - \boldsymbol{\mu})$$

is an independent unit normal ( $\mathbf{y} \sim \text{MVN}(\mathbf{0}, \mathbf{I})$ ). Solving for  $\mathbf{z}$  gives

$$(\mathbf{z} - \boldsymbol{\mu}) = \mathbf{U} \boldsymbol{\Lambda}^{1/2} \mathbf{y}$$

Using this and recalling Equation A2.11a,

$$\begin{aligned} (\mathbf{z} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}_{\mathbf{z}}^{-1} (\mathbf{z} - \boldsymbol{\mu}) &= (\mathbf{U} \boldsymbol{\Lambda}^{1/2} \mathbf{y})^T (\mathbf{U} \boldsymbol{\Lambda}^{-1} \mathbf{U}^T) (\mathbf{U} \boldsymbol{\Lambda}^{1/2} \mathbf{y}) \\ &= \mathbf{y}^T \boldsymbol{\Lambda}^{1/2} (\mathbf{U}^T \mathbf{U}) \boldsymbol{\Lambda}^{-1} (\mathbf{U}^T \mathbf{U}) \boldsymbol{\Lambda}^{1/2} \mathbf{y} \\ &= \mathbf{y}^T \mathbf{y} = \sum_{i=1}^n y_i^2 \end{aligned} \quad (\text{A2.21})$$

Thus if  $\mathbf{z} \sim \text{MVN}$ , the quadratic form given by Equation A2.21 is the sum of  $n$  independent squared unit normal random variables. By definition, this sum is a  $\chi^2$  random variable with  $n$  degrees of freedom (e.g., Morrison 1976), suggesting that one test for multivariate normality is to compare the goodness of fit of the scaled distances

$$d_i^2 = (\mathbf{z}_i - \bar{\mathbf{z}})^T \mathbf{S}_{\mathbf{z}}^{-1} (\mathbf{z}_i - \bar{\mathbf{z}}) \quad (\text{A2.22})$$

to a  $\chi_n^2$ . Here  $\mathbf{z}_i$  is the vector of observations from the  $i$ th individual,  $\bar{\mathbf{z}}$  the vector of sample means, and  $\mathbf{S}_{\mathbf{z}}^{-1}$  the inverse of the sample variance-covariance matrix. (We use the term distance because  $\boldsymbol{\Sigma}_{\mathbf{y}} = \mathbf{I}$ , giving the variance of any linear combination  $\mathbf{c}^T \mathbf{y}$  as  $\mathbf{c}^T \boldsymbol{\Sigma}_{\mathbf{y}} \mathbf{c} = \mathbf{c}^T \mathbf{I} \mathbf{c} = \|\mathbf{c}\|^2$ . Thus, regardless of orientation, any two  $\mathbf{y}$  vectors having the same length also have the same variance, which equals their squared Euclidean distance.) We can order these distances as

$$d_{(1)}^2 \leq d_{(2)}^2 \leq \dots \leq d_{(m)}^2$$

where  $m$  is the number of individuals sampled. Note that  $d_{(i)}^2$  is the  $i$ th smallest distance, whereas  $d_i^2$  is the distance associated with the vector of observations for the  $i$ th individual. Let  $\chi_n^2(\alpha)$  correspond to the value of a chi-square random variable  $x$  with  $n$  degrees of freedom that satisfies  $\text{Prob}(x \leq \chi_n^2(\alpha)) = \alpha$ . Under multivariate normality, we expect the points

$$\left( d_{(i)}^2, \chi_n^2 \left( \frac{i - 1/2}{m} \right) \right)$$

to fall along a line, as the  $i$ th ordered distance has  $i/m$  observations less than or equal to it (the factor of  $1/2$  is added as a correction for continuity). As with normal probability plots, departures from multivariate normality are indicated by departures from linearity. Complete tables of the  $\chi^2$  may be difficult to locate, in which case the appropriate  $\chi_n^2(\alpha)$  values can be numerically obtained using the incomplete gamma function (see Press et al. 1986 for details).

**Example 1.** Consider again the data of Jolicoeur and Mosimann (1960) on carapace characters in 24 male turtles. Are the characters  $z_1$  (carapace length) and  $z_2$  (maximum carapace width) jointly bivariate normally distributed? Here  $n = 2$  and  $m = 24$  and

$$\bar{\mathbf{z}} = \begin{pmatrix} 113.13. \\ 88.29 \end{pmatrix}, \quad \mathbf{S}_{\mathbf{z}} = \begin{pmatrix} 13.77 & 79.15 \\ 79.15 & 50.04 \end{pmatrix}, \quad \mathbf{S}_{\mathbf{z}}^{-1} = \begin{pmatrix} 0.0737 & -0.1165 \\ -0.1165 & 0.2043 \end{pmatrix}$$

where  $\mathbf{S}_{\mathbf{z}}$  is the sample covariance matrix. A partial list of the 24 vectors of observations are

$$\mathbf{z}_1 = \begin{pmatrix} 93 \\ 74 \end{pmatrix}, \quad \dots, \quad \mathbf{z}_{11} = \begin{pmatrix} 113 \\ 88 \end{pmatrix}, \quad \dots, \quad \mathbf{z}_{24} = \begin{pmatrix} 135 \\ 106 \end{pmatrix}$$

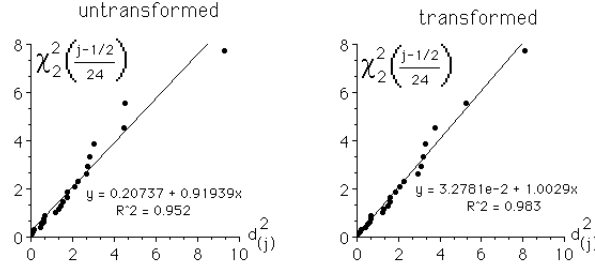
Applying Equation A2.22, these observations translate into the distances

$$d_1^2 = 4.45, \quad \dots, \quad d_{11}^2 = 0.002, \quad \dots, \quad d_{24}^2 = 9.277$$

After rank ordering, these correspond to  $d_{(23)}^2$ ,  $d_{(1)}^2$ , and  $d_{(24)}^2$ , respectively. For  $d_{(23)}^2$ , the matching value when distances are chi-squared distributed is

$$\chi_2^2 \left( \frac{23 - 1/2}{24} \right) = \chi_2^2(0.913.)$$

From chi-square tables, we find  $\text{Prob}(\chi_2^2 \leq 5.561) = 0.913$ . so that the data point generated from  $\mathbf{z}_1$  is  $(4.45, 5.561)$ . Likewise, the chi-square values for  $d_{(1)}^2$  and  $d_{(24)}^2$  are 0.043 and 7.727, respectively. Proceeding similarly for the other values, we obtain the curve plotted in Figure A2.2. This curve departs somewhat from linearity. Further, under the assumption of multivariate normality, the line is expected to pass through the origin, while the best linear fit of these data departs from the origin. Transforming the data by taking logs gives a slightly better fit to a MVN (Figure A2.2).



**Figure A2.2.** Plots of ranked distance data ( $d_{(j)}^2$  being the  $j$ th smallest distance) versus the expected corresponding  $\chi^2$  value for the data in Example 1. **Left:** The untransformed data does not appear to depart significantly from linearity. **Right:** Log-transforming the data gives a slightly better linear fit ( $r^2 = 0.983$  versus  $r^2 = 0.952$  for the untransformed data), with the best-fitting line passing through the origin as expected if the distance data follow a  $\chi^2$  distribution.

### Mardina's Test: Multivariate Skewness and Kurtosis

As was the case for univariate normality, we can test for multivariate normality by examining the skewness and kurtosis of the sample. Mardina (1970) proposed multivariate extensions of skewness and kurtosis and suggested a large sample test based on the asymptotic distribution of these statistics. Let  $\mathbf{z}_i$  be the  $i$ -th vector of observations,  $\bar{\mathbf{z}}$  the vector of sample means, and  $\mathbf{S}_{\mathbf{z}}$  sample covariance matrix. If there are  $m$  vectors of observations (with each vector measuring  $n$  characters), then the multivariate skewness is estimated by

$$b_{1,n} = \frac{1}{m^2} \sum_{i=1}^m \sum_{j=1}^m ((\mathbf{z}_i - \bar{\mathbf{z}})^T \mathbf{S}_{\mathbf{z}}^{-1} (\mathbf{z}_j - \bar{\mathbf{z}}))^3 \quad (\text{A2.23a})$$

while the multivariate kurtosis is estimated by

$$b_{2,n} = \frac{1}{m} \sum_{i=1}^m ((\mathbf{z}_i - \bar{\mathbf{z}})^T \mathbf{S}_{\mathbf{z}}^{-1} (\mathbf{z}_i - \bar{\mathbf{z}}))^2 \quad (\text{A2.23b})$$

If  $\mathbf{z} \sim \text{MVN}$ , then  $b_{1,n}$  and  $b_{2,n}$  have expected values 0 and  $n(n+2)$ . For large  $m$ , Mardina (1970) showed that the (scaled) multivariate skewness is asymptotically distributed as a chi-square random variable with  $f$  degrees of freedom, viz.,

$$\frac{m}{6} b_{1,n} \sim \chi_f^2, \quad \text{where } f = \frac{n(n+1)(n+2)}{6} \quad (\text{A2.24a})$$

Likewise for large  $m$ , the multivariate kurtosis (following appropriate scaling) is distributed as a unit-normal, viz.,

$$\frac{b_{2,n} - n(n+2)}{\sqrt{8n(n+2)/m}} \sim N(0,1) \quad (\text{A2.24b})$$

If either Equation A2.24a or A2.24b is significant, then multivariate normality is rejected.

**Example 2.** Again, let us examine the data of Jolicoeur and Mosimann (1960). Does the data considered in Example 1 display significant skewness or kurtosis? Here  $n = 2$  and  $m = 24$ . Applying Equations A2.24a/b gives  $b_{1,2} = 0.6792$ ,  $b_{2,2} = 7.6043$ . Considering skewness first,

$$\frac{m}{6} b_{1,2} = \frac{24}{6} 0.6792 = 2.717$$

is approximately chi-square distributed with  $f = 2(2+1)(2+2)/6 = 4$  degrees of freedom. Since  $\text{Prob}(\chi_4^2 \geq 2.717) \simeq 0.606$ , this is not significant. Turning to kurtosis, Equation A2.24b gives

$$\frac{b_{2,n} - n(n+2)}{\sqrt{8n(n+2)/m}} = \frac{7.6043 - 8}{1.633} \simeq -0.2423$$

which is also not significant as  $\text{Prob}(|N(0,1)| \geq 0.2423) \simeq 0.81$ . Transforming the data by taking logs gives  $b_{1,2} = 0.2767$  and  $b_{2,2} = 7.1501$ , improving the departure from skewness but increasing the departure from kurtosis. Applying Equations A2.24a/b gives 1.068 and  $-0.5204$ , again these are not significant. Reyment (1971) gives a number of other biological examples using Mardia's test.

## THE MULTIVARIATE BREEDERS' EQUATION

Recall from Chapter 7 that conditional distributions of subvectors from a multivariate normal are also multivariate normal. In particular, if we partition a MVN distributed vector  $n$ -dimensional column  $\mathbf{x}$  into two components, an  $m$ -dimensional column vector  $\mathbf{x}_1$  and an  $(n - m)$ -dimensional column vector  $\mathbf{x}_2$  of the remaining variables, e.g.,

$$\mathbf{x} = \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix}$$

where the mean vector and variance-covariance matrix are similarly be partitioned as

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix} \quad \text{and} \quad \mathbf{V} = \begin{pmatrix} \mathbf{V}_{\mathbf{x}_1\mathbf{x}_1} & \mathbf{V}_{\mathbf{x}_1\mathbf{x}_2} \\ \mathbf{V}_{\mathbf{x}_1\mathbf{x}_2}^T & \mathbf{V}_{\mathbf{x}_2\mathbf{x}_2} \end{pmatrix} \quad (\text{A2.25})$$

The conditional random variable  $\mathbf{x}_1|\mathbf{x}_2$  is distributed MVN with ( $m$ -dimensional) mean vector

$$\boldsymbol{\mu}_{\mathbf{x}_1|\mathbf{x}_2} = \boldsymbol{\mu}_1 + \mathbf{V}_{\mathbf{x}_1\mathbf{x}_2} \mathbf{V}_{\mathbf{x}_2\mathbf{x}_2}^{-1} (\mathbf{x}_2 - \boldsymbol{\mu}_2) \quad (\text{A2.26a})$$

and ( $m \times m$ ) variance-covariance matrix

$$\mathbf{V}_{\mathbf{x}_1|\mathbf{x}_2} = \mathbf{V}_{\mathbf{x}_1\mathbf{x}_1} - \mathbf{V}_{\mathbf{x}_1\mathbf{x}_2} \mathbf{V}_{\mathbf{x}_2\mathbf{x}_2}^{-1} \mathbf{V}_{\mathbf{x}_1\mathbf{x}_2}^T \quad (\text{A2.26b})$$

Likewise (Equation 7.28), the regression of  $\mathbf{x}_1$  on  $\mathbf{x}_1$  is given by

$$\mathbf{x}_1 = \boldsymbol{\mu}_1 + \mathbf{V}_{\mathbf{x}_1\mathbf{x}_2} \mathbf{V}_{\mathbf{x}_2\mathbf{x}_2}^{-1} (\mathbf{x}_2 - \boldsymbol{\mu}_2) + \mathbf{e} \quad (\text{A2.27a})$$

where

$$\mathbf{e} \sim \text{MVN}_m(\mathbf{0}, \mathbf{V}_{\mathbf{x}_1|\mathbf{x}_2}) \quad (\text{A2.27a})$$

A direct application of these results is the multivariate breeders' equation,  $\Delta\boldsymbol{\mu} = \mathbf{G}\boldsymbol{\beta}$ . Assume the vector  $\mathbf{z} = (z_1, z_2, \dots, z_n)^T$  of phenotypic values of characters in an individual can be written as  $\mathbf{z} = \mathbf{G} + \mathbf{e}$ , the sum of a vector of additive genetic (breeding) values  $\mathbf{g}$  plus an independent vector of environmental (and nonadditive genetic) values  $\mathbf{e}$ . Assuming  $\mathbf{x} \sim \text{MVN}(\boldsymbol{\mu}, \mathbf{G})$  and  $\mathbf{e} \sim \text{MVN}(\mathbf{0}, \mathbf{E})$ , then  $\mathbf{z} \sim \text{MVN}(\boldsymbol{\mu}, \mathbf{P})$  where  $\mathbf{P} = \mathbf{G} + \mathbf{E}$ .

In order to compute the expected change in  $\mathbf{z}$  due to selection, consider the distribution of breeding values conditioned on the observed phenotypic value. Since we assume  $\mathbf{g}$  and  $\mathbf{e}$  are independent,

$$\sigma(\mathbf{g}, \mathbf{z}) = \sigma(\mathbf{g}, \mathbf{g} + \mathbf{e}) = \sigma(\mathbf{g}, \mathbf{g}) = \mathbf{G}$$

the joint distribution of  $\mathbf{g}$  and  $\mathbf{z}$  is

$$\begin{pmatrix} \mathbf{g} \\ \mathbf{z} \end{pmatrix} \sim \text{MVN} \left( \begin{pmatrix} \boldsymbol{\mu} \\ \boldsymbol{\mu} \end{pmatrix}, \begin{pmatrix} \mathbf{G} & \mathbf{G} \\ \mathbf{G} & \mathbf{P} \end{pmatrix} \right) \quad (\text{A2.28})$$

In the notation of Equation A2.25,  $\mathbf{V}_{\mathbf{G}\mathbf{G}} = \mathbf{V}_{\mathbf{G}\mathbf{z}} = \mathbf{G}$  and  $\mathbf{V}_{\mathbf{z}\mathbf{z}} = \mathbf{P}$ . From Equations A2.26a/b, the conditional distribution of  $\mathbf{g}$  given  $\mathbf{z}$  is MVN with mean

$$\boldsymbol{\mu}_{\mathbf{G}|\mathbf{z}} = \boldsymbol{\mu} + \mathbf{G}\mathbf{P}^{-1}(\mathbf{z} - \boldsymbol{\mu}) \quad (\text{A2.29a})$$

and variance-covariance matrix

$$\mathbf{V}_{\mathbf{e}} = \mathbf{G} - \mathbf{G}\mathbf{P}^{-1}\mathbf{G} \quad (\text{A2.29b})$$

Alternatively, this can be restated as the regression of the vector of breeding values on the vector of phenotypic values,

$$\mathbf{g} - \boldsymbol{\mu} = \mathbf{GP}^{-1}(\mathbf{z} - \boldsymbol{\mu}) + \mathbf{e} \quad (\text{A2.30a})$$

where

$$\mathbf{e} \sim \text{MVN}(\mathbf{0}, \mathbf{V}_e) \quad (\text{A2.30b})$$

Given a vector of phenotypic observations  $\mathbf{z}$ , the expected vector of breeding values is  $\boldsymbol{\mu} + \mathbf{GP}^{-1}(\mathbf{z} - \boldsymbol{\mu})$ , while the actual vector of breeding values is distributed about this mean vector as a Gaussian with covariance matrix  $\mathbf{V}_e$ . The variance-covariance matrix of the residual vector  $\mathbf{e}$  is independent of the actual value of  $\mathbf{z}$ , and hence the regression of  $\mathbf{G}$  on  $\mathbf{z}$  is both linear (from Equation A2.29a) and homoscedastic (Equation A2.29b). Using our univariate notation of  $g = A$  and  $\mathbf{G} = \sigma_A^2$ ,

$$A - \mu = \sigma_A^2 \sigma_z^{-2} (z - \mu) + e = h^2 (z - \mu) + e \quad (\text{A2.30c})$$

where

$$\sigma_e^2 = \sigma_A^2 - \sigma_A^2 \sigma_z^{-2} \sigma_A^2 = \sigma_A^2 (1 - h^2) \quad (\text{A2.30d})$$

Taking expectations over all selected individuals, and assuming that all between-generation changes in character value are due to changes in breeding value,

$$\begin{aligned} \Delta \boldsymbol{\mu} &= E[\mathbf{GP}^{-1}(\mathbf{z} - \boldsymbol{\mu}) + \mathbf{e}] \\ &= \mathbf{GP}^{-1} E[(\mathbf{z} - \boldsymbol{\mu})] + E(\mathbf{e}) \\ &= \mathbf{GP}^{-1} \mathbf{s} = \mathbf{G}\boldsymbol{\beta} \end{aligned} \quad (\text{A2.31})$$

as obtained (in various forms) by Young and Weiler (1960), Harvey and Bearden (1962) and Lande (1979). It is important to note that all the caveats of the univariate breeder's equation (it is compromised if epistasis,  $G \times E$ , maternal effects, etc. are present) also apply to the multivariate breeders' equation.

**Example 3.** What is  $\mathbf{G}^*$ , the variance-covariance matrix of breeding values after selection (but before recombination and random mating) under the assumptions leading to the multivariate breeders' equation? From the definition of a covariance matrix,

$$\mathbf{G}^* = E((\mathbf{g} - \boldsymbol{\mu}^*)(\mathbf{g} - \boldsymbol{\mu}^*)^T)$$

where  $\boldsymbol{\mu}^*$  is the vector of phenotypic means following selection. Using, respectively, Equation A2.30a, the matrix identity  $(\mathbf{ABC})^T = \mathbf{c}^T \mathbf{B}^T \mathbf{A}^T$  (recalling that

$\mathbf{G}$  and  $\mathbf{P}^{-1}$  are symmetric), and expanding gives

$$\begin{aligned}\mathbf{G}^* &= E\left( [\mathbf{GP}^{-1}(\mathbf{z} - \boldsymbol{\mu}^*) + \mathbf{e}] [\mathbf{GP}^{-1}(\mathbf{z} - \boldsymbol{\mu}^*) + \mathbf{e}]^T \right) \\ &= E\left( [\mathbf{GP}^{-1}(\mathbf{z} - \boldsymbol{\mu}^*) + \mathbf{e}] [(\mathbf{z} - \boldsymbol{\mu}^*)^T \mathbf{P}^{-1} \mathbf{G} + \mathbf{e}^T] \right) \\ &= E\left( \mathbf{GP}^{-1}(\mathbf{z} - \boldsymbol{\mu}^*)(\mathbf{z} - \boldsymbol{\mu}^*)^T \mathbf{P}^{-1} \mathbf{G} \right) + E\left( \mathbf{GP}^{-1}(\mathbf{z} - \boldsymbol{\mu}^*) \mathbf{e}^T \right) \\ &\quad + E\left( \mathbf{e}(\mathbf{z} - \boldsymbol{\mu}^*)^T \mathbf{P}^{-1} \mathbf{G} \right) + E(\mathbf{e} \mathbf{e}^T)\end{aligned}$$

Using Equation 7.15b and the independence of  $\mathbf{e}$  and  $\mathbf{z}$ , this reduces to

$$\begin{aligned}\mathbf{G}^* &= \mathbf{GP}^{-1} E((\mathbf{z} - \boldsymbol{\mu}^*)(\mathbf{z} - \boldsymbol{\mu}^*)^T) \mathbf{P}^{-1} \mathbf{G} + \mathbf{GP}^{-1} E(\mathbf{z} - \boldsymbol{\mu}^*) E(\mathbf{e}^T) \\ &\quad + E(\mathbf{e}) E((\mathbf{z} - \boldsymbol{\mu}^*)^T) \mathbf{P}^{-1} \mathbf{G} + E(\mathbf{e} \mathbf{e}^T)\end{aligned}$$

This can be further simplified by noting that  $E(\mathbf{e}) = \mathbf{0}$  and that  $E[(\mathbf{z} - \boldsymbol{\mu}^*)(\mathbf{z} - \boldsymbol{\mu}^*)^T] = \mathbf{P}^*$  is the phenotypic variance-covariance matrix after selection. Finally, from Equation A2.29b we have  $E(\mathbf{e} \mathbf{e}^T) = \mathbf{V}_e$ , giving

$$\mathbf{G}^* = \mathbf{GP}^{-1} \mathbf{P}^* \mathbf{P}^{-1} \mathbf{G} + \mathbf{0} + \mathbf{0} + (\mathbf{G} - \mathbf{GP}^{-1} \mathbf{G})$$

Writing  $\mathbf{GP}^{-1} \mathbf{G} = \mathbf{GP}^{-1} \mathbf{P} \mathbf{P}^{-1} \mathbf{G}$  and factoring gives the within-generation change in the variance-covariance matrix of breeding values as

$$\begin{aligned}\mathbf{G}^* - \mathbf{G} &= \mathbf{GP}^{-1} \mathbf{P}^* \mathbf{P}^{-1} \mathbf{G} - \mathbf{GP}^{-1} \mathbf{P} \mathbf{P}^{-1} \mathbf{G} \\ &= \mathbf{GP}^{-1} (\mathbf{P}^* - \mathbf{P}) \mathbf{P}^{-1} \mathbf{G}\end{aligned}\tag{A2.32}$$

as obtained by Lande and Arnold (1983).

**Gaussian fitness functions.** As we observed in Chapter 28, selection generally introduces non-normality even if the initial distribution is Gaussian. Ideally, we would like to have a class of fitness functions that on one hand models directional, stabilizing, disruptive, and correlational selection and yet still preserves normality. The **Gaussian fitness function**,

$$\begin{aligned}W(\mathbf{z}) &= \exp\left( \boldsymbol{\alpha}^T \mathbf{z} - \frac{1}{2} (\mathbf{z} - \boldsymbol{\theta})^T \mathbf{W} (\mathbf{z} - \boldsymbol{\theta}) \right) \\ &= \exp\left( \sum_i \alpha_i z_i - \frac{1}{2} \sum_i \sum_j (z_i - \theta_i)(z_j - \theta_j) W_{ij} \right)\end{aligned}\tag{A2.33}$$

where  $\mathbf{W}$  is a symmetric matrix, is one such class. While this dates back to Weldon (1901) and Haldane (1954), this slightly more general form is due to Felsenstein (1977).

Directional selection occurs when  $\alpha \neq \mathbf{0}$  and/or  $\mu \neq \theta$ , while the elements of  $\mathbf{W}$  measure quadratic selection. If  $\mathbf{W}$  is a diagonal matrix, then  $W_{ii} > 0$  implies stabilizing selection on  $z_i$  about an optimal value of  $\theta_i$ , while  $W_{ii} < 0$  implies disruptive selection about  $\theta_i$ . The larger the magnitude of  $W_{ii}$ , the stronger selection. As we will see in Chapter 34, some care must be taken in interpreting the nature of the fitness surface when  $\mathbf{W}$  has non-zero off-diagonal elements. Note from our discussions on the canonical axes of a quadratic form (Equation A2.16) that, provided  $\mathbf{W}^{-1}$  exists, we can always transform the original vector of characters  $\mathbf{z}$  to a new vector  $\mathbf{y}$  such that this matrix is now diagonal with the diagonal elements being the eigenvalues of  $\mathbf{W}$ . The signs of these eigenvalues indicate whether selection is stabilizing or disruptive (positive eigenvalues indicate stabilizing selection, negative eigenvalues disruptive selection), while their magnitudes indicate the strength of selection (the larger the magnitude, the stronger the effect). If  $\mathbf{W}$  has  $k$  zero eigenvalues, the fitness surface has no curvature (is a plane) in  $k$  dimensions.

To see that the gaussian fitness function preserves normality, first note that if  $p(\mathbf{z})$  is the phenotypic distribution before selection, then the phenotypic distribution after selection is

$$p^*(\mathbf{z}) = \frac{p(\mathbf{z}) W(\mathbf{z})}{\int p(\mathbf{z}) W(\mathbf{z}) d\mathbf{z}} = c p(\mathbf{z}) W(\mathbf{z})$$

where  $c$  is a constant such that  $\int p^*(\mathbf{z}) d\mathbf{z} = 1$ . If the vector of phenotypic values before selection  $\mathbf{z} \sim \text{MVN}(\mu, \mathbf{P})$ , the distribution of phenotypes after selection is

$$\begin{aligned} p^*(\mathbf{z}) &= c p(\mathbf{z}) W(\mathbf{z}) \\ &= c \exp\left(\alpha^T \mathbf{z} - \frac{1}{2} (\mathbf{z} - \theta)^T \mathbf{W} (\mathbf{z} - \theta)\right) \cdot \exp\left(-\frac{1}{2} (\mathbf{z} - \mu)^T \mathbf{P}^{-1} (\mathbf{z} - \mu)\right) \\ &= c \exp\left(\alpha^T \mathbf{z} - \frac{1}{2} [(\mathbf{z} - \theta)^T \mathbf{W} (\mathbf{z} - \theta) + (\mathbf{z} - \mu)^T \mathbf{P}^{-1} (\mathbf{z} - \mu)]\right) \end{aligned} \quad (\text{A2.34a})$$

If this distribution is Gaussian, it can be written as

$$p^*(\mathbf{z}) = (2\pi)^{-n/2} |\mathbf{P}^*|^{-1/2} \exp\left(-\frac{1}{2} (\mathbf{z} - \mu^*)^T (\mathbf{P}^*)^{-1} (\mathbf{z} - \mu^*)\right) \quad (\text{A2.34b})$$

where  $\mu^*$  and  $\mathbf{P}^*$  are the mean vector and phenotypic covariance matrix after selection. By matching terms in Equations A2.34a/b and doing a little algebra we find that these two distributions are indeed identical, with

$$\mu^* = \mathbf{P}^* (\mathbf{P}^{-1} \mu + \mathbf{W} \theta + \alpha) \quad (\text{A2.35a})$$

and

$$\mathbf{P}^* = (\mathbf{P}^{-1} + \mathbf{W})^{-1} \quad (\text{A2.35b})$$

If  $\mathbf{W}^{-1}$  exists, we can alternatively write  $\mathbf{P}^*$  as

$$\mathbf{P}^* = \mathbf{W}^{-1}(\mathbf{W}^{-1} + \mathbf{P})^{-1}\mathbf{P} \quad (\text{A2.35c})$$

$$= \mathbf{P}(\mathbf{W}^{-1} + \mathbf{P})^{-1}\mathbf{W}^{-1} \quad (\text{A2.35d})$$

These follow from standard matrix identities (e.g., Searle 1982) and will prove useful in further analysis of this model. Note that since  $\mathbf{P}^*$  is a covariance matrix, it must be positive-definite. This is always the case if  $\mathbf{W}$  is non-negative definite, but can fail if  $\mathbf{W}$  has at least one sufficiently large negative eigenvalue, as would occur with sufficiently strong disruptive selection on at least one character. In this case, the Gaussian fitness function is not appropriate (in effect, it means for at least one character, fitness approaches infinity as that character gets arbitrarily large or small at a faster rate than the frequency of that character value approaches zero, resulting in a mean population fitness of infinity).

Finally, it will be useful to have an expression for the mean population fitness under Gaussian selection, which is a function of the mean vector and covariance matrix. By definition,

$$\bar{W}(\boldsymbol{\mu}, \mathbf{P}) = \int W(\mathbf{z}) p(\mathbf{z}) d\mathbf{z}$$

To compute this integral, we use the standard trick of expressing this as a probability distribution times a function independent of  $\mathbf{z}$ . Since

$$\int (2\pi)^{-n/2} |\mathbf{P}^*|^{-1/2} \exp\left(-\frac{1}{2}(\mathbf{z} - \boldsymbol{\mu}^*)^T (\mathbf{P}^*)^{-1} (\mathbf{z} - \boldsymbol{\mu}^*)\right) d\mathbf{z} = 1$$

we find (after a little algebra) that when  $\mathbf{z} \sim \text{MVN}$ ,

$$\bar{W}(\boldsymbol{\mu}, \mathbf{P}) = \sqrt{\frac{|\mathbf{P}^*|}{|\mathbf{P}|}} \cdot \exp\left(-\frac{1}{2} \left[ \boldsymbol{\theta}^T \mathbf{W} \boldsymbol{\theta} + \boldsymbol{\mu}^T \mathbf{P}^{-1} \boldsymbol{\mu} - (\boldsymbol{\mu}^*)^T (\mathbf{P}^*)^{-1} \boldsymbol{\mu}^* \right]\right) \quad (\text{A2.36a})$$

Using Equation A2.35a, we can alternately express the term in the exponent as

$$\boldsymbol{\theta}^T \mathbf{W} \boldsymbol{\theta} + \boldsymbol{\mu}^T \mathbf{P}^{-1} (\mathbf{I} - \mathbf{P}^* \mathbf{P}^{-1}) \boldsymbol{\mu} - 2 \cdot \mathbf{b}^T \mathbf{P}^{-1} \boldsymbol{\mu} - \mathbf{b}^T \mathbf{P}^{-1} \mathbf{b} \quad (\text{A2.36b})$$

where  $\mathbf{b} = \mathbf{W} \boldsymbol{\theta} + \boldsymbol{\alpha}$ , and we have used  $\mathbf{b}^T \mathbf{P}^{-1} \boldsymbol{\mu} = (\mathbf{b}^T \mathbf{P}^{-1} \boldsymbol{\mu})^T = \boldsymbol{\mu}^T \mathbf{P}^{-1} \mathbf{b}$  (Equation 7.18).

The careful reader will observe that we have only shown that **phenotypes** remain Gaussian-distributed. Our main interest is how selection changes the distribution of the vector of additive genetic values  $\mathbf{g}$  as these determine the between-generation change. Provided  $\mathbf{z} = \mathbf{g} + \mathbf{e}$ , where  $\mathbf{g}$  and  $\mathbf{e}$  are independent and MVN,

then the distribution of additive genetic values after selection also remains MVN. To see this, consider the expected fitness of an individual with additive genetic value  $\mathbf{g}$ ,

$$w(\mathbf{g}) = \int w(\mathbf{z}) p(\mathbf{z} | \mathbf{g}) d\mathbf{z}$$

The conditional distribution of  $\mathbf{z}$  given  $\mathbf{g}$  is  $\mathbf{z} | \mathbf{g} = (\mathbf{g} + \mathbf{e}) | \mathbf{g} \sim \text{MVN}(\mathbf{g}, \mathbf{E})$ , as phenotypes are distributed around their genotypic value  $\mathbf{g}$  according to the distribution of environmental values, which is  $\text{MVN}(\mathbf{0}, \mathbf{E})$ . Using the same approach leading to Equation A2.34, we can see that  $w(\mathbf{g})$  is also of a Gaussian form and hence  $p^*(\mathbf{g}) = w(\mathbf{g}) p(\mathbf{g})$  is also Gaussian.

To conclude, let us combine these results with those obtained above for the multivariate breeders' equation to examine changes in the vector of means and the covariance matrix. Consider changes in mean first. Since the distribution of additive genetic and phenotypic values remains Gaussian, then from the breeders' equation the expected change in the mean in generation  $t$  is  $\Delta\boldsymbol{\mu}_t = \mathbf{G}_t \mathbf{P}_t^{-1} \mathbf{s}_t$ . Noting that  $\mathbf{s}_t = \boldsymbol{\mu}_t^* - \boldsymbol{\mu}_t$ , from Equations A2.35a-c we have

$$\begin{aligned} \mathbf{s}_t &= \mathbf{W}^{-1} (\mathbf{W}^{-1} + \mathbf{P}_t)^{-1} \mathbf{P}_t (\mathbf{P}_t^{-1} \boldsymbol{\mu}_t + \mathbf{W}\boldsymbol{\theta} + \boldsymbol{\alpha}) - \boldsymbol{\mu}_t \\ &= \mathbf{W}^{-1} (\mathbf{W}^{-1} + \mathbf{P}_t)^{-1} (\boldsymbol{\mu}_t + \mathbf{P}_t \mathbf{W}\boldsymbol{\theta} + \mathbf{P}_t \boldsymbol{\alpha}) - \boldsymbol{\mu}_t \end{aligned}$$

Using the identity

$$\mathbf{I} = \mathbf{W}^{-1} (\mathbf{W}^{-1} + \mathbf{P})^{-1} (\mathbf{W}^{-1} + \mathbf{P}) \mathbf{W} = \mathbf{W}^{-1} (\mathbf{W}^{-1} + \mathbf{P})^{-1} (\mathbf{I} + \mathbf{P}\mathbf{W})$$

this reduces to

$$\begin{aligned} \mathbf{s}_t &= \mathbf{W}^{-1} (\mathbf{W}^{-1} + \mathbf{P}_t)^{-1} (\boldsymbol{\mu}_t + \mathbf{P}_t \mathbf{W}\boldsymbol{\theta} + \mathbf{P}_t \boldsymbol{\alpha}) - \mathbf{W}^{-1} (\mathbf{W}^{-1} + \mathbf{P}_t)^{-1} (\mathbf{I} + \mathbf{P}_t \mathbf{W}) \boldsymbol{\mu}_t \\ &= \mathbf{W}^{-1} (\mathbf{W}^{-1} + \mathbf{P}_t)^{-1} ([\mathbf{I} - \mathbf{I} - \mathbf{P}_t \mathbf{W}] \boldsymbol{\mu}_t + \mathbf{P}_t \mathbf{W}\boldsymbol{\theta} + \mathbf{P}_t \boldsymbol{\alpha}) \\ &= \mathbf{W}^{-1} (\mathbf{W}^{-1} + \mathbf{P}_t)^{-1} \mathbf{P}_t (\mathbf{W}(\boldsymbol{\theta} - \boldsymbol{\mu}_t) + \boldsymbol{\alpha}) \end{aligned} \quad (\text{A2.37a})$$

At equilibrium  $\hat{\mathbf{s}} = \mathbf{0}$ . Assuming  $\mathbf{W}$  and  $\mathbf{P}$  are nonsingular, this implies  $\mathbf{W}(\boldsymbol{\theta} - \hat{\boldsymbol{\mu}}) + \boldsymbol{\alpha} = \mathbf{0}$ , giving

$$\hat{\boldsymbol{\mu}} = \boldsymbol{\theta} + \mathbf{W}^{-1} \boldsymbol{\alpha} \quad (\text{A2.37b})$$

Surprisingly, this equilibrium value is independent of the additive genetic and phenotypic covariance matrices (provided  $\mathbf{G}$  and  $\mathbf{P}$  are nonsingular) and simply depends on the elements of the fitness function. This interesting result was first noted by Zeng (1988) for a slightly different fitness function.

Turning now to the change in  $\mathbf{G}$ , we first note that

$$\begin{aligned} \mathbf{P}^* &= \mathbf{P} (\mathbf{P} + \mathbf{W}^{-1})^{-1} \mathbf{W}^{-1} = \mathbf{P} (\mathbf{P} + \mathbf{W}^{-1})^{-1} ([\mathbf{P} + \mathbf{W}^{-1}] - \mathbf{P}) \\ &= \mathbf{P} - \mathbf{P} (\mathbf{P} + \mathbf{W}^{-1})^{-1} \mathbf{P} \end{aligned} \quad (\text{A2.13a})$$

Thus

$$\mathbf{P}^{-1} (\mathbf{P}^* - \mathbf{P}) \mathbf{P}^{-1} = - (\mathbf{W}^{-1} + \mathbf{P})^{-1} \quad (\text{A2.13b})$$

Substituting this result into Equation A2.32 gives

$$\mathbf{G}^* = \mathbf{G} - \mathbf{G} (\mathbf{W}^{-1} + \mathbf{P})^{-1} \mathbf{G} \quad (\text{A2.13c})$$

We defer discussion on how this **within-generation** change is passed between generations until later chapters.