Chapter 3

Basic Principles of Response to Selection

3.1 Introduction

When comparing different breeding programs the first question usually asked is "what are the expected responses to selection of the various plans". A considerable part of this course will focus on methods of designing breeding programs, which maximize response to selection. Although breeding plans are often quite complex, most can usually be understood in terms of a few simple principles of response to selection. In this chapter we briefly review these principles as a foundation for what follows in the rest of the course.

As in many fields of science, there are often many different ways of deriving a particular result. If you are familiar with the basic principles of quantitative genetics (e.g. as in Falconer and Mackay, 1996), the results given here should be familiar to you. However, the approach used here is slightly different to that given in other texts. You should be familiar with the derivations given in texts such as Falconer and Mackay (1996), as those derivations are generally more rigorous and go back to first principles. However, the derivations given in this course will often be more useful when it comes to designing breeding strategies and deriving statistics necessary for such designs.

3.2 Predicting Genetic Merit of Progeny

The basic guiding principle behind genetic improvement and predicting response to selection is that parents with high additive genetic values (breeding values) tend to have progeny with high additive genetic values (and therefore high phenotypes). This follows from the quantitative genetic model for the additive genetic value of progeny:

\[ g_o = \frac{1}{2} g_s + \frac{1}{2} g_d + g_m \]  

(3.1)

where \( g_s \) and \( g_d \) are the additive genetic values of the sire and dam and \( g_m \) is the Mendelian sampling contribution, as described in the previous chapter.

Since \( E(g_m) = 0 \), the expectation of the progeny additive genetic value, \( E(g_o) \), from a given pair of parents is given by

\[ E(g_o) = \frac{1}{2} g_s + \frac{1}{2} g_d \]

i.e., the expected additive genetic value of the progeny is equal to the mean additive genetic value of the two parents.
For determining response to selection, we are interested in the mean of the genetic value of the progeny generation, \( E(\overline{g}_o) \). This can be obtained from the average genetic value of the selected parents \( \overline{g}_s \) and \( \overline{g}_d \), where * indicates that the variable refers to selected individuals:

\[
E(\overline{g}_o) = \frac{1}{2} \overline{g}_s + \frac{1}{2} \overline{g}_d
\]  

(3.2)

For the purpose of understanding and predicting response to selection, it is useful to express the mean genetic value of selected parents in terms of a deviation from the mean genetic value of all individuals from which they were selected (\( g_s \) and \( g_d \)):

Thus:

\[
E(\overline{g}_o) = \frac{1}{2}(\overline{g}_s - g_s) + \frac{1}{2}(\overline{g}_d - g_d)
\]

Here, \( S \) is the genetic superiority of the selected parents, which is defined as the difference between the mean genetic value of the selected individuals from the mean of the group they were selected from, e.g.:

\[
S = \overline{g}_s - g_s
\]  

(3.4)

Response to selection is defined as the difference of the mean genetic value of progeny of selected parents from the mean genetic value of progeny of all possible parents. Response is often denoted as \( R \) or \( \Delta g \). Using the \( R \) notation, the expectation of \( R \) is given by:

\[
E(R) = \overline{g}_o - \overline{g}_p
\]  

(3.5)

Where

\[
\overline{g}_p = \frac{1}{2}(\overline{g}_s + \overline{g}_d)
\]

Using this and the expression of \( \overline{g}_o \) in terms of means of the parental generation and genetic superiorities of the selected parents (equation 3.3), expected response from the current to the next generation simplifies to:

\[
E(R) = \frac{1}{2}(\overline{g}_s + \overline{g}_d) + \frac{1}{2}(S_s + S_d) - \frac{1}{2}(\overline{g}_s + \overline{g}_d)
\]

(3.6)

Thus, expected response from the current to the next generation is determined entirely by genetic superiority of the selected parents.

Note that for the simple case of equal selection in males and females, \( S_s = S_d = S \) and \( E(R) = S \).
In general we do not know the genetic value of parents. But we may have a prediction of their genetic value through an estimated breeding value (EBV), \( \hat{g} \). Usually this prediction is based on a recognized method of genetic evaluation using different sources of phenotypic information. Examples are simple phenotypic selection, family index selection, pedigree index selection, BLUP, and so on. Whatever the method used, provided the estimate is unbiased, i.e. that

\[
E(g | \hat{g}) = \hat{g}
\]

then the expectation of the genetic value of an individual progeny is equal to the mean of the parental predictions, i.e.

\[
E(g_o) = \frac{1}{2} \hat{g}_s + \frac{1}{2} \hat{g}_d = \hat{g}_p
\]

where \( \hat{g}_p \) is the mean estimated genetic value of the two parents.

Then, the expected mean genetic value of the progeny generation can be written in terms of the mean EBV of the selected and all parents by replacing \( \bar{g} \) in (3.2) and (3.3) by \( \bar{g} \) as:

\[
E(\bar{g}_o) = \frac{1}{2} \bar{g}_s^* + \frac{1}{2} \bar{g}_d^*
\]

\[
= \frac{1}{2}(\bar{g}_s^* + \hat{S}_s) + \frac{1}{2} (\bar{g}_d^* + \hat{S}_d)
\]

(3.7)

Where \( \hat{S} \) is the estimated genetic superiority of the selected parents, which can be obtained from (3.4) as:

\[
\hat{S} = \bar{g}^* - \bar{g}
\]

(3.8)

Similarly, knowing the EBV of the parents, response from the current to the next generation can be predicted based on (3.5) and (3.6) as:

\[
E(R) = \hat{g}_o - \hat{g}_p = \frac{1}{2}(\hat{S}_s + \hat{S}_d)
\]

(3.9)

It should be noted that equation (3.1) can be extended back so that the sire and dam terms are replaced by their respective sire and dam terms (i.e. grandsires and grandams of individual \( i \)) and so on back through the ancestor pathways, e.g.

\[
g_o = \frac{1}{2} (\frac{1}{2} g_{ss} + \frac{1}{2} g_{ds} + g_{ms}) + \frac{1}{2} (\frac{1}{2} g_{sd} + \frac{1}{2} g_{dd} + g_{md}) + g_m
\]

(3.10)

where \( ss \) is sire of the sire, \( ds \) is dam of the sire, etc., and \( g_{ms} \) and \( g_{md} \) are the sire and dam Mendelian sampling terms.

However, the expectation of \( g_o \) in terms of \( \hat{g}_s \), \( \hat{g}_d \) in cannot easily be pushed back to include grandparental \( \hat{g} \) terms since the expectation of these terms depends on the degree of selection of the parents. However, solutions to most problems of design of breeding programs can be found using the parent-offspring relationships.
3.3 Predicting Response per Generation

The previous section allows us to predict response to selection if we have a particular group of chosen parents. This can be useful where we have an existing population of real animals and we want to predict the effects of choosing different combinations of animals as parents from that population. For example, in dairy cattle we might have several hundred bulls available for use, each with an estimated breeding value for milk yield. Assuming that the genetic evaluation procedure is unbiased, we could ask the consequences of using different numbers of bulls. Should we use the best 10 available or the best 20? Semen price is often (but not always!) related to quality, so that the top 10 bulls will often be more expensive than the next best 10 bulls. We could then ask how much genetic improvement would we expect when using the cheaper second set of 10 bulls rather than using the more expensive 10 best bulls. We will return to this problem later.

In many cases we are not interested in a particular group of existing animals but in predicting response to selection in future generations or in the consequences of different designs of animal breeding programs. We might ask, if we had a population of 100 bulls (which do not yet exist), what would be the expected response to selection if we use only the best 10 in comparison to using the best 20 every generation? The problem is then to predict the genetic superiority ($S$) of different types of possible parents in a hypothetical population as a result of a particular selection program.

A selection program typically is described by the fraction or number of males and females that are selected and by the criterion on which they are selected. Our objective here is to develop theory that can be used to predict the genetic superiority of selected parents based on this information.

We can assume that in this hypothetical population we have an estimate of each animal’s genetic value, which we will call an index value that is used as the selection criterion. We do not need to know at this stage how this index is derived. But we will assume that there is a linear relationship between the index value and the true genetic value. We can then derive predictions of genetic superiorities of selected parents based on standard regression theory.

A standard equation for the regression of a dependent variable, $y$, on an independent variable, $x$, takes the form

$$y_i = a + b_{yx}x_i + e_i$$  \hspace{1cm} (3.11)

and a prediction of $y$ given $x$ is

$$\hat{y}_i = \bar{y} + b_{yx}(x_i - \bar{x})$$  \hspace{1cm} (3.12)

where $\bar{y}$ is the mean value of $y$ over all values of $x$, $\bar{x}$ is the mean value of $x$ in the population of all possible values, and $x_i$ is the observed value of $x$ for the $i^{th}$ individual for whom we wish to predict a value of $y$. From standard regression theory, the regression coefficient, $b_{yx}$, of $y$ on $x$ is given by
\[ b_{yx} = \frac{\sigma_{xy}}{\sigma_x^2} = r_{xy} \frac{\sigma_y}{\sigma_x} \]  
(3.13)

where \( \sigma_{xy} \) is the covariance of \( x \) and \( y \), \( \sigma_x^2 \) is the variance of \( x \), and \( r_{xy} \) is the correlation between \( y \) and \( x \), which is given by

\[ r_{xy} = \frac{\sigma_{xy}}{\sqrt{\sigma_y^2 \sigma_x^2}} \]  
(3.14)

In our breeding problem, we want to predict the genetic value of an individual (that will become a parent) given a recorded or estimated index value, \( I_i \). Hence from (3.12),

\[ \hat{g}_i = g + b_{gl} (I_i - \bar{I}) \]  
(3.15)

where \( I_i \) is the index value of individual \( i \), \( \bar{g} \) is the mean genetic value of individuals in the population, \( \bar{I} \) is the mean index value of individuals in the population, and \( b_{gl} \) is the regression of genetic values on index values.

If we are predicting the average genetic value of a group of selected (chosen) animals, we get:

\[ \bar{g}^* = \bar{g} + b_{gl} (\bar{I}^* - \bar{I}) \]  
(3.16)

To obtain a prediction of the genetic superiority of the selected parents, we can substitute (3.16) into (3.8), recalling that it is the genetic value of parents we are predicting, to get:

\[ \hat{S} = \hat{g}^* - \bar{g} = b_{gl} (\bar{I}^* - \bar{I}) \]  
(3.17)

The right-hand side of equation (3.17) in parentheses, \((\bar{I}^* - \bar{I})\), is the deviation of index values of selected animals from the mean index value of all animals in the population. We can define the intensity of selection, \( i \), as the deviation of selected from average animals in standard deviation units, i.e.

\[ i = (\bar{I}^* - \bar{I}) / \sigma_I \]  
(3.18)

where \( \sigma_I \) is the standard deviation of index values. It then follows from (3.18) that

\[ (\bar{I}^* - \bar{I}) = i \sigma_I \]  
(3.19)

and substituting (3.18) into (3.17) we get

\[ \hat{S} = b_{gl} i \sigma_I \]  
(3.20)

From standard regression theory (equation 3.13), we recall that

\[ b_{gl} = r_{gl} \frac{\sigma_g}{\sigma_I} \]  
(3.21)
hence,
\[ S = r_{gl} \frac{\sigma_g}{\sigma_I} (i \sigma_i) = i r_{gl} \sigma_g \]  (3.22)

Equation (3.22) gives a general formula to predict genetic superiorities of selected parents, which are needed to predict the response to selection. This formula applies whenever the value on which animals are selected, \( I \), is linearly related to their additive genetic value. Predicted superiorities can be used to model the genetic level of future generations in a recursive manner using equation (3.7):

\[
E(\hat{g}_o) = \frac{1}{2}(\hat{g}_s + \hat{S}_d) = \frac{1}{2}(\hat{g}_s + i_s r_{gi} \sigma_g) + \frac{1}{2}(\hat{g}_d + i_d r_{gi} \sigma_d) \]  (3.23)

or model response per generation using equation (3.9):

\[
R = \frac{1}{2}(S_s + S_d) = \frac{1}{2}(i_s r_{gi} \sigma_g + i_d r_{gi} \sigma_d) \]  (3.24)

Methods to derive the accuracy of selection, \( r_{gl} \), based on various sources of information will be reviewed and developed in Chapter 4. To illustrate, its derivation for the simplest case, phenotypic selection based on own phenotype, will be given in section 3.4. The intensity of selection, \( i \), can be obtained from Normal distribution theory and will be further discussed in section 3.6. For the moment, we will assume that the genetic standard deviation, \( \sigma_g \), is known and remains constant over generations. The latter assumption will be relaxed in Chapter 5.

In the remainder of this chapter, we will first illustrate equation (3.22) for phenotypic selection, then present how equation (3.23) fits in a general diagram for a deterministic simulation model, followed by a discussion of approximations for intensity of selection, and finally develop extensions of this equation to prediction of response with selection across multiple age groups, response per unit of time, and correlated response to selection.

### 3.4 Example of Phenotypic Selection

The generality of equation (3.22) can be seen by considering the specific and familiar case of phenotypic selection. In this case, the index value, \( I \), is simply the phenotype of the animal. Assuming only additive genetic and random environmental effects, and assuming phenotype is adjusted for fixed effects (e.g. the mean), we can write the phenotypic value of an animal, \( y_i \) as

\[ y_i = g_i + e_i \]

where \( e_i \) is the environmental effect, assumed uncorrelated with the additive genetic effect, \( g_i \). Then,

\[ \sigma_{g_i} = \sigma_{gy} = \sigma_{g,y,e} = \sigma_g^2 \]

Thus

\[ r_{gl} = r_{gy} = \frac{\sigma_g^2}{\sqrt{\sigma_g^2 \sigma_p^2}} = \frac{\sigma_g}{\sigma_p} = h \]  (3.25)
Where \( h \) is the square root of heritability.

Thus, from (3.22),

\[
\hat{S} = i h \sigma_g
\]   (3.26)

Recalling that heritability is \( h^2 = \frac{\sigma^2_g}{\sigma^2_p} \), we get

\[
\hat{S} = i h^2 \sigma_p
\]   (3.27)

Equation (3.27) should be familiar as the standard form for prediction of response to phenotypic selection. What we have shown here is that this standard response to phenotypic selection is just a special case of the general form of response to selection given by equation (3.22).

### 3.5 Simple Deterministic Model for Predicting Response to Selection with Multiple Age Groups

A general schematic for a simple deterministic simulation of a breeding program is given in Figure 3.1. Comparing to Figure 2.1 for a stochastic simulation, it should be clear that while the general flow of deterministic and stochastic simulations are similar, their fundamental nature is quite different. Whereas stochastic simulations model individual animals and their genetic and phenotypic characteristics, deterministic simulations model means and variances of genetic and phenotypic characteristics of groups of individuals. Recurrence equations such as equation (3.23) for computing the mean genetic value of progeny are used to compute characteristics of progeny. Other recursive equations, such as those for variances, will be presented in later Chapters. Another important component of deterministic simulations is the derivation of the means and variances of the selection criterion that is used. Variance of the selection criterion depends on the accuracy of selection. Methods to derive accuracy of selection are presented in Chapter 4.

**Figure 3.1** General schematic of a deterministic simulation of a breeding program.

1. Define means and variances of base population.
2. Derive means and variances of selection criteria.
3. Derive proportions selected from each available group of animals.
4. Derive means and variances of selection criteria of all groups of selected parents.
5. Derive means and variances of underlying traits of selected parents.
6. Derive means and variances of resulting progeny → \( if \ time < t \)

\[
if \ time = t
\]

7. Output results and stop program.
It is clear that, by modeling means and variances, deterministic simulations are computationally less demanding than stochastic models, besides the fact that deterministic models give expected responses and are not subject to stochastic variation in response. However, to accurately model all aspects of a breeding program deterministically does require more complicated models. Some of these will be described in the remainder of this chapter, while others follow in later chapters.

3.6 Selection Intensity with Truncation Selection

The prediction of response to selection given by (3.24) does not require that we know how animals are selected, merely that we know the mean index value of selected animals and hence are able to derive the intensity of selection, \( i \).

Generally in animal breeding we consider the special case of truncation selection. In this form of selection, all animals above a certain index value, \( x \), are chosen for breeding and all animals below this value are discarded. Usually the truncation point is determined by the proportion, \( p \), of animals to be used for breeding. In many cases, index values will be normally distributed. If so, and under the assumption of large population size, the relationships between \( p \), \( x \) (measured in s.d. units), and \( i \) can be derived from the properties of the normal distribution to be equal to:

\[
  i = z/p
\]

where \( z \) is the height of the normal distribution at the truncation point \( x \) and is given by

\[
  z = \frac{e^{-x^2/2}}{\sqrt{2\pi}}
\]

and \( \pi \), to 9 decimal places, is 3.141592654.

For individual cases it is often convenient to look up the intensity of selection corresponding to a particular proportion selected from tables, such as those supplied by Falconer and MacKay (1996). When simulating breeding programs on the computer, many computer languages supply a routine that returns the truncation point, \( x \), corresponding to a particular proportion selected, \( p \).

Realized selection intensity in small populations will be less than predicted by \( i = z/p \) as a result of order statistics (Hill 1976). Special tables are provided in Falconer and MacKay (1996) for specific population sizes. Analytically, intensities for finite population size can be approximated by adjusting \( p \) to \( p^* \) as follows:

\[
  p^* = \frac{(s + 1/2z)}{n + s/2n}
\]

where \( s \) is the number selected and \( n \) is the population size (i.e. uncorrected \( p = s/n \)), and then estimating the adjusted \( i \), \( i^* \) as

\[
  i^* = \frac{z^*}{p^*}
\]

where \( z^* \) is the height of the normal distribution at the truncation point \( x^* \) corresponding to \( p^* \).
Derivation of selection intensity from the standard Normal distribution

\[ i = \frac{z}{p} \]

Effect of % Selected on Selection Intensity
Large population size

Effect of Population Size
(Falconer and MacKay, 1996)
The second assumption that is made in the standard equation for selection intensity (3.28) is that there is no correlation between the selection criterion (EBV) of the different candidates of selection. Correlations between the selection criterion of different candidates are generally due to: 1) genetic relationships between candidates of selection; and 2) the use of the same information in calculating the EBV for different animals.

The most extreme example of such a correlation occurs when the population consist of \( n_{fs} \) full sib families with \( n_w \) individuals per family and selection based on pedigree information (\( \hat{g}_s = \frac{1}{2} \hat{g}_s + \frac{1}{2} \hat{g}_d \)). Note that the same pedigree information is used for all member of the family and, because this is the only information used, the correlation between their EBV is equal to 1.

The impact of a correlation between the selection criterion of candidates on intensity is related to the impact of population size on intensity. This is easy to see from the above example by noting that the number of alternative values the selection criterion has among all candidates is not \( n = n_{fs} n_w \) but only \( n_{fs} \). Thus, if \( n_c \) individuals are to be selected, selection is of \( n_c / n_w \) families out of \( n_{fs} \), rather than of \( n_c \) individuals out of \( n_{fs} n_w \).

Rawlings (1976) proposed a method of adjusting intensity for correlations between EBV, as well as finite population size based on:

\[
i^* = \sqrt{1 - t_{av}} \ i
\]

where \( t_{av} \) is the average correlation between the selection criterion across all possible pairs of selection candidates. For a population with unrelated full sib families, \( t_{av} \) can be derived based on the correlation of the EBV of full sibs, \( t_{fs} \), and the correlation of the EBV of unrelated individuals (=0), each weighted by the number of full-sib pairs and unrelated pairs that exist in the population (Rawlings, 1976). The result is:

\[
t_{av} = t_{fs} \frac{n_w - 1}{n_w n_{fs} - 1}
\]

The correlation between the selection criterion of full sibs (\( t_{fs} \)) that is required for these computations can be derived based on the information that contributes to the selection criterion of each full sib. Computation of these correlations for more complex selection criteria will be covered in section 6.1, once selection index methods to derive EBV have been developed.

Meuwissen (1991) extended the method of Rawlings (1976) for populations where full sib families are nested within half sib families. This situation is more common in livestock populations and originates from mating each of \( n_{hs} \) sires to \( n_{fs} \) dams and where each dam produces \( n_w \) offspring. The resulting population consists of \( n_{hs} \) half-sib families with \( n_{fs} \) full sib families of \( n_w \) progeny per half-sib family. The selection intensity adjusted for finite population size and correlated EBV can then be approximated as a weighted average of the correlation between EBV of full-sibs (\( t_{fs} \)), the correlation between EBV of half-sibs (\( t_{hs} \)), and the correlation between EBV of unrelated individuals (0). Weighting each correlation by the number of pairs
that have that specific relationship results in the following equation for the average correlation between all possible pairs of individuals:

\[
t_{av} = \frac{t_{fs}(n_w - 1) + t_{hs}n_w(n_f - 1)}{n_nn_nn_hn - 1}
\]  

(3.33)

Meuwissen (1991) compared this approximation with Monte Carlo simulation for a range of correlations and population sizes and found that the approximation worked well when low correlations between EBV were present or when the number of half-sib families was greater than 10. The approximation, however, overestimated the Monte Carlo results by up to 32% for a scheme with high correlations. A modified approximation for situations with high correlations between EBV was suggested by Meuwissen (1991).

Modern sire and dam evaluation methods use all available information for the prediction of breeding values. The use of more family information increases correlations between EBV of family members. In some breeding schemes, selection focuses on young animals because older animals tend to lag behind genetically. However, young animals have little information on individual or on progeny performance. In that case, family information dominates the prediction of EBV and correlations between EBV of relatives are expected to be high. For a correct comparison of schemes, it is therefore important to consider the effect of correlations between EBV, especially when the number of families is limited. In some animal selection experiments or in the nucleus herd of an animal breeding program, the population is often reproduced by rather few families, perhaps as few as 10, or at least half sibs. Even when the total size is larger, breeding may be carried out through the year with selection only among contemporaries at any time, and these may represent few families. In calculating the selection intensity in those cases, the correlation between family members should not be ignored (Hill, 1976).

### 3.7 Modeling Selection Across Multiple Age Groups

In many breeding populations, candidates for selection may come from several distinct groups, each with a different genetic mean and a different variance for the selection criterion. Examples might be: 1) dairy sires of various ages, where older sires have lower average genetic merit but will be more accurately evaluated and hence have higher variance for the selection criterion when their second crop of daughters become available; 2) selection of boars of different ages, where older boars will have lower average genetic merit; 3) selection of cows, where older cows have more lactations and therefore more accurate evaluations.

Genetic means of progeny generations and responses to selection can in these cases be derived by extending the principle obtained before. Considering sires and dams separately, assume that sires can be selected from three age groups, with the relative number of selection candidates in each age group equal to \(w_{s1}, w_{s2}, \text{ and } w_{s3} \) (\(\Sigma w_j = 1\)). Fractions selected from each age group are \(p_{s1}, p_{s2}, \text{ and } p_{s3}\), for a total proportion selected of

\[
P_s = p_{s1}w_{s1} + p_{s2}w_{s2} + p_{s3}w_{s3}
\]  

(3.34)
Let the genetic mean in age group \(i\) be denoted by \(g_{si}\) and the accuracy of the selection criterion by \(r_{si}\). For the moment we will assume the genetic standard deviation is the same in each age group and equal to \(\sigma_g\). This assumption we be relaxed in later chapters.

Then, the genetic mean of selected sires in age group \(i\) is equal to:

\[
\bar{g}_{si}^* = \bar{g}_{si} + S_{si}
\]  

where \(S_{si}\) is the genetic superiority of the selected sires from age group \(i\) over the mean of all males in that age group, and can be predicted as before based on

\[
\hat{S}_{si} = i_{si} r_{si} \sigma_g
\]

where \(i_{si}\) is the intensity that corresponds to a fraction selected \(p_{si}\).

Using a weighted average based on the relative number of sires from each age group, the mean genetic value of selected sires can be computed as:

\[
\bar{g}_s = \frac{1}{P_s} \left\{ p_{s1} w_{s1} \bar{g}_{s1}^* + p_{s2} w_{s2} \bar{g}_{s2}^* + p_{s3} w_{s3} \bar{g}_{s3}^* \right\}
\]

\[
= \frac{1}{P_s} \sum p_{si} w_{si} (\bar{g}_{si} + S_{si})
\]  

(3.37)

Similarly, the mean genetic value of dams can be derived as:

\[
\bar{g}_d = \frac{1}{P_d} \sum p_{di} w_{di} (\bar{g}_{di} + S_{di})
\]

(3.38)

and the average genetic value of the progeny as

\[
E(\bar{g}_o) = \frac{1}{2} \bar{g}_s^* + \frac{1}{2} \bar{g}_d^*
\]

\[
= \frac{1}{2} \frac{1}{P_s} \sum p_{si} w_{si} (\bar{g}_{si} + S_{si}) + \frac{1}{2} \frac{1}{P_d} \sum p_{di} w_{di} (\bar{g}_{di} + S_{di})
\]

(3.39)

These equations allow for recursive prediction of the genetic mean of the population in successive time periods. In Chapter 8, we will formalize these recursive equations in the form of gene flow.

In the previous, the proportions selected from each age group were pre-determined. These proportions may, however, not maximize the average genetic value of the selected parents and, thereby, the genetic value of progeny. Thus, referring to sires, the problem is to determine the proportions to select from each age group such that the average genetic value of the selected group is maximized, but subject to the constraint that the total proportion selected is equal to \(P_s\).

To address this problem, we’ll assume that the selection criterion \(I_i\) for each age group \(i\) is unbiased. This implies that \(E(g_i|I_i) = I_i\) and also that the selection criterion can be compared across age groups. Thus, individuals with the same value \(v\) of the selection criterion in different age groups are expected to have the same genetic value \(v\).
The general problem is illustrated in Figure 3.2. Given the assumptions for the selection criterion, individuals should be selected by truncating across the distributions of the selection criterion; replacing an individual in age group 1 that falls just above the truncation point with an individual from age group 2 that falls just below the truncation point will reduce the expected genetic value of selected parents. Thus, the same truncation point should be used for all distributions. In practice, this would be equivalent to ranking all individuals based on their EBV regardless of the age group they belong to, and selecting the top ones.

Thus, to maximize the genetic value of selected parents, the objective is to find the truncation point $T$ where selection of sires across all available distributions yields a total proportion selected of $P$. There is no algebraic solution to this problem and the answer must be found iteratively. Bisection is a general, simple, and effective optimization method that can be used for this problem. A schematic of a simple computer subroutine to do this is illustrated below.

1. Find for all $i$ the (unstandardized) truncation point, $T_i$, of the $i^{th}$ distribution that corresponds to a proportion $P$ selected from that distribution ($T_i = \bar{g}_i + x_i \sigma_i$, where $x_i$ is the standardized truncation point and $\sigma_i$ the standard deviation of the $i^{th}$ distribution ($\sigma_i = r_{yi} \sigma_y$ for our case))
2. Choose the lowest \( T_1 \) as a lower bound for \( T \rightarrow T_1 \)
   Choose the highest \( T_i \) as a upper bound for \( T \rightarrow T_u \). (\( T \) must lie between \( T_1 \) and \( T_u \)).

3. Compute the mean of the upper and lower bound \( \Rightarrow T_m = \frac{1}{2} (T_u + T_l) \)

4. For each distribution \( i \), find the proportion selected, \( p_i \), that corresponds to truncation at \( T_m \).

5. Find the total proportion selected for truncation at \( T_m \): \( P_m = \sum p_i w_i \)

6. If \( |P_m - P| < \varepsilon \), where \( \varepsilon \) is a pre-set convergence criterion, exit the routine and return \( T_m \) as the optimized truncation point.

7. If \( P_m < P \) then \( T_m \) becomes the new upper bound \( \Rightarrow \) set \( T_u = T_m \)
   If \( P_m > P \) then \( T_m \) becomes the new lower bound \( \Rightarrow \) set \( T_l = T_m \)

8. Return to step 3.

Even with a large number of distributions, this program will iterate to a solution with high accuracy fairly rapidly. For most applications no more than 5 or 6 rounds of iteration should be required.

The proportion of animals in each distribution, \( w_i \), might reflect structural differences in numbers (different numbers produced in different groups as designed in the breeding program) and losses from groups over time due to death, disease, sales, etc. Differences between groups in reproductive capacity (fertility) could be incorporated directly into \( w_i \), or treated as a separate factor affecting the effective numbers (in terms of contributions to progeny) in each group after selection.

### 3.8 Asymptotic Response per Unit Time

Response defined by equations (3.22) and (3.24) is the response from one generation to the next. If conditions remain constant over generations, it is also the response per generation. **Generation interval** is generally defined as the average age of the parents when their progeny are born or as the average time between birth of parents and birth of progeny.

Generation intervals vary widely across species. For example, a generation interval for poultry and swine can be as short as 1 year, whereas for progeny testing schemes in cattle, generation intervals for sires are often 7 years or more. Generation intervals can also be altered within species by changing the age at which animals are selected and bred.

In general, it is more useful to estimate response per unit time, usually response per year. Response per year is often given the same notation as response per generation, \( R \).
When selection is equal in males and females and, therefore, response per generation is equal to 

\( R = S = ir_g \sigma_g \), response per year is obtained by dividing equation (3.22) by the generation interval, \( L \), to get

\[
R = \frac{ir_g \sigma_g}{L}
\]  

(Note, in general, as here, we must be careful to know whether response, \( R \), is expressed per generation, per year, or in some other unit of time).

Equation (3.40) holds the key to designing breeding programs. Response per unit of time is proportional to the intensity of selection, the accuracy of genetic evaluation, and the square root of the genetic variance, and is inversely proportional to the generation interval.

3.8.1 Multiple Pathways of Selection

The derivations leading to equation (3.40) assumed that males and females are treated alike. In practice this is often not the case. For example, in most species, males have a higher reproductive rate than females, thus we need fewer males for breeding and consequently can have a higher intensity of selection in males than females. In some species, traits of interest are recorded only in one sex, obvious examples being milk yield in dairy cattle, litter size in swine, and rate of egg production in poultry. This can lead to different accuracies of evaluation in the two sexes, since one sex has it’s own performance contributing to it’s evaluation while in the other sex genetic evaluation must be based entirely on information from relatives. Similarly, different sexes can have different generation intervals for a variety of reasons, e.g. the sex with the highest reproductive rate (usually males) may take less time to produce replacement offspring and hence potentially have the shortest generation interval.

In these cases, response per unit of time can be derived by deriving the sum of genetic superiorities in males and females (\( S_s \) and \( S_d \)) by the sum of their generation intervals (\( L_s \) and \( L_d \)):

\[
R = \frac{S_s + S_d}{L_s + L_d}
\]  

This is referred to as the ‘steady state’ or ‘asymptotic’ response to selection, which is the expected response per unit of time after the breeding program has been in operation for several years. The reason for this assumption will be made clear in the derivation of the equation, which follows.

In practice it may take several generations to approach this steady state, and in some cases a true steady state may never be reached. It is therefore generally safer to think of \( R \) predicted by equation (3.41) as the prediction of the average rate of response per year, recognizing that predicted response may well vary from one year to the next. Even where a steady state response rate is eventually achieved, genetic response will usually be variable from one year to the next in the early generations of the breeding program.
Note that responses from year to year can always be predicted from the recursive equation (3.23). A comparison of this approach with the asymptotic response is given in Figure 3.3. Note that, starting from an unselected population, expected responses fluctuate during the initial years but stabilize to the asymptotic response after several years of selection.

**Figure 3.3** Example of predicted annual versus asymptotic responses
Starting from an unselected population

**Figure 3.4** Asymptotic response to selection for breeding programs with overlapping generations
To derive equation (3.41), we start by describing the genetic mean of progeny in terms of the average of the genetic mean of the selected parents, from equation (3.23):

$$
\overline{g}_o = \frac{1}{2} \overline{g}_s^* + \frac{1}{2} \overline{g}_d^* = \frac{1}{2}(\overline{g}_s + S_s) + \frac{1}{2}(\overline{g}_d + S_d) 
$$

(3.42)

Now, referring to Figure 3.4, note that if the asymptotic response of $R$ per year has been achieved, the genetic mean of male selection candidates is expected to be $L_s R$ lower than the genetic mean of the progeny generation. This is because males are on average $L_s$ years older than their progeny and the gain per year is equal to $R$. Thus, the genetic mean of male candidates can be expressed as:

$$
\overline{g}_s = \overline{g}_o - L_s R
$$

and similarly,

$$
\overline{g}_d = \overline{g}_o - L_d R
$$

Substituting into equation (3.42) we get:

$$
\overline{g}_o = \frac{1}{2} \overline{g}_o - L_s R + S_s + \frac{1}{2} \overline{g}_o - L_d R + S_d
$$

Rearranging and solving for $R$ results in equation (3.41).

Equation (3.41) applies to a so-called two-path selection program, in which selection differs between males and females.

### 2 Pathway Program

**Example**

Selection of sheep for weaning weight (WW)

- **Sires**: top 5% at 9 months, selected based on own WW record
- **Dams**: top 60% at 9 months

\[ \sigma_{WW} = 1.97 \text{ kg} \]

Rendel and Robertson (1950) and Robertson and Rendel (1950) pointed out that in any breeding program there are actually four basic pathways of genetic improvement, corresponding to the four sources of parental genes of male and female progeny. These four pathways are:

- male parents of male progeny (sires of males, $sm$)
- female parents of male progeny (dams of males, $dm$)
- male parents of female progeny (sires of females, $sf$)
- female parents of female progeny (dams of females, $df$).

\[ \Delta G_{WW} = 2.92/2.34 = 1.25 \text{ kg/yr} \]
Robertson and Rendel showed that where each of the four pathways of genetic improvement were separately recognized, response per generation as predicted by equation (3.41) can be rewritten as:

$$ R = \frac{S_{sm} + S_{dm} + S_{sf} + S_{df}}{L_{sm} + L_{dm} + L_{sf} + L_{df}} = \sum_i S_i \sum_i L_i $$

(3.42)

For each path, genetic superiorities can be derived as shown before as:  

$$ S_i = i r_i \sigma_i $$

When for a particular path selection is across multiple age groups, genetic superiority for that path can be computed as a weighted average of genetic superiorities achieved within each age group. To illustrate, referring to the example of selection across three age groups of section 3.6, the superiority of that path would be computed as:

$$ S_s = \frac{1}{P_s} \{ p_{s1} w_{s1} S_{s1} + p_{s2} w_{s2} S_{s2} + p_{s3} w_{s3} S_{s3} \} $$

(3.43)

Similarly, the generation interval for this path would be computed as:

$$ L_s = \frac{1}{P_s} \{ p_{s1} w_{s1} L_{s1} + p_{s2} w_{s2} L_{s2} + p_{s3} w_{s3} L_{s3} \} $$

(3.44)

### Selection Across Age Groups

<table>
<thead>
<tr>
<th>Age Group</th>
<th>Age at Birth of Progeny</th>
<th>% of Bull Dams Selected</th>
<th>Genetic Superiority</th>
<th>Genetic Super.</th>
<th>Pooled Generation Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heifers</td>
<td>2 yr</td>
<td>50%</td>
<td>2.34</td>
<td>707.9</td>
<td></td>
</tr>
<tr>
<td>1st Lact.</td>
<td>3 yr</td>
<td>30%</td>
<td>2.53</td>
<td>946.2</td>
<td></td>
</tr>
<tr>
<td>2nd Lact.</td>
<td>4 yr</td>
<td>20%</td>
<td>2.53</td>
<td>1001.9</td>
<td></td>
</tr>
</tbody>
</table>

To illustrate a breeding program in which all four pathways of improvement are recognized, we can consider a conventional progeny testing program for improvement of milk production in dairy cattle with the use of artificial insemination. For simplicity we assume all cows reproduce naturally without the aid of embryo transfer. In such a scheme, young bulls are tested by mating to a (hopefully) random sample of cows, the resulting heifers are reared, and their first lactation performance is recorded. This daughter lactation information is then used to produce a genetic evaluation on each young bull, often referred to as the "first proof" of a bull. At this stage the best bulls can be selected for breeding and the remainder discarded. In contrast, heifers and cows are evaluated largely based on their own lactation performance. In a population of several hundred thousand recorded dairy cows, several hundred young bulls, perhaps up to a thousand, would be tested each generation.
We can now consider each of the four pathways of genetic improvement in a highly efficient hypothetical progeny-testing program.

**Sires of males:** Since we only test a few hundred young bulls, and every sire can produce tens of thousands of doses of semen, we need only a few sires to produce these young bulls each generation. Thus we need to select only the top 1 or 2% of tested bulls as sires of sons. These sires have high accuracy of genetic evaluation, since progeny tests generally give high accuracy. The generation interval will, however, be at least 6 years because of the time from birth of the young bull, through the birth of his first crop of test daughters, through their first lactation to the birth of his sons.

**Sires of females:** Since there are several hundred thousand cows to be bred, many more bulls are required to produce the necessary amount of semen each generation. In an efficient scheme, the top 10-15% of young bulls can be selected, giving a lower selection intensity than for sires of sons. Accuracy of selection is the same as for sires of sons because they are chosen on the basis of the same information. The generation interval is, however, about a year longer because it takes time to breed a large population of cows and the better bulls will be used by farmers for a little longer than the not so good bulls.

**Dams of males:** Since there are several hundred thousand cows and only a few hundred sons are tested, dams of sons can be selected very intensely, perhaps only the best 0.1 to 0.5% being required. But evaluation is based on their own performance, which has lower accuracy than a progeny test. These cows could be bred in their second lactation based on their first lactation performance and part of their second lactation performance, so that they would be around 4½ to 5 years old at the birth of their sons.

**Dams of females:** Dairy cows have a very low reproductive rate, producing less than one live calf per year, after allowing for average calving intervals and mortality of fetuses and calves. Allowing for disease and other losses of growing heifers and for the fact that only half the calves are females, only about 1 in 3 calvings results in a potential replacement heifer for the dairy herd. Since average life in the herd in many western countries is often not much over three lactations, the average cow barely has sufficient time to produce a
replacement before she leaves the herd. There is thus very little room for selection of
dams of cows, with perhaps 90% of all cows required for breeding. Accuracy of
selection would be very similar to that for dams of sires. However, generation interval is
generally increased by a year or two, since the average cow takes close to three calving to
produce a replacement.

The parameters applying to each pathway are summarized in Table 3.1.

Table 3.1. Intensity and accuracy of selection and generation interval in a highly efficient
hypothetical progeny-testing program for improving milk yield in dairy cattle.

<table>
<thead>
<tr>
<th>Pathway</th>
<th>Proportion Selected (p_i)</th>
<th>Intensity (i_i)</th>
<th>Accuracy (r_i)</th>
<th>Genetic Superiority (S_i = i_r_iσ_g)</th>
<th>Generation Interval (yr) (L_i)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sires of males</td>
<td>2 %</td>
<td>2.42</td>
<td>0.90</td>
<td>2.178 σ_g</td>
<td>6</td>
</tr>
<tr>
<td>Sires of females</td>
<td>10 %</td>
<td>1.75</td>
<td>0.90</td>
<td>1.575 σ_g</td>
<td>7</td>
</tr>
<tr>
<td>Dams of males</td>
<td>0.5 %</td>
<td>2.89</td>
<td>0.60</td>
<td>1.743 σ_g</td>
<td>5</td>
</tr>
<tr>
<td>Dams of females</td>
<td>90 %</td>
<td>0.19</td>
<td>0.60</td>
<td>0.114 σ_g</td>
<td>6</td>
</tr>
<tr>
<td>TOTAL</td>
<td></td>
<td></td>
<td></td>
<td>ΣS = 5.601σ_g</td>
<td>ΣL = 24</td>
</tr>
</tbody>
</table>

If we assume that genetic variance is the same for all pathways (a common assumption but not
always strictly true; see Chapter 5), then we can use the parameter values in Table 3.1 to obtain
an estimated annual rate of response for this particular breeding program, of

\[ R = \frac{5.601}{24} \sigma_g = 0.233 \sigma_g \text{ per yr} \]

Response could of course be expressed in many units, but the three most common and probably
most useful are in genetic standard deviations, \( \sigma_g \), per year (as above), absolute units per year
(e.g. kg milk per year), or as a percentage of the mean per year.

Imagine that the dairy cattle population above has a mean yield of 6000 kg, that the heritability
\( (h^2) \) of milk yield is 0.25, and that coefficient of variation (CV) is 0.18, all fairly typical values
for intensive dairy production. Since

\[ \sigma_g^2 = h^2 \sigma_p^2 \]

and

\[ \sigma_p^2 = (cv \times \bar{x})^2, \]

then

\[ \sigma_p^2 = (0.18 \times 6000)^2 = (1080)^2. \]

Hence

\[ \sigma_g^2 = 0.25(1080)^2 \]

And

\[ \sigma_g = \sqrt{\sigma_g^2} = 0.5 \times 1080 = 540 \text{ kg.} \]

Hence

\[ R = 0.233 \times 540 = 125.82 \text{ kg per year} \]

or, alternatively,

\[ R = 125.82/6000 = 2.1\% \text{ per year.} \]
The choice of units will depend on how the results are to be used. Use of genetic standard deviation units may be useful to geneticists who think in such terms and allow results to be readily converted from one population to the next if it is believed that the major variation between populations is in the absolute amount of genetic variance. For example, this would be true if \( h^2 \) and \( cv \) were the same for different populations but the mean level of performance differed.

Absolute units, such as kg milk per year, are often the most intelligible to people familiar with the species and trait(s) in question. For example, there would probably be little point in presenting results in \( \sigma_x \) per year if the audience is made up of non-geneticists, such as dairy farmers, industry, or government officials.

Expressing results in terms of percentage change per year is likely to be understood by a wide audience. It also has the advantage of allowing relatively meaningful comparisons of response for different traits across species. A good example is given by Smith (1984), who compared the theoretical response rate for typical breeding programs for sex-limited traits in poultry, swine, sheep, and cattle. The traits were egg production in poultry, litter size in swine, litter size in sheep, and milk production in cattle. His estimates of absolute response rates were 5.46 eggs per year, 0.3 piglets per year, 0.04 lambs per year, and 75 kg milk per year. Expressed in absolute units, it is clearly very difficult to interpret these results or make any comparison across species. However, expressed as percentage change per year, the same results were 2.1, 3.0, 2.1, and 1.5\% per year for poultry, swine, sheep, and dairy cattle. Although not perfect, this does allow us to draw such general conclusions, as that selection for sex-limited traits should give roughly similar relative rates of response in different species. It may come as a surprise to those working with dairy cattle, that the relative rates of response are lowest for milk production in cattle.

**Accounting for use of young bulls**

In the previous, the sire to female path only accounted for the use of progeny-tested sires to breed cows to produce herd replacements. However, young bulls also contribute to the next generation of females; in a practical breeding program, semen from young bulls can represent as much as 20\% of all inseminations. To account for this, the genetic superiority and generation interval for sires of females must be computed as a weighted average. Assuming \( y \) is the proportion of females produced from young bulls, genetic superiority of the sire to female path is computed as:

\[
S_{sf} = y S_{yb,f} + (1-y) S_{pb,f}
\]

where \( S_{yb,f} \) and \( S_{pb,f} \) are genetic superiorities of young and progeny-tested bulls that are used to breed female replacements. In most cases, \( S_{yb,f} = 0 \) because \( P_{yb,f} = 1 \) and thus \( i_{yb,f} = 0 \), unless there is additional selection of young bulls that are entered into the progeny tests, above and beyond selection of their parents (which is already covered through the \( sm \) and \( dm \) pathways). An example where \( S_{yb,f} > 0 \) is preselection of young bulls based on genetic markers (see Chapter 12).

Similarly, the generation interval for the \( sf \) pathway is computed as a weighted average of the generation intervals for the \( yb,f \) and \( pb,f \) pathways:

\[
L_{sf} = y L_{yb,f} + (1-y) L_{pb,f}
\]
Predicting Genetic Progress
Accounting for use of young bulls

An example is given in Table 3.2, which assumes \( y = 0.2 \)

**Table 3.2.** Intensity and accuracy of selection and generation interval in a highly efficient hypothetical progeny-testing program for improving milk yield in dairy cattle with accounting for 20% use of young bulls to breed female replacements.

<table>
<thead>
<tr>
<th>Pathway</th>
<th>Proportion Selected ((p_i))</th>
<th>Intensity ((i_i))</th>
<th>Accuracy ((r_i))</th>
<th>Genetic Superiority ((S_i = i_i r_i \sigma_g))</th>
<th>Generation Interval (yr) ((L_i))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sires of males</td>
<td>2 %</td>
<td>2.42</td>
<td>0.90</td>
<td>2.178 (\sigma_g)</td>
<td>6</td>
</tr>
<tr>
<td>Sires of females</td>
<td>- Young</td>
<td>100 %</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>- Proven</td>
<td>10 %</td>
<td>1.75</td>
<td>0.50</td>
<td>1.260 (\sigma_g)</td>
</tr>
<tr>
<td>Dams of males</td>
<td>0.5 %</td>
<td>2.89</td>
<td>0.60</td>
<td>1.734 (\sigma_g)</td>
<td>5</td>
</tr>
<tr>
<td>Dams of females</td>
<td>90 %</td>
<td>0.19</td>
<td>0.60</td>
<td>0.114 (\sigma_g)</td>
<td>6</td>
</tr>
<tr>
<td>TOTAL</td>
<td></td>
<td></td>
<td></td>
<td>(\Sigma S = 5.268 \sigma_g)</td>
<td>(\Sigma L = 23)</td>
</tr>
</tbody>
</table>

Now response per year becomes: \[ R = \frac{5.268}{23} \sigma_g = 0.230 \sigma_g \text{ per yr} \]

Note that, compared to Table 3.1, response is slightly lower. By changing \( y \), this approach can be used to optimize the proportion of the population to inseminate with young bulls. Note, however, that increasing \( y \) also increases the number of young bulls that can be tested or, alternatively, the number of progeny per young bulls. This has consequences for other parameters of the breeding program. Nevertheless, this method provides a means to look at the impact of various factors on genetic gain. A spreadsheet to evaluate alternative program parameters is provided.
3.9 Correlated Response to Selection

Selection for trait $i$ will not only result in genetic change in trait $i$ ($R_i$) but also in traits that are genetically correlated to the selected traits. Genetic change in trait $j$ to selection on trait $i$ is referred to as correlated response to selection and will be denoted $R_{ji}$, in contrast to direct response, which is denoted by $R_i$. Similarly, genetic superiorities of parents selected on trait $i$ will be denoted by $S_i$ and superiorities for trait $j$ by $S_{ji}$.

Following equation (3.22), genetic superiority of parents for trait 2 as a result of selection on an index for trait 1, $I_1$, can be obtained based on the general equation:

$$S_{2.1} = i r_{g2.1} \sigma_g$$  \hspace{1cm} (3.45)

Here $r_{g2.1}$ is the correlation of the genetic value for trait 2 with the criterion that selection is based on, i.e. $I_1$. When the selection criterion $I_1$ is only based on records for trait 1 (single trait
evaluation), this correlation can be expressed in terms of the accuracy of selection for trait 1 and
the genetic correlation as: \( r_{g_{2|1}} = r_{g_{2|1}}r_{g_{1|1}} \)

Then:

\[
S_{2.1} = i r_{g_{2|1}} r_{g_{1|1}} \sigma_x^2 = r_{g_{2|1}} \frac{\sigma_y}{\sigma_x} i r_{g_{1|1}} \sigma_x = r_{g_{2|1}} \frac{\sigma_y}{\sigma_x} S_1 = b_{g_{2|1}} \text{ } S_1 \tag{3.46}
\]

where \( b_{g_{2|1}} \) is the regression of genetic values for trait 2 on genetic values for trait 1. This
regression coefficient quantifies the expected genetic change in trait 2 for every unit genetic
change in trait 1. When the selection criterion is not exclusively based on records for trait 1, e.g.
the index is a multiple-trait index, the same principle holds but derivation of the regression
coefficient becomes more complex. This will be dealt with in Chapter 4.

Correlated response to selection can now be predicted from direct response by simple regression
techniques:

\[
R_{2.1} = b_{g_{2|1}} R_1 = r_{g_{2|1}} \frac{\sigma_y}{\sigma_x} R_1 \tag{3.47}
\]

Where \( R_1 \) can be predicted using equation (3.41).
Example

Selection of sheep on weaning weight

- Direct response: \( \Delta G_{WW} = 1.25 \text{ kg/yr} \)
- Correlated response in birth weight:
  \[
  \sigma_{g_{WW}} = .5 \text{ kg} \\
  r_{g_{WW}g_{BW}} = .3 \\
  \Delta G_{BW,WW} = b_{g_{WW}g_{BW}} \Delta G_{WW}
  \]

Predicting Response in WW

<table>
<thead>
<tr>
<th>Path</th>
<th>%</th>
<th>i</th>
<th>( r )</th>
<th>( \text{Genetic Superiority} )</th>
<th>Gen. Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sire</td>
<td>5</td>
<td>2.06</td>
<td>.55</td>
<td>2.23</td>
<td>1.17 yr</td>
</tr>
<tr>
<td>Dam</td>
<td>60</td>
<td>.64</td>
<td>.55</td>
<td>.69</td>
<td>1.17 yr</td>
</tr>
</tbody>
</table>

\[
\Delta G_{WW} = \frac{2.92}{2.34} = 1.25 \text{ kg/yr}
\]

Prediction of Correlated Response

\[
b_{g_{WW}g_{BW}} = \frac{r_{g_{WW}g_{BW}} \sigma_{g_{WW}}}{\sigma_{g_{BW}}} = (0.3) \frac{5}{197} = .076 \text{ kg/kg}
\]

\[
\Delta G_{BW,WW} = (0.076)(1.25) = .095 \text{ kg/yr}
\]

Indirect Selection

- Genetic improvement of a trait of economic importance through selection on EBV for a correlated trait, e.g.
  - Select on Somatic Cell Count to improve mastitis resistance.
  - Select on conformation traits to improve herd life.
  - Select on scrotal circumference to improve fertility (sheep).

Indirect Selection (cont’d)

- Advocated over direct selection if:
  - Correlated trait is recorded and direct trait not.
  - Correlated trait is less expensive to measure.
  - Correlated trait is measured earlier in life
  - Correlated trait has higher \( h^2 \).

Efficiency of Indirect vs Direct Selection

- \( 1 = \text{correlated trait} \)
- \( 2 = \text{economic trait} \)
- Direct selection:
  \[
  \Delta G_2 = \frac{\Sigma r_{g_i} \sigma_{g_i}}{\Sigma L}
  \]

Efficiency of Indirect vs Direct Selection (cont’d)

- Indirect selection: (correlated response in trait 2 to selection on trait 1)
  \[
  \Delta G_{2,1} = \frac{r_{g_{i1}} \sigma_{g_{i1}}}{\sigma_{g_{i1}}} (\Delta G_1) \\
  \uparrow \\
  \Delta G_1 = \frac{ir_{g_{i1}} \sigma_{g_{i1}}}{\Sigma L}
  \]

Efficiency = \[
\frac{\Delta G_{2,1}}{\Delta G_2}
\]
3.10 Design of Breeding Programs

The prediction of rate of response to selection given by equation (3.40) and in its more complete form by equation (3.42) holds the key to understanding many of the basic principles of design of breeding programs. In general, response is positively related to intensity and accuracy of selection and to amount of genetic variation, and is negatively related to generation interval. Altering a breeding program will often affect several parameters simultaneously and it is the net effect of all these changes that determines the predicted response to selection.

Consider the dairy cattle progeny testing scheme outlined in section 3.8.1. We could, for example, ask the consequence of waiting until potential dams of sires were older and thus had more lactation records than in the scheme originally outlined. This would increase accuracy of evaluation in this pathway somewhat, because of the increase in information available, but would also increase the generation interval. Later in this course you will have the tools to predict the expected change in accuracy, but at this stage we will simply state that by waiting for an extra year, the accuracy of evaluation in the dams of sires pathway would increase from 0.6 to 0.64 while the generation interval increases from 5 to 6 years. Thus the predicted rate of response is now

\[ R = \frac{2.42 \times 0.9 + 1.75 \times 0.9 + 2.89 \times 0.64 + 0.19 \times 0.6}{6 + 7 + 6 + 6} \sigma_g = 0.229 \sigma_g \text{ per year} \]

which is less than the predicted response of 0.233 \( \sigma_g \) per year when selecting younger dams of sires. Assuming our parameters are appropriate, we would conclude that we should not wait for extra lactation records on our potential dams of sires.

As another example, we could go on to ask what would happen if we tested more young bulls in our progeny test program each generation. If testing resources were limited by having more young bulls to test, we would have to produce fewer daughters per bull. Thus accuracy of selection would decrease (due to having fewer daughters) and intensity of selection would increase (due to having more young bulls to choose among) in both sire pathways. But also, if we had more young bulls tested, we would need more dams to produce these bulls, which would increase the proportion selected and reduce intensity of selection in the dams of sons pathway. In such a situation we could vary the number of young bulls tested per generation, calculating the appropriate selection intensities and accuracies in each pathway and hence derive the expected rate of response to selection for each number tested. The number of bulls tested that maximized response rate could then be identified.

As we will see later in this course, the above approach is only an approximation to the real world. But in many cases this approximation can be quite reliable in its own right. Adapting this approximation to more complex (realistic?) situations is not necessarily particularly difficult.

Another consideration is that the design that maximizes genetic response is not necessarily the design that maximizes economic progress. To evaluate the optimum design from an economic perspective requires that the economic costs be weighed against the economic benefits of the designs considered. In some cases a wide range of designs can give similar rates of genetic progress, but often at widely differing costs. In such cases the economically optimum design may give slightly less than maximum genetic response.