DESIGN AND OPTIMISATION OF ANIMAL BREEDING PROGRAMMES

Lecture notes prepared by:

Jack C.M. Dekkers
Iowa State University
Ames, USA

John P Gibson
Institute for Genetics and Bioinformatics
Armidale, Australia.

Piter Bijma and Johan A.M. van Arendonk
Animal Breeding and Genetics group
Wageningen University
Wageningen, The Netherlands

for AnS 652 A and B, S05, Iowa State University
by Jack Dekkers

Preface

The first version of the notes were written by John Gibson for a Nordic graduate course in Denmark in 1992. When Johan was asked to give a similar course in Denmark four years later, it was decided to combine the material into a joint set of lecture notes. From that point onwards, the (parts of the) notes have been used in several other courses. In recent years, Jack Dekkers has made a considerable contribution to the notes. Over time, the notes have changed which reflects the developments in the fields as well as feed back we have had in teaching animal breeding and genetics.

For the Chapter on Inbreeding, we have made use of material prepared by John Woolliams and Theo Meuwissen for an international course in 2000 in The Netherlands. We greatfully acknowledge their contribution.

We would like to thank the students attending courses in which these or an earlier version of the notes were used for making teaching an enjoyable experience and for all their comments and suggestions towards improving these notes.

Wageningen, July 2004

Jack Dekkers
Department of Animal Science
Iowa State University
239D Kildee Hall
Ames IA 50011-3150 USA
e-mail: jdekkers@iastate.edu

John Gibson
Institute for Genetics and Bioinformatics
University of New England
Armidale NSW 2351, Australia.
e-mail: john.gibson@pobox.une.edu.au

Johan van Arendonk and Piter Bijma
Animal Breeding and Genetics Group
Wageningen University
PO Box 338
6700 AH Wageningen, The Netherlands
e-mail: johan.vanarendonk@wur.nl and Piter.Bijma@wur.nl

Chapter 1

Introduction

(Based on Bijma and van Arendonk, 2004)

There are two fundamental questions faced by animal breeders. The first asks: **"What is the best animal?"** Is the best Labrador the one with show-winning conformation or the one with exceptional retrieving instinct? Is the best dairy cow the one that gives the most milk; the one with the best feet, legs and udder support; or the one that combines performance in these traits in some optimal way? These are matters of intense debate among breeders, and, in truth, no one has all the answers. The question is an important one, however, because the answers determine the desired direction of genetic change for breeding organisations and people keeping farm or companion animals. The second question asks, "**How do you breed animals so that their descendants will be, if not "best", at least better than today's animals?"**. In other words, how can we genetically improve animal populations? This question involves genetic principles and animal breeding technology, and is the subject of this course.

1. What is the best animal

"Best" is a relative term. There is no best animal for all situations. The kind of animal that works best in one environment may be quite different from the best animal under another set of circumstances.

When we describe animals, we usually characterise them either in terms of appearance or performance or some combination of both. In any case, we talk about **traits**. A trait is any observable or measurable characteristic of an animal.

Some examples of *observable* traits –traits we would normally mention in describing the appearance of an animal- are coat colour, size, muscling, leg set, udder conformation, and so on. Some examples of *measurable* traits –traits we would likely refer to in describing how an animal has performed- are body weight, daily milk production, time to run a mile, etc. There are hundreds of traits of interest in domesticated animals. Note that in none of the examples of traits mentioned above is the appearance or performance of a particular animal described. An animal may be red and weigh 343 kilograms at 1 years of age, but *red* coat colour and *343 kg* yearling weight are not the traits- the traits are simply coat colour and weaning weight. *Red* and *343 kg* are the observed categories or measured levels of performance for the traits of coat colour and yearling weight. They are the **phenotypes** for these traits.

In animal breeding, we are mainly concerned with changing animal populations genetically. From a genetic point of view, therefore, we want to know not only the most desirable phenotypes, but the most desirable **genotypes** as well. That is because an animal's genotype provides the genetic background for its phenotypes and it is the genetic material that is passed on from parents to its offspring. Summarised in an equation:

$$\mathbf{P = G + E}$$

where P represents an individual's phenotype, G represents its genotype, and E represents the **environmental effects**- the effects that external (nongenetic) factors have on an animal's performance[1]. In other words, its genotype and the environment it experiences determine an animal's phenotype.

The word *genotype* is used in several ways. We can speak of an animal's genotype in general, referring to all the genes and gene combinations that affect the array of traits of interest to us. An example used later on in this section involves a "tropically adapted" genotype. In this case, the genotype includes all the genes and gene combinations affecting heat resistance, parasite resistance, and other traits that make up tropical adaptation. This sense of the word *genotype* is generally implied in this chapter. We can also speak of an animal's genotype for a particular trait, referring to just those genes and gene combinations that affect that trait (e.g., heat resistance). Or, as we will see later in this course, we can limit the definition of genotype even further in which case it refers to a particular gene only (e.g., an animal has genotype AA for the kappa-casein gene). In any case, the genotypes of our animals' descendants are what we can change with breeding methods. Favourable changes in genotypes result in improved phenotypes.

To answer the question "What is the best animal?" we need to determine what traits are of primary importance and what genotypes are most desirable for those traits. Most breeders, if they have some experience, have an opinion about the key traits and better genotypes. A Thoroughbred breeder, for example, might describe the perfect animal as "…. fast, but with enough endurance and heart for the longer distances, and easily rated". A pig breeder version might be "…. a healthy pig with a good growth and good carcass quality." There are probably as many opinions of this sort as there are breeders and for the most part they are quite subjective. In order to develop a sense of the important traits and best genotypes in a more objective way it is important to understand the role of the genotype in the system of the farm. This means that the importance of traits will depend on the physical environment under which animals are kept, the management system as well as economic factors. If you think about it, it will become clear that a number of the components of the system will interact with each other. For example, the best preventive health program (management) depends on the kinds of pathogens in the area (physical environment) and the costs of vaccines, dewormers, etc. (economics). To determine which health program is the most cost-effective, you must have knowledge of alternative programs, local pathogens, and treatment costs and understand how treatment programs interact with these other factors to affect profitability. Similarly, the best genotype depends on the local environment, the management practises in use, and the costs of inputs and prices of animal products. To determine the best genotype, you must have knowledge of environmental, management, and economic components and understand how they interact with the genotype to affect profitability.

> The genotype of domestic animals determines the degree to which the animals are suited for their function in society. The key to determining the traits of importance and optimal genotypes for those traits is a thorough analysis of the function of the animal in the entire system and an understanding of the many interactions among components of the system.

Knowledge of the function of the animal and the interactions between the genotype and other components of the system is necessary if we want to develop sensible goals for breeding programs, in other words, if we want to develop appropriate **breeding objectives**. Knowing, for

[1] This mathematical expression is oversimplified but it will do fine for the purposes of this discussion. Later on we will see that there might also be an interaction between the G and E.

example, that parasite resistance is critically important in tropical climates, breeding objectives in the Tropics emphasise traits such as tick count (a measure of tick resistance). In temperate regions, on the other hand, less emphasis is placed on parasite resistance and more emphasis is placed on other traits.

2. Population structure and breeding objective

In the process of determining the best animal, you might ask, "Best for whom?". The answer to this question depends on the function of the animal, the structure of the population and the role of the "breeder"[2] within that structure. Most populations can be thought of as having a pyramidal structure: a relatively small number of breeders at the top selling breeding stock to a larger number of multipliers who in turn sell animals to a great number of end users.

The pyramid suggests a flow of **germ plasm** – genetic material in the form of live animals, semen, or embryos – from the top down, the elite breeders producing the most advanced animals, breeders at the multiplier level replicating those animals, and end users benefiting from the genetic improvement occurring at the higher levels. Ideally, breeders at each level try to produce animals that will be in the greatest demand by their customers at the next level down, with the ultimate result that the best animal is the animal that is the most useful or profitable for the end user. *End users* can thus be defined as the individuals whose particular needs should form the basis for determining breeding objectives.

In food and fibre producing species (sheep, cattle, swine, and poultry), the end users are commercial producers. These are the persons whose primary products are commodities for public consumption. Commercial dairy farmers produce milk; commercial swine producers produce pork; commercial poultry farmers produce eggs, chicken and turkey. Commercial producers are in most cases not the end of the production chain; beyond them are the processors (dairy plant, slaughterhouses), the retailers and consumers. But the commercial producers are end users because their particular needs reflect the requirements of the entire production chain. They need animals that are physically and reproductively sound, healthy and perform efficiently in their environment. They also need animals that possess the product and performance characteristics required by the retailers and consumers. The importance of these latter characteristics should be reflected – when the market systems functions well - in the prices paid to the commercial producers for their products. In the Western world, the interest of consumers in the system of production has increased over time. This increased awareness of consumers has resulted in an increased emphasis on health and welfare traits in the breeding objective of farm animals and reduced emphasis on primary production traits (e.g. amount of milk, growth rate and litter size).

The breeding industries for recreational and companion animal species (horses, dogs, cats, etc.) differ somewhat in structure from the livestock industries. The pyramid arrangement is still present, and markets for specialised types of animals exist, but seedstock/commercial divisions are usually less clear and the end users may not be breeders at all. Consider, for example, Labrador retrievers. The end users of Labs are hunters and pet owners. These persons may or may not choose to breed their animals, and the qualities that are important to them are those that contribute to retrieving ability, companionship, health, aesthetics, or some combination of these

[2] Person answering the question

traits. Among Labrador breeders there are elite breeders and multipliers, but the term commercial producers does not really fit here because no consumable commodity like meat, eggs or milk is being produced. The various horse industries provide similar examples. End users of horses range from owners of the most valuable racing animals to causal riders to those that keep miniature horses as pets.

3. How are animal populations improved?

The purpose of animal breeding is not to genetically improve individual animals- once an individual is conceived, it is too late to change the genotype of that animal- but to improve animal **populations**, to improve future generations of animals. To this task breeders bring two basic tools: selection and mating. Both involve decision-making. In selection, it is decided which individuals become parents, how many offspring they may produce, and how long they remain in the breeding population. In mating, it is decided which of the males we have selected will be bred to which of the females we have selected.

Selection

Selection is used to make long-term genetic change in animals. It is the process that determines which individuals become parents, how many offspring they may produce, and how long they remain in the breeding population. Most of us are familiar with the term **natural selection**. Natural selection is the great evolutionary force that fuels genetic change in all living organisms. We commonly think of natural selection as affecting wild animals and plants, but in fact it affects both the wild and domestic species. All animals with lethal genetic defects, for example, are naturally selected against- they never live to become parents. Natural selection cannot be ignored but the kind of selection of primary interest in animal breeding is **artificial selection**. The idea behind selection is simply this: to let individuals with the best sets of genes reproduce so that the next generation has, on average, more desirable genes than the current generation of animals. The animals with the best sets of genes are said to have the best **breeding values**. They are –from a genetic point of view- the individuals with the greatest value as parents. In selection, we try to choose those animals with the best breeding values: the animals that will contribute the best genes to the next generation. The result of successful selection is then to genetically improve future generations of a population by increasing over time the proportion of desirable genes.

To see how selection works, consider the simplest form of selection: **phenotypic selection** or **mass selection**. In this type of selection, the performance of the individual is the only information used in making selection decisions. No attention is paid to the pedigree of the animal or the performance of its sibs (brothers and sisters) or of any progeny it may have produced. For example, if you were using phenotypic selection for weaning weight to determine whether a particular ewe lamb was to be kept for breeding, you would base your decision strictly on her own weaning weight. In practise (meaning outside of scientific laboratories), phenotypic selection in its pure form is increasingly rare, but it makes a good example, as we will also see later on during this course.

Figure 1.1 depicts phenotypic selection for increased body size in mice. The largest mice in each generation are chosen to become parents of the next generation, and the result over time is an increase over time in average body size. The idea of using the phenotype for body size as the

selection criteria is based on the expectation that phenotype for size is a reasonable indicator of the genes affecting body size. It is the genes, after all, which are transmitted from parent to offspring. In other words, it is assumed that phenotype for body size in mice is somehow related to breeding value for body size. If that were not the case, phenotypic selection for this trait would be a waste of time. The relationship between phenotype and breeding value is therefore a very important one, and this relationship is reflected by the **heritability**. When heritability of a trait is high, phenotypes are generally good indicators of underlying breeding values, and phenotypic selection will be effective in changing the level of the trait. When heritability is low, phenotypes reveal little about breeding values, and phenotypic selection will be ineffective. Judging by the rapid increase in body size of the mice in Figure 1.1, body size must be quite heritable. Not all traits are as heritable. The heritability of fertility in mammals, for example, is generally quite low. Estimating the heritability of a trait involves statistical techniques to estimate the extent to which relatives resemble each other for the trait of interest, compared with unrelated animals. The actual methodology involved and a description of the methods is beyond the scope of this course.

Most animal breeders are unlikely to limit themselves to individual performance information alone in making selection decisions. They will use information on relatives as well. For example, when a dog breeder purchases an eight-week old puppy from another breeder, she probably does not base here choice on just the conformation and personality characteristics evident in such a young puppy. She wants to evaluate those same traits in the littermates, the dam and the sire. She might want to see a copy of the puppy's extended pedigree to learn more about its ancestors. Similarly, when beef cattle breeders evaluate a sire to use via artificial insemination (A.I.) they look further than the sire's own performance for growth rate. They want to know something about the growth performance of his progeny.

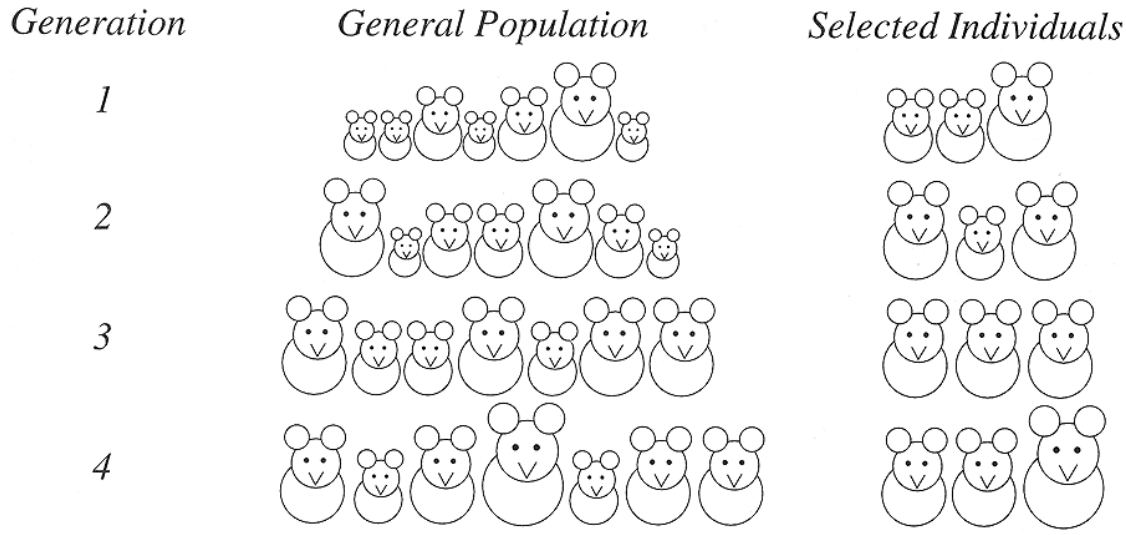


Figure 1.1. Illustration of phenotypic selection for increased body size in mice

The above examples illustrate that selection decisions are based on a combination of information. In this course we will outline how the different sources of information can be combined into a single prediction of the breeding value of the animal. The strength of the relationship between the true breeding value and its prediction is measured by the **accuracy**. When accuracy is high, predictions of breeding values will normally be good ones – they will closely reflect the

differences in true breeding values of the animals being evaluated. And because the predictions of breeding values are accurate, we can do a good job in selection.

The traits mentioned so far in this chapter – such as weaning weight in sheep, body size in mice, fertility, conformation and personality in dogs, milk production in dairy cattle- have all been **polygenic traits**. Many genes affect polygenic traits, and no single gene is thought to have an overriding influence. The genetic variation in these traits is due to segregation at many loci. Until recently, we knew little about the specific genes affecting these traits – we just know there were lots of them. As long as we cannot identify specific genes, we have to rely on phenotypic performances, predictions of breeding value to characterise the genotypes of animals. There are good grounds for believing that there is a range in the size of effects of genes for any trait, from a few with large effect, down to a large number having very small effects. We will see in this course that the developments in molecular biology now make it feasible to identify individual genes that affect quantitative trait. Information on genetic markers linked to individual genes can be used in selection programmes to improve the accuracy of selection (so-called marker-assisted selection). Once an individual gene has been identified, its biochemical and physiological roles can be studied. The results of these studies will greatly increase our understanding of the nature of genetic variation in traits.

Most traits in animals are polygenic in nature. Some traits, however, are **simply-inherited** – they are affected by a single or only a few genes. A good example is the horned/polled character in cattle of European origin (Polled means naturally without horns). A single gene determines whether a cow is horned or polled. There are also a large number of single-gene disorders that are considered to be serious problems but do not prevent affected individuals from reproducing. Well known examples include the inherited eye disorder is dogs, the malignant hyperthermia syndrome ("halothane gene") in pigs. Because only a few genes influence simply-inherited traits, selection for simply-inherited traits is different from selection for typical polygenic traits. With simply-inherited traits, we do not deal with breeding values and their predictions, or even with the concept of heritability. Rather, we are interested only in knowing whether an individual possesses the specific allele or alleles of interest, and we select animals based on that knowledge. If the disorder can be detected either by clinical examination or by DNA-testing prior to reproductive age, it is possible to select against the disorder effectively. The detection of the gene for malignant hyperthermia syndrome in pigs and the subsequent development of a DNA-test have greatly increased the opportunity for pig breeders to eliminate the disorder from the population. Malignant hyperthermia in pigs is an autosomal recessive disorder which means that it is not possible to discriminate between a phenotype of animals with two normal alleles (homozygous animals) and animals carrying one defect allele (heterozygous animals, so-called carriers). The power of the DNA-test lies in the fact that it facilitates the detection of carriers- animals that are heterozygous at the gene causing the genetic disorder- prior to reproductive age.

When we think of selection, we normally envision selection of individual animals within a breed. It is also possible to select between breeds. In setting-up a farm or breeding program, we need to choose a breed to work with. **Between-breed selection** provides a way of using breed differences to make very rapid genetic change. For many traits, breed differences can be very large. By taking advantage of such large differences, between-breed selection can produce genetic change much faster than the gradual change possible from selection within a breed. For example, the milk production of Black and White cattle in The Netherlands has increased enormously in the

1970's – not through selection within the Dutch Friesian population, but through importation of semen from the more productive Holstein-Friesian in the United States and Canada.

Mating

Selection is the first of the two basic tools used by animal breeders to make genetic change. The second tool is **mating**. Mating is the process that determines which (selected) males are bred to which (selected) females. It is distinctly different from selection. In selection, you choose the group of animals you want to be parents; in mating, you match males and females from the selected group.

There are many different methods for mating animals, and each method can be defined by a set of mating rules: a **mating system**. There are three reasons for using mating systems: (1) to produce offspring with extreme breeding value, (2) to make use of complementarity, and (3) to obtain hybrid vigour. Extreme phenotypes can be obtained by mating parents with extreme breeding values (high*high and low*low). If an animal of intermediate size is desired, mating large animals to small animals is one way to produce it. The parental genotypes are quite different, and neither one is optimal, but the mating is complementary because the offspring is optimal. Mating a Charolais to an Angus is an example of **crossbreeding**; the mating of sires of one breed to dams of another. In crossbreeding often used to produce breed complementarity, and in fact, the Charolais x Angus mating is a complementary one. Charolais are large French cattle known for their fast growth and heavy muscling, Angus are smaller British cattle known for their maternal ability, and the crossbred offspring benefit for having both kinds of parents. Another reason for crossing these two breeds is to produce **hybrid vigour** or **heterosis**. Hybrid vigour is an increase in performance of crossbred or hybrid animals over that of the pure-breds. Hybrid vigour occurs to a greater or lesser degree in many traits, but it is most noticeable in traits like fertility and survivability.

4. Multiple trait selection

In this course, a lot of the discussion of selection and the examples used for illustration will be limited to **single-trait selection**, selection for just one trait. That is because single-trait selection provides a simple framework within which to learn the principles of animal breeding. But in the real world of animal breeding, selection for a single trait is rare. Breeders are typically interested in improving a number of traits. They practise **multiple-trait selection.** Dairy farmers select for traits related to milk production, health, reproduction, type and longevity.

Selection for one trait rarely affects just that one trait. Usually other traits are affected as well. Genetic change in a trait resulting from selection on another trait is termed **correlated response to selection**. Correlated response to selection is probably caused by a number of genetic mechanisms and results in so-called genetic correlation between traits.

Genetic correlations between traits and the correlated response to selection brought about by them can be beneficial. However, if we are unaware of or choose to ignore unfavourable genetic correlations, selection for one trait can lead to undesirable response in others. In cattle, for example, blind selection for growth rate leads to larger birth weights and more dystocia. If we

want faster growth, but cannot tolerate increased dystocia, we must avoid simply selecting for growth or against dystocia. We need a way to select for growth rate and against dystocia at the same time. We need a method for multiple-trait selection as introduced in this course.

5. Inbreeding

Inbreeding is the mating of related individuals. That is the simplest definition anyway. Because all animals within a population are related to some degree, a more technically correct definition of inbreeding is the mating of individuals more closely related than average for the population. Inbreeding has a number of effects, but the chief one and the one from which all the others stem is an increase in homozygosity- an increase in the number of homozygous loci in inbred animals and an increase in the frequency of homozygote genotypes in an inbred population. Because inbred individuals have fewer heterozygous loci than non-inbreds, they cannot produce as many different kinds of gametes. The result is fewer different kinds of zygotes and therefore less variation in the offspring. This illustrates, as we will see in more detail furtheron in this course, that inbreeding (more precisely the level of inbreeding in the population) is related to the amount of genetic variation. A second consequence of inbreeding is the expression of deleterious recessive alleles with major effects, and it is this aspect of inbreeding, more than any other, that gives inbreeding a bad reputation. People associate inbreeding with genetic defects. It is true that defects caused by recessive alleles often surface in inbred populations. But inbreeding does not create deleterious recessive alleles; they must already have been present in a population. Inbreeding by itself simply increases homozygosity, and it does so without regard to whether the newly formed homozygous combinations contain dominant or recessive alleles. It therefore increases the chance of deleterious alleles becoming homozygous and expressing themselves. Expression of deleterious recessive alleles with major effects, particularly lethal genes, is a very visible consequence of inbreeding. It is an example of the effect of inbreeding can have on certain simply-inherited traits. Less obvious is the expression of unfavourable recessive alleles influencing polygenic traits. The individual effects of these genes are small but, taken together, can significantly decrease performance- a phenomenon known as **inbreeding depression**.

6. Biodiversity

An important issue arises in situations where a breed that is native to a particular area appears to have lost its function in that area or elsewhere, and consequently is in danger of becoming extinct. The question to be raised in this situation is whether such a breed should be preserved. The arguments in favour of preservation are that we do not know what type of animals will be required in the future, and that we should therefore preserve the available genetic variation between breeds (bio-diversity) as an insurance against the unknown future. On the other hand, it is argued that people who aim to earn a living from animals cannot afford to look too far into the future; they appreciate the arguments in favour of preservation, but are unable to meet the relatively high cost of preserving populations that they are unlikely ever to utilise during their own lifetimes. At both the national and international level, e.g. FAO and Rare Breeds International, concerted efforts are being made to gather relevant data on breeds that seem threatened by extinction, and to act, where possible, to save them. Interestingly, the two areas that are probably of greatest concern are at the either end of the spectrum of animal improvement. At one end we have a large variety of locally adapted native populations (often in developing countries) that are under threat from the influx of “improved” breeds and strains from developed

countries. And at the other end we have an increasing number of poultry selection lines that are discarded when yet another independent poultry breeding company is taken over by a larger and often multinational breeding company.

7. Technology and animal breeding

The face of animal breeding has changed significantly over the past decades. Animal breeding used to be in the hands of a few distinguished "breeders", individuals who seem to have specific arts and skills to "breed good animals". Nowadays, breeding in particular in livestock species is dominated by science and technology. In some livestock species, animal breeding is in the hands of a few large companies, and the role of the individual breeders seems to have decreased. There are several reasons for this change. Firstly, the breeding industry has adopted scientific principles. Looking was replaced by measuring, and an intuition was partly replaced by calculations and scientific prediction. Other major developments grew from the introduction of biotechnology.

Biotechnology can be broadly defined as the application of biological knowledge to practical needs. These technologies fall generally into two categories, reproductive and molecular. Not all of this is new. Artificial insemination was introduced in cattle in the fifties. There is no doubt that technology had a major impact on rates of genetic improvement in dairy cattle and is just as important to the structure of animal breeding programs. Nowadays, technologies like ovum pick up, in vitro fertilisation, embryo transfer, cloning of individuals, and selection with the use of DNA-information is all on the ground. Some of the technologies are already applied, others are further developed, or waiting application. Finally, rapid development of computer and information technology has greatly influenced data collection and genetic evaluation procedures in animal populations, now allowing comparison of predicted breeding values across farms, breeds or countries.

It is important to recognise that the introduction and exploitation of new technologies have large social impacts. The introduction of breeding methods typically needs to find the right balance between what is possible from a technological point of view and what is accepted by the decision makers and users within the socio-economic context of the production system. Ultimately it is the consumer who decides which technology is desirable and which is not. In most western societies, consumers are increasingly aware of health, environmental and animal welfare issues. Food safety and methods of food production are part of their buying behaviour. However, price and production efficiency are still major factors determining the sustainability of a livestock sector. Successful animal breeding programs need to find and apply the accepted technologies that help them remain competitive. This course is mostly concerned with the technical issues involved in the application of new technologies in animal breeding.

8. Components of breeding programs

Very generally, the aim of animal breeding is to genetically improve populations of livestock so that they produce more efficiently under the expected future production circumstances. Genetic improvement is achieved by selecting the best individuals of the current generation and by using them as parents of the next generation. A breeding program is the organized structure that is put into place to genetically improve livestock populations. This chapter deals with the set-up and evaluation of animal breeding programs.

Message

A breeding program is the organized structure that is set up in order to realize the desired genetic improvement of the population.

Successful genetic improvement requires breeding programs to have (at least) the following components: *i*) A system to record data on selection candidates. Without data on selection candidates it is impossible to identify the best individuals. *ii*) Methods and tools to estimate the genetic merit (breeding value) of selection candidates. This step is referred to as "breeding value estimation" or "genetic evaluation system". *iii*) A system to select the animals that become parents of the next generation, and mate them to produce the next generation. *iv*) A structure to disseminate the genetic improvement of the breeding program into the production population. In most cases, the breeding population and the production population are (partly) separated. Since the aim is to improve livestock production, genetic improvement created in the breeding population should be disseminated into the production population.

Data recording and collection*:* Estimation of breeding values requires phenotypic data on selection candidates. Thus a system has to be set up to routinely record data on selection candidates. The way data is collected depends on the species and the traits in the breeding goal. For example, the product of a dairy cattle breeding company is a straw of semen from a bull. However, milk yield cannot be recorded on bulls. Thus to identify bulls of high genetic merit for milk yield, one has to collect data on daughters of bulls. Dairy cattle breeding schemes therefore have a system to record data on daughters of test bulls. Milk yield of those daughters is recorded on common dairy herds, meaning that farmers are involved in the data recording. In beef cattle breeding, growth performance of bulls can be recorded on the selection candidates themselves, meaning that progeny testing is not necessary. In beef cattle breeding, data collection therefore takes place at testing stations where the performance of selection candidates is recorded. *The quality of the data is fundamental to the success of breeding programs.* Without high quality data, it is impossible to accurately estimate genetic parameters and breeding values.

Breeding value estimation: After data are recorded, breeding values have to be estimated. The common procedure to estimate breeding values in applied livestock breeding is called "BLUP". BLUP and selection index theory have the same theoretical basis; both are based on regression of breeding values on phenotypes. Compared to selection index theory however BLUP has the following advantages; *i*) It accounts for systematic environmental effects. *ii*) BLUP is more flexible than selection index theory and therefore more suitable as an operational tool. *iii*) BLUP takes account of selection.

Selection and mating: Selection and mating takes place after breeding values are estimated. Selection refers to the process of choosing parents to produce the next generation, whereas mating refers to the pairing of selected individuals. Thus selection precedes mating. The selection process determines the genetic improvement of the population over time, whereas the mating process determines how maternal and paternally derived alleles are combined within individuals. This chapter will introduce a number of selection and mating procedures and present theory to understand the effects of the different procedures.

Dissemination of genetic progress: In most species, the breeding and production populations are distinct. Genetic progress is created in the breeding population, but the final aim is to improve livestock production in the entire population. Thus genetic improvement created in the breeding population has to be disseminated into the production population.
In dairy cattle, the breeding and production populations are not strictly separated. Superior cows from the production population can enter the breeding population, meaning that they are selected as bull dams. Genetic progress created in the breeding program is transferred to the dairy farms by the sale of semen of progeny tested bulls to the farmers. The sale of semen is the primary source of income for dairy cattle breeding companies. In addition, a limited number of embryos from the breeding population are sold to the dairy farmers.

The situation is different in pig and poultry breeding. Pig and poultry production are based on crossbreeding systems. The breeding populations consist of purebred lines, which are mated together to produce crossbred offspring. Crossbred offspring are sold to fattening farms or egg producers. The breeding and production populations are therefore completely separated; crossbred production animals cannot enter the purebred breeding populations. Dissemination of genetic superiority of the purebred breeding populations takes place by the sale of crossbred offspring.

Message
A breeding program has the following components: *i*) a data recording system, *ii*) methods and tools for breeding value estimation, *iii*) a selection and mating system and *iv*) a structure to disseminate the genetic improvement into the production population.

9. Design and evaluation of breeding programs

Design of breeding programs: The structure of breeding programs depends on both the species and the breeding goal. The optimum design of a breeding program will differ between species with large reproductive capacity and species with small reproductive capacity, between breeding programs that aim to improve production or reproduction traits, and low heritable traits versus high heritable traits.

Judging the quality of breeding programs: Choosing the best breeding scheme among a number of alternatives requires yardsticks to measure the quality of breeding schemes. Such yardsticks can be developed only when there is a well-defined breeding goal. Given that the breeding goal is clearly defined, there are three criteria that summarize the quality of a breeding program. These are:

1. Selection response for the breeding goal traits.
2. Maintenance of genetic diversity as measured by the rate of inbreeding.
3. Costs of the breeding program.

Selection response for the breeding goal traits is the revenue of a breeding program, whereas loss of genetic diversity and financial costs are the expenses of a breeding program. Selection response, loss of genetic diversity and financial costs are expressed in different units. The problem therefore is to combine them into a single criterion for the quality of a breeding program.

A comparison of breeding schemes based on selection response and the rate of inbreeding can be done as follows. To avoid long-term loss of genetic diversity an upper limit can be set to the rate of inbreeding. Next, alternative breeding schemes can be judged by comparing their selection response at the same rate of inbreeding. The scheme with the highest selection response at the same rate of inbreeding (e.g. 1%/generation) is the best scheme.
It is more difficult to combine selection response and cost into a single criterion. The question is whether the revenues from an increase in selection response, for example in the form of increased market share, makes up for the cost of increased selection response. Hence, this is not a genetic issue but primarily a commercial and business issue.

Evaluation of breeding programs: Once a breeding program is operational it is essential to routinely evaluate the results. Evaluation may consist of comparing realized genetic improvement and rates of inbreeding with values expected when designing the breeding program. When there are clear differences between expected and realized selection response and inbreeding, then one needs to find the causes of those discrepancies and if possible improve the breeding program. Reasons that breeding programs do not yield the expected genetic improvement are: *i*) the use of inappropriate models for breeding value estimation, for example when the models do not include systematic environmental effects that are present in the data; *ii*) overestimation of the genetic parameters (e.g. h^2) resulting in biased EBVs and overprediction of the expected response; iii) preferential treatment among selection candidates resulting in selection of individuals that received "good treatment" instead of genetically superior individuals, and *iv*) unexpected correlated response in other traits.

Message
The quality of alternative breeding schemes can be judged by comparing selection response, rate of inbreeding and costs of the alternatives.

Methods to design and evaluate breeding programs: To compare alternative breeding programs we need methods to quantify expected rates genetic improvement and inbreeding of the alternatives. In other words, we need methods to predict rates of gain and inbreeding of breeding programs. From a methodological point of view, quantifying the expected rates of gain and inbreeding can be done in two manners, either stochastically or deterministically. Stochastic simulation is often the easiest way, but in most cases deterministic simulation gives more insight.
With stochastic simulation, the breeding program is simulated in detail on a computer. Stochastic simulation consists of the following cycle. 1. Breeding values and phenotypes of individuals in the base generation are simulated. 2. Breeding values are estimated for the base generation animals by performing BLUP analyses on their simulated phenotypes. 3. Based on the estimated breeding values coming from the BLUP analyses, a number of animals is selected to become parents of the next generation. 4. The selected animals are mated and offspring from the matings are simulated. Next, steps 2, 3 and 4 are repeated until the desired number of generations is simulated. Because in stochastic simulation we simulate an entire population of "real" animals, the rates of gain and inbreeding can simply be estimated from the simulated data. Hence, after simulating the breeding scheme, the next step is to analyze the simulated data to quantify the rate of gain and inbreeding of the breeding scheme. Multiple replicates of the population are simulated, and the rates of gain and inbreeding are averaged over replicates.

The advantage of stochastic simulation is that one can mimic the true breeding program in detail, because the individual animal is simulated. Hence, stochastic simulation can be very precise. However, there are two disadvantages related to the use of stochastic simulation to evaluate breeding schemes. First, stochastic simulation is time consuming, particularly when large populations are simulated. Even with modern computers, simulation of a sufficient number of replicates of a large breeding scheme may take several hours or even days. Hence, stochastic simulation is less suited as an operational tool to quickly evaluate a number of alternatives. Second, with stochastic simulation the user often does not gain as much insight into the breeding scheme as with deterministic simulation. For example, with stochastic simulation the user will observe that shorter generation intervals generally go together with higher gain, but the deterministic equation $\Delta G = ir_{IH}\sigma_H/L$ directly shows that gain is inversely proportional to the generation interval. Hence, because stochastic simulation does not explicitly model mechanisms like accuracy, generation interval, etc, it can be difficult to extend the result to other breeding schemes that were not simulated themselves.

Instead of using stochastic simulation, one can use deterministic methods to quantify expected gain and inbreeding from alternative breeding schemes. Deterministic methods do not mimic the breeding program on the individual animal level, but use (deterministic) equations to predict gain and inbreeding. For example, prediction of the rate of gain by using the expression that $\Delta G = ir_{IH}\sigma_H/L$ is a deterministic methods. Hence, modeling the mechanisms that determine gain and inbreeding as mathematical equations allows us to quantify the expected outcome of a breeding program. To use deterministic methods one needs to know/derive the mechanisms determining gain and inbreeding; it requires more insight into quantitative population genetics than stochastic simulation.

Advantages of deterministic methods are 1). It takes limited computation time, so that many alternatives can be compared within limited time, and 2). Because the mechanisms are modeled explicitly, it gives a lot of insight into gain and inbreeding in breeding programs. In some cases, however, it may be difficult to derive accurate deterministic methods. Hence, there is a risk that deterministic methods are not precise if they do not properly model the mechanisms determining gain and inbreeding in populations. In complicated cases, stochastic simulation may be used to check the accuracy and validity of the deterministic equations. Hence, stochastic simulation may be helpful to validate and improve deterministic models, and in this way we improve our understanding of the mechanisms determining genetic improvement and rates of inbreeding in populations.

In this course we will mainly deal with deterministic models. The reason is that for many important situations deterministic methods are available and they provide more insight than stochastic models.

Message

The expected selection response and inbreeding of breeding schemes can be determined by using either stochastic simulation or deterministic methods. Deterministic methods provide more insight and are computationally fast. Stochastic simulation is precise and useful to validate deterministic methods.

10. This course

The course "Animal breeding strategies" introduces the quantitative genetic principles underlying the design and implementation of genetic improvement programs in livestock species. Those principals will also apply to companion animals, populations of endangered breeds and zoo populations. The basic quantitative genetic principles used in this course are handled in Falconer and Mackay (1996)[3].

The lectures start with a general overview of the field. This course focuses on the definition of breeding objectives and the genetic evaluation of breeding strategies. To achieve this, much of the course is devoted to the general principles involved in deriving economic weights of the various traits that might be genetically improved, making selection decisions between animals, designing breeding strategies and determining which strategies will make optimum progress. What is presented is a selection of some of the more common tools used in defining breeding objectives and designing and evaluating breeding strategies. These tools should be adequate to tackle many basic practical problems in animal breeding and provide background to using more complex methods.

Estimation of breeding values using best linear unbiased prediction (BLUP) is an important element in animal breeding but this lies outside the scope of this course (see AnS562). Attention will be paid to selection index theory but the emphasis lies on prediction of genetic gain and not on genetic evaluation of animals.

Lecture notes provide students with detailed knowledge on issues related to the design of breeding programmes for farm animals. Lectures will 'guide' the student through these notes. In addition problems will be supplied.

[3] Falconer D.S. and T.F.C. Mackay, 1996. Introduction to quantitative genetics. Longman fourth edition.

Genetic andEconomic Aspects of Animal Breeding Strategies

Chapter I Introduction

- Building blocks and structure of animal breeding programs
- Methods to model/evaluate breeding programs/strategies
- Design of breeding programs/strategies
- Marker-assisted Selection

What Do We Mean By Breeding Strategies?

- Tactics designed to integrate new technologies and to improve old ones, for the purpose of maximizing performance of existing stock (Charles Smith).
- Integration of the components of a breeding program into a structured system for genetic improvement, with the aim to maximize an overall objective.

General aim for animal breeding strategies:

Obtain future generations of animals that will produce more efficiently under future production circumstances

Basic Principle of making genetic progress in a population

Mate the "best" to the "best" and do that as quickly as possible.

$$\text{Genetic Gain / Yr} = \frac{\text{genetic superiority selected parents}}{\text{generation interval}}$$

$$= \frac{\text{intensity} \times \text{accuracy} \times \text{genetic st. dev.}}{\text{generation interval}}$$

Mate the "Best" to the "Best" and Do That As Quickly As Possible

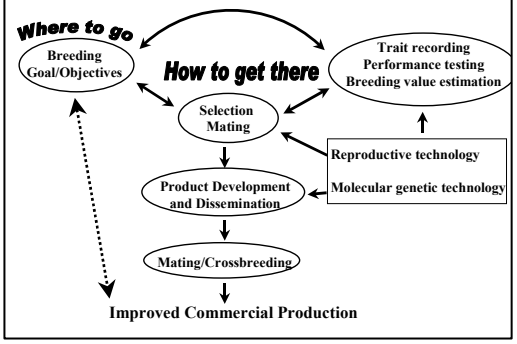
Some Questions

- How to find/identify the "best"?
- "Best" for what?
- What are the limits to use of only the "best"?
- How can we shorten the generation interval?
- What are the limits?
- How can a breeding company make a profit from this?
  - "Breeding is a business" Lush, 1945
- How do technologies enter into this?

↓

BREEDING STRATEGIES

Basic Components of a Successful Breeding Program/Strategy



Basic Components of a Successful Breeding Strategy

- Breeding Goal or Objectives - where should we go?
  - Which traits must be improved? - Economic traits
  - How important is each trait? - Economic values
    - Focus on improvement of Economic efficiency/profit
    - Consider (future) consumer demands
- Trait recording, Performance testing, Breeding value estimation

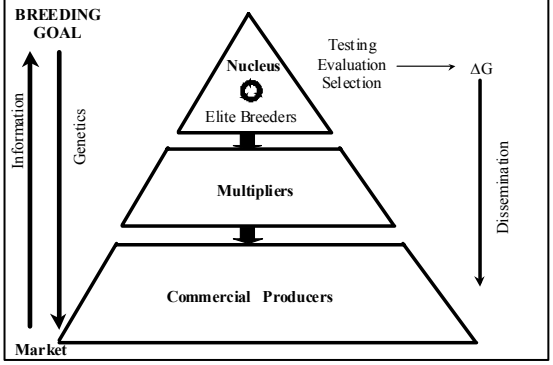
  Identify animals with "best" genetics - relative to breeding goal
  - trait performance recording and testing programs
    - which traits should be recorded and on which animals?
      - field recording
      - performance test stations / nucleus herds
      - progeny testing
  - pedigree registration

Genetic Evaluation ↔ **Selection Index (Total merit index)**

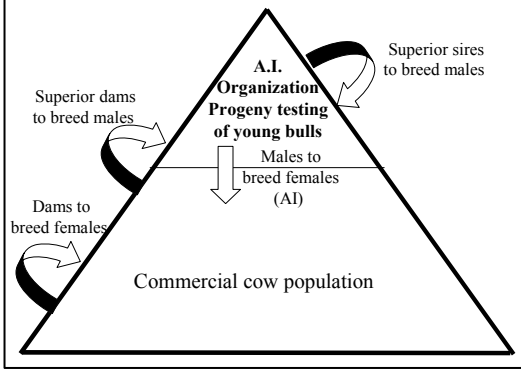
Components of a Successful Breeding Strategy (cont'd)

- Selection and mating
  - use best animals to breed next generation → genetic improvement
    - How many and which animals should we select?
    - How should we mate them?
    - Should reproductive technology be used to increase # progeny/parent?
    - balancing rate of genetic gain and inbreeding (and cost)
- Product Development and Dissemination
  - program for marketing and distribution of superior genes into the commercial sector
    - progeny testing, AI
    - multipliers
- Mating/Crossbreeding
  - optimize combinations of genetic material in commercial animals

Typical Structure of Livestock Breeding Programs



Dairy Cattle Progeny Testing Program



Basic steps in the design of breeding programs (Harris '84)

1) Describe the production system(s)
2) Formulate the objective -simplified and comprehensive- of the system
3) Choose a breeding system and breeds
4) Estimate selection parameters and (discounted) economic values
5) Design an animal evaluation system
6) Develop selection criteria
7) Design matings for selected animals
8) Design a system for expansion - dissemination - of genetic superiority
9) Compare alternative programs

Development of Breeding Strategies Summary

- Integration of the components of a breeding program into a structured system for genetic improvement, with the aim to maximize an overall objective (genetic gain, market share).
- Evaluate opportunities for improving upon current strategies.
- Evaluate the potential of new technologies.
  - How can they best be incorporated into current strategies?
  - Can their benefits best be capitalized on in a redesigned breeding structure?

Breeding Strategies - Summary

What tools are necessary to develop optimum strategies?

- Quantitative genetics theory
  - Predicting response to selection, selection index, inbreeding, etc.
- Systems analysis
  - Predicting and optimizing response in overall objective
- Common sense
- An open mind

Chapter 2

Stochastic Methods to Model Breeding Programs

2.1 Introduction

The objective of genetic improvement of livestock is to enhance the genetic level for traits of interest in a population through genetic selection such that some overall goal is achieved or enhanced. The overall goal can usually be described in economic terms (e.g. maximize profit per animal per year) and will be discussed further in chapter 7.

There are many factors that determine the success of a breeding program. These include design and implementation issues. In this course, we will primarily focus on factors related to the design of genetic improvement programs, which include factors such as population size, numbers of animals to select, criteria for selection, etc.. Because of the number of factors involved, the number of alternative programs is numerous. However, ultimately only one program can be implemented; animal breeders don't have the luxury of trying out different options and then deciding which one to go with. Thus, we need some other means of deciding *a priori* which breeding program will maximize our overall objective. This requires the ability to model breeding programs and to predict outcomes from alternative breeding programs. Furthermore, if a good understanding can be developed of the impact alternative design factors have on program outcomes, this will lead to the development and choice of better breeding programs. The development of this knowledge and associated methods and tools are the focus of this course.

2.2 Quantitative Genetic Model

Because most traits of interest in livestock are multifactorial in nature, i.e. affected by a potentially large number of individual genes along with environmental factors, quantitative genetic theory has become the primary basis for the development of methods to develop, model, and evaluate alternative breeding programs. The basis of this theory is the infinitesimal genetic model (Falconer and Mackay, 1996). The purpose of this section is to briefly review this theory as a basis for developing methods to model breeding programs.

The quantitative genetic model for the phenotype of animal i is: $y_i = \mu + g_i + e_i$ (2.1)

where μ is an overall mean (or sum of fixed effects), g_i is the animal's genetic value, and e_i it's random environmental effect. For the purposes of the majority of this course, we will assume we are dealing with additive traits such that g_i refers to the additive genetic or breeding value.

Variables g_i and e_i are assumed normally distributed with means zero and standard deviations σ_g and σ_e. Strictly, these assumptions hold for g_i only for an unselected (base) population and both the mean and variance will change as a result of selection, as will be described later on in the course.

With the exception of the sex chromosomes, which we will ignore for the moment, all animals carry two copies of every gene. One copy is inherited by random sampling from the two copies carried by the male parent (sire) and the other copy is inherited by random sampling from the two copies carried by the female parent (dam). It follows that the additive genetic value of an offspring, g_o, can be partitioned into three sources, and modeled as follows:

$$g_o = ½\, g_s + ½\, g_d + g_m \tag{2.2}$$

where g_s and g_d are the additive genetic values of the sire and dam and g_m is the Mendelian sampling contribution. The Mendelian sampling contribution reflects the random selection of copies of parental genes. Since genes are inherited at random from the parents, the average values of g_m over a large number of progeny is expected to be zero.

Mathematically, it is said that the expectation of g_m, $E(g_m)$, is zero. But for any particular individual, g_m has a real value which varies between individuals. The range of values of g_m is determined by its variance, which in the absence of inbreeding, is expected to be

$$E(\sigma^2_{g_m}) = ½\, \sigma^2_{g_0} \tag{2.3}$$

where $\sigma^2_{g_0}$ is the initial genetic variance in the population prior to any selection. The reason for noting the requirement that there be no prior selection in the population will become clear later in the course.

With inbreeding, the expected variance of Mendelian sampling terms is reduced by a factor $[1- ½(F_s+F_d)]$, where F_s and F_d are the inbreeding coefficients of the sire and dam. Thus:

$$E(\sigma^2_{g_m}) = ½\, [1- ½(F_s+F_d)]\, \sigma^2_{g_0} \tag{2.4}$$

2.3 Stochastic Models for Evaluation of Breeding Programs

The simple quantitative genetic models described in the previous paragraph can be used to simulate a breeding program and evaluate its outcomes. Simulations in animal breeding can be divided into three types:

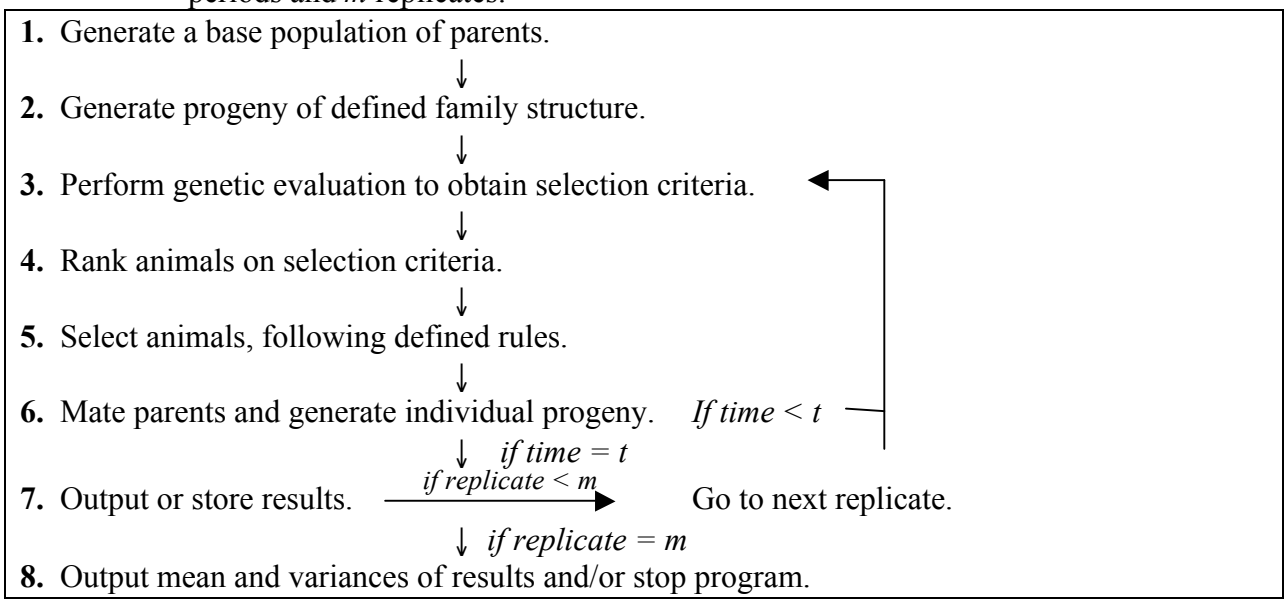
1) stochastic simulation (or sometimes called Monte Carlo simulation)
2) deterministic simulation
3) combination of stochastic and deterministic simulation.

Stochastic simulations use random number generators to simulate variability. The two most common types of random generators needed are those for the uniform and the normal distribution. Most statistical software programs have functions that can generate these. Excel has a uniform random number generator: RAND(), which returns a uniform number between 0 an 1.

Using the Inverse Transform method (http://www.mathwave.com/articles/random-numbers-excel-worksheets.html) this function can be used in combination with inverse cumulative distribution functions to generate numbers from other distributions in Excel. For example, to generate a random number from a standard normal distribution, use: NORMINV(RAND(),0,1). The function NORMINV(p, mean, st.dev.) returns the truncation point for a normal distribution that has a fraction p below it. So by drawing p from a random uniform distribution (0,1), a random truncation point is generated based on the cumulative distribution function.

With stochastic simulations in animal breeding, which will be described here, a population of animals is simulated by generating records for each animal in the population by random sampling from pre-defined distributions which are determined by the rules of inheritance and origins of environmental effects imposed on the model. A model for stochastic simulation of a breeding program is schematically represented in Figure 2.1. The steps involved are described in further detail in what follows.

Figure 2.1 General schematic of a stochastic simulation of a breeding program with *t* time periods and *m* replicates.



2.3.1 Generating Base Population Parents

A base population is generated according to the rules of inheritance and structure of the population defined by the program control variables. For example, if the phenotype of a single trait, explained by the simple additive inheritance model plus a random environment effect, is

$$y_i = \mu + g_i + e_i$$

and there are n_m males and n_f females in the base assumed to be randomly selected, unrelated, and non-inbred, then the effects for an animal in the base population could be defined by the following programming steps:

Do for each animal i:

1. *r = random number from normal distribution with mean 0 and variance 1*
2. $g_i = r * \sigma_{g_o}$ *where* σ_{g_o} *is the additive genetic st. deviation in the base population.*
3. *r = new random number from normal distribution with mean 0 and variance 1*
4. $e_i = r * \sigma_e$ *where* σ_e *is the standard deviation of environmental effects.*
5. $p_i = \mu + g_i + e_i$, where μ is the pre-defined population mean, i.e. a constant.
6. *Store* p_i, g_i; *and* e_i

This can be repeated for all animals in the base population. In order to enable the construction of a pedigree file, animals should be given a unique identification number. The simulation can be extended to include other genetic effects, such as dominance or systematic environmental effects such as age, herd or year. Virtually all programming languages have a random number generator or an associated library of subroutines containing a routine for random number generation.

2.3.2 Generating Progeny

Once parents are generated, mating pairs are allocated and progeny generated. Recalling from equation 2.1 that the phenotype of progeny *k* of male parent *i* and female parent *j* is

$$y_{ijk} = \mu + \tfrac{1}{2} g_{s_i} + \tfrac{1}{2} g_{d_j} + g_{m_{ijk}} + e_{ijk} \qquad (2.5)$$

where g_{s_i} and g_{d_j} are the known additive genetic values of the sire and dam, $g_{m_{ijk}}$ is the Mendelian sampling contribution for individual *k* and e_{ijk} is the environmental effect. The contributions of $g_{m_{ijk}}$ and e_{ijk} are obtained for each progeny in turn by sampling from a random normal distribution with mean *0* and variance 1 and multiplying the random number by σ_{g_m} or σ_e, where $\sigma^2_{g_m} = \tfrac{1}{2}\sigma^2_{g_o}$ in the absence of inbreeding,
or $\sigma^2_{g_m} = \tfrac{1}{2}(1 - \tfrac{1}{2} F_{s_i} - \tfrac{1}{2} F_{d_j})\sigma^2_{g_o}$ in the presence of inbreeding, where F_{s_i} and F_{d_j} are the inbreeding coefficients of the two parents. Fixed effects can then be added to p_{ijk} according to the structure specified by the design.

2.3.3 Deriving the Selection Criterion

The selection criterion, such as the phenotypic record, a selection index, or BLUP evaluation, would be estimated for each simulated animal as if in real life. A subroutine of the program would be written to perform the evaluations. The nature of the selection criterion will determine the amount of data to be stored. For example, a selection index involving only collateral relatives would not require the parental records to have been stored, whereas animal model BLUP evaluation would require all animals and relationships back to the base population to be stored. In contrast to selection indexes, BLUP evaluation will be expensive for computing time because of the iterative nature.

Selection index or BLUP requires defined variances of traits for single trait evaluation and variance/covariance matrixes for multiple traits. Usually these would be set to the base population values, though false values may be given deliberately if estimation of sensitivity to parameter for BLUP is under investigation. If relationships back to the base generation are included, BLUP automatically allows for change in genetic variance due to selection (see Chapter 5).

With selection indexes, the appropriate variance/covariance among traits and relatives at each generation are required. A decision will therefore have to be taken as to whether to use constant parameters over time or to allow them to change. When the same set of parameters is used over time it seems logical to use the parameters from the base population, which were also used in simulating the data. In real life, the base population parameters can only be estimated and it might therefore be interesting to investigate the consequences of using other than the true parameters. Population parameters will change over time as a result of selection. These changes can be allowed for in constructing the selection index. In that case a method is needed to obtain the parameters at each point in time. The parameters could be estimated from the phenotypic and the true additive genetic values (g_{ijk}, g_{s_i}, g_{d_j}). This, however, would not be possible in real life and hence would not give realistic results. Alternatively, parameters could be estimated using phenotypic records or changes in parameters could be predicted from the selection strategy. Interpretation of the results will obviously depend on the assumptions made.

2.3.4 Selecting and Mating Animals for Breeding

In order to produce the next generation of offspring, one needs to define the method of selecting the animals to be used as parents and the procedure used in mating the selected parents. In the previous step, the selection criterion has been estimated for all candidates for selection. Truncation selection is commonly used for selection, in which the animals with the highest value for the selection criteria are selected. This requires that males and females are separately ranked in order of merit for the selection criteria. Efficient ranking routines are available in most language libraries. Apart from the method of selection, the user has to specify the number of animals to be selected and the category of animals, which are eligible for selection. One might, for example, restrict the selection to animals of one particular age class only or have no restriction other than that animals need to be old enough to be able to reproduce. In the latter case, selection will be across age groups and it is important to specify up to what age animals are eligible for selection.

In the absence of restrictions on selection, selection is simply a process of designating the required number of top ranking animals as parents. With complete assortative mating, the top ranked male is allocated to the n top ranked females, the second ranked male to the next n females and so on; where n is the number of females per male. With random mating, each selected female is allocated a random deviate, and the females are then ranked on the random deviate and mating proceeds as above.

An advantage of stochastic simulation is that restrictions can be imposed on selection and mating. Common examples would be restrictions defining the maximum number of full and half

sibs that can be selected as parents, and restrictions that full and half sibs may not be mated together. The imposition of restrictions may make some animals ineligible for mating so that more animals must be available for mating than indicated by the defined proportions to be selected.

2.3.5 Inbreeding Coefficients

Traditional methods of estimating inbreeding coefficients of individual animals by tracing path coefficients, or directly from a complete relationship matrix rapidly become time consuming and expensive of storage space as population sizes and number of generation's increase. With this method it was often impractical to estimate inbreeding coefficients in stochastic simulations. Several algorithms have been developed, however, for efficiently deriving inbreeding coefficients from a pedigree file (e.g. Tier, 1990). Use of these algorithms reduces computer time 10-100 fold compared to traditional methods. An additional trick is to recognize that all full sibs have the same inbreeding coefficient so that only one member of the family needs to have the coefficient estimated. Even so, calculation of inbreeding coefficients can still be expensive of computing time when simulating several thousand animals in each of several generations.

2.3.6 Completing the Cycle

Once mating pairs are allocated, progeny can be produced and the cycles repeated until the desired number of time periods has been achieved. At this point, summary statistics can be printed or stored, and the next replicate started. The number of replicates required will depend principally on the required accuracy of estimates of response and variance of response, which are largely dependent on the size of the population and the number of generations simulated. Large populations have low variance of response and therefore require fewer replicates for a given level of accuracy.

Stochastic simulations are often used to validate deterministic simulations. In this case it is desirable to have very accurate estimates of output parameters to estimate biases in the deterministic program. Typically, with smaller populations, several hundreds to 1000 replicates are run. But when using stochastic simulations to evaluate alternative breeding programs, very small differences between alternatives are rarely of practical interest so that often fewer, say 100, replicates can suffice. In practice the number of replicates required can be determined once a few initial runs have indicated the variance to be expected between runs for a particular size and type of population.

2.3.7 Multiple Trait Simulations

Multiple trait simulations are a little more difficult because they require simulation of correlated random variables. For Excel, a user-defined function is available from http://homepage2.nifty.com/hashimoto-t/misc/mnormrand-e.html#download that allows you to generate correlated random variables based on a defined vector of means and a variance-covariance matrix. See Excel file mnormrand.xls .

Alternatively, simulation of correlated random variables can be achieved by deriving the *n* uncorrelated principal components of the genetic and environmental variance covariance matrix among the *n* traits, generating random deviates for each principal component in turn and then back-transforming these to obtain random deviates for the original traits. Alternatively, an approach using Cholesky decomposition of the original variance covariance matrixes can be used which has advantages in terms of computing ease and time. The Cholesky decomposition approach is explained in Appendix C and some examples of simulating correlated traits and records for related individuals are given by Van Vleck (1993). These same methods can deal with simulations involving other covariances among random variables, such as *g* x *e* covariance and additive x dominance genetic covariances.

2.3.8 Genome-level models

In the previous, the genetic component was modeled as a normally distributed variable, using the infinitesimal genetic model. This model assumes that the trait is affected by a large number of unlinked loci, each of small effect. Stochastic models also allow the modeling of a more realistic genetic architecture of the trait by simulating individual loci and their placement on chromosomes within the genome, along with genetic markers. These models require specification of the number of loci, the number and length of chromosomes that these loci are located on, and their position (in centi-Morgans, cM) on these chromosomes. Then, the following parameters must be specified for each locus:

1) Locus position - loci could be positioned on chromosomes at random by sampling from a uniform distribution, or evenly distributed across the genome.
2) Number of alleles.
3) Allele frequencies in the base population – these could be set to be equal or sampled from some distribution
4) Genotypic effects associated with each genotype - these can, for example, for a locus with two alleles B, b, be based on the standard single locus genetic model with genotypic values of $+a_l$, d_l, and $-a_l$ for genotypes BB, Bb, and bb at locus *l* (Falconer and MacKay, 1996). Genotypic values assigned to each locus could be sampled from an assumed distribution of gene effects, such as from a gamma distribution (e.g. Hayes and Goddard, 2003), in an attempt to reflect reality. In addition, epistatic effects could be allowed for by assigning genotypic effects to combinations of genotypes at multiple loci.

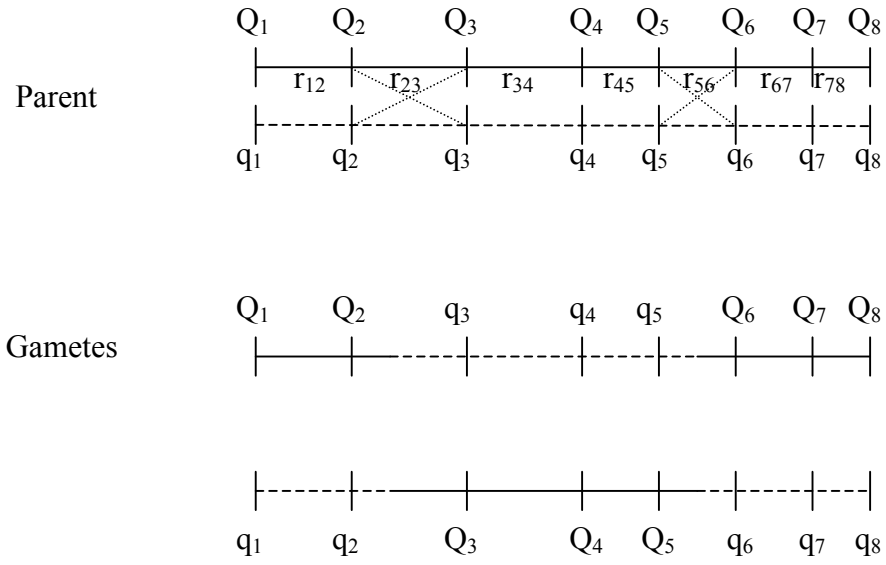
For the base population, alleles at locus *l* for individual *i* can then be assigned by drawing two random numbers *u* from a uniform (0,1) distribution. For example, for a locus *l* with allele frequency f_j^l for alleles B_j (j=1, . . , m_l), allele *j* is assigned if $\sum_{k=1}^{j-1} f_k^l < u < \sum_{k=1}^{j} f_k^l$. This random sampling of alleles assumes the base population is in Hardy-Weinberg and gametic phase equilibrium (Falconer and MacKay, 1996).

The genetic value of individual *i* then is the sum of genetic effects at each of the *q* loci:

$g_i = \sum_{l=1}^{q} g_i^l$, where g_i^l is the genotypic value at locus *l* for individual *i*, which is based on the simulated genotype of for locus *i* and the genotypic value that is associated with this genotype.

If all loci are unlinked, progeny genotypes at each locus can be simulated by randomly drawing one of the two alleles of the sire and one of the two alleles of the dam. If loci are linked, recombination must be allowed for. Consider the two haplotypes for a parent in Figure 2.1.

Figure 2.1. Simulation of Mendelian inheritance with linked loci with recombination in intervals 23 and 56.



To create a progeny from this parent, the first step is to simulate the production of two gametic chromosomes through meiosis. This can be simulated as follows

1) Starting with the first interval, 12, the probability of recombination (r_{12}) or not ($1-r_{12}$) is drawn from a uniform normal distribution. If $u[0,1] < r_{12}$, then a recombination takes place and we end up with the following two recombinant haplotypes: $Q_1,q_2,q_3,q_4,q_5,q_6,q_7,q_8$ and $q_1,Q_2,Q_3,Q_4,Q_5,Q_6,Q_7,Q_8$, since all alleles downstream from the cross-over are switched. If $u[0,1] > r_{12}$ then the parental chromosomes stay intact.

2) Proceed to the next interval and draw presence or absence of a recombination event in that interval: if $u[0,1] < r_{23}$ then there is a recombination event and we end up with the following two recombinant haplotypes (assuming there also was recombination in interval 12): $Q_1,q_2,Q_3,Q_4,Q_5,Q_6,Q_7,Q_8$ and $q_1,Q_2,q_3,q_4,q_5,q_6,q_7,q_8$. If there is no recombination event, then the haplotypes generated in step 1 remain intact.

3) Proceed through all intervals consecutively as described above.

Once a pair of recombinant gametes has been created, a random one of the two is sampled to generate the progeny. A similar procedure is used to generate the other parental chromosome.

Note that this method assumes that recombination events in adjacent intervals are independent (no interference – Haldane mapping function). If there is interference, probabilities of recombination in interval *i* must be adapted, depending on presence or absence of a recombination event in interval *i*-1.

Simulation of genomic selection programs or data for genome-wide association analysis also requires simulation of historical generations of the population, in order to generate linkage disequilibrium between loci. A useful freely available software program for this purpose is QMSim (http://www.aps.uoguelph.ca/~msargol/qmsim/QMSim_documentation.pdf Sarargolzaei and Schenkel, 2009, University of Guelph). After download, you can run this program from command line, using ./QMSim [parameterfile] -o The download provides several example input parameter files.

2.4 Advantages and Disadvantages of Stochastic Models

Stochastic simulation depends on relatively simple rules determining inheritance from one generation to the next, along with description of the criteria on which all animals will be selected for breeding. Thus, for a given degree of complexity of the breeding program, stochastic simulations are often relatively easy to write compared to the deterministic models that will be described later. In addition, stochastic models allow alternative genetic models to be evaluated, while deterministic models are primarily restricted to the infinitesimal genetic model. However, see Chapter 12 for deterministic models with individual genes along with an infinitesimal polygenic component.

With stochastic simulation, the result of any one run reflects random sampling events so that to obtain the mean expected response, many replicate runs must be made; but this also allows the variance of the response to be estimated. Because each animal in the population is individually identified, stochastic programs can take up a large amount of storage space and involve a very large number of mathematical operations for every run. This, combined with the need to replicate, means that stochastic programs take much longer, often very much longer, to run than deterministic programs.

Stochastic simulation also does not provide much insight into the impact of various factors on response to selection and does not lend itself easily to optimization of breeding programs. Hence, in the remainder of this course, the main focus will be on deterministic models, to facilitate an understanding of the factors that affect the outcomes of breeding programs. With the tremendous increases in computing power, however, stochastic models have become more and more attractive and used for the evaluation and analysis of breeding programs in both research and practice.

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:

$$g_o = \tfrac{1}{2}g_s + \tfrac{1}{2}g_d + g_m \qquad (3.1)$$

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, $E(g_i)$, from a given pair of parents is given by

$$E(g_o) = \tfrac{1}{2}g_s + \tfrac{1}{2}g_d$$

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, $E(\bar{g}_o)$. This can be obtained from the average genetic value of the selected parents $\bar{g}^*_s$ and $\bar{g}^*_d$, where * indicates that the variable refers to selected individuals:

$$E(\bar{g}_o) = \tfrac{1}{2}\bar{g}^*_s + \tfrac{1}{2}\bar{g}^*_d \tag{3.2}$$

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 ($\bar{g}_s$ and $\bar{g}_d$):

Thus:

$$\begin{aligned} E(\bar{g}_o) &= \tfrac{1}{2}(\bar{g}^*_s - \bar{g}_s + \bar{g}_s) + \tfrac{1}{2}(\bar{g}^*_d - \bar{g}_d + \bar{g}_d) \\ &= \tfrac{1}{2}(\bar{g}_s + \bar{g}^*_s - \bar{g}_s) + \tfrac{1}{2}(\bar{g}_d + \bar{g}^*_d - \bar{g}_d) \\ &= \tfrac{1}{2}(\bar{g}_s + S_s) + \tfrac{1}{2}(\bar{g}_d + S_d) \\ &= \tfrac{1}{2}(\bar{g}_s + \bar{g}_d) + \tfrac{1}{2}(S_s + S_d) \end{aligned} \tag{3.3}$$

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.:

$$S_s = \bar{g}^*_s - \bar{g}_s \tag{3.4}$$

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 Δg. Using the R notation, the expectation of R is given by:

$$E(R) = \bar{g}_o - \bar{g}_p \tag{3.5}$$

Where

$$\bar{g}_p = \tfrac{1}{2}(\bar{g}_s + \bar{g}_d)$$

Using this and the expression of $\bar{g}_o$ 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:

$$\begin{aligned} E(R) &= \tfrac{1}{2}(\bar{g}_s + \bar{g}_d) + \tfrac{1}{2}(S_s + S_d) - \tfrac{1}{2}(\bar{g}_s + \bar{g}_d) \\ &= \tfrac{1}{2}(S_s + S_d) \end{aligned} \tag{3.6}$$

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

$$E(g \mid \hat{g}) = \hat{g}$$

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

$$E(g_o) = \tfrac{1}{2}\hat{g}_s + \tfrac{1}{2}\hat{g}_d = \hat{g}_p$$

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 $\bar{\hat{g}}$ as:

$$\begin{aligned} E(\bar{g}_o) &= \tfrac{1}{2}\bar{\hat{g}}_s^* + \tfrac{1}{2}\bar{\hat{g}}_d^* \\ &= \tfrac{1}{2}(\bar{\hat{g}}_s + \hat{S}_s) + \tfrac{1}{2}(\bar{\hat{g}}_d + \hat{S}_d) \end{aligned} \qquad (3.7)$$

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

$$\hat{S} = \bar{\hat{g}}^* - \bar{\hat{g}} \qquad (3.8)$$

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:

$$E(R) = \hat{g}_o - \bar{g}_p = \tfrac{1}{2}(\hat{S}_s + \hat{S}_d) \qquad (3.9)$$

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.

$$g_o = \tfrac{1}{2}(\tfrac{1}{2}g_{ss} + \tfrac{1}{2}g_{ds} + g_{ms}) + \tfrac{1}{2}(\tfrac{1}{2}g_{sd} + \tfrac{1}{2}g_{dd} + g_{md}) + g_m \qquad (3.10)$$

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 $\hat{g}_s$ and $\hat{g}_d$ 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

$$y_i = a + b_{yx} x_i + e_i \qquad (3.11)$$

and a prediction of y given x is

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

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

$$b_{yx} = \frac{\sigma_{xy}}{\sigma_x^2} = r_{xy} \frac{\sigma_y}{\sigma_x} \tag{3.13}$$

where σ_{xy} is the covariance of x and y, σ_x^2 is the variance of x, and r_{xy} is the correlation between y and x, which is given by

$$r_{xy} = \frac{\sigma_{xy}}{\sqrt{\sigma_y^2 \sigma_x^2}} \tag{3.14}$$

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),

$$\hat{g}_i = \bar{g} + b_{gI} (I_i - \bar{I}) \tag{3.15}$$

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_{gI} 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:

$$\bar{\hat{g}}^* = \bar{g} + b_{gI} (\bar{I}^* - \bar{I}) \tag{3.16}$$

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:

$$\hat{S} = \bar{\hat{g}}^* - \bar{g} = b_{gI} (\bar{I}^* - \bar{I}) \tag{3.17}$$

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.

$$i = (\bar{I}^* - \bar{I}) / \sigma_I \tag{3.18}$$

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

$$(\bar{I}^* - \bar{I}) = i \sigma_I \tag{3.19}$$

and substituting (3.18) into (3.17) we get

$$\hat{S} = b_{g.I}\ i\ \sigma_I \tag{3.20}$$

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

$$b_{g.I} = r_{gI} \frac{\sigma_g}{\sigma_I} \tag{3.21}$$

hence,
$$\hat{S} = r_{gI} \frac{\sigma_g}{\sigma_I} (i\ \sigma_I) = i\ r_{gI}\ \sigma_g \tag{3.22}$$

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):

$$\begin{aligned} E(\bar{g}_o) &= \tfrac{1}{2}(\bar{g}_s + \hat{S}_s) + \tfrac{1}{2}(\bar{g}_d + \hat{S}_d) = \\ &= \tfrac{1}{2}(\bar{g}_s + i_s\, r_{g,I_s}\, \sigma_g) + \tfrac{1}{2}(\bar{g}_d + i_d\, r_{g,I_d}\, \sigma_g) \end{aligned} \tag{3.23}$$

or model response per generation using equation (3.9):

$$R = \tfrac{1}{2}(S_s + S_d) = \tfrac{1}{2}(i_s\, r_{g,I_s}\, \sigma_g + i_d\, r_{g,I_d}\, \sigma_g) \tag{3.24}$$

Methods to derive the accuracy of selection, r_{gI} , 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, σ_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,
$$\sigma_{gI} = \sigma_{gy} = \sigma_{g,g+e} = \sigma_g^2$$

Thus
$$r_{gI} = r_{gy} = \frac{\sigma_g^2}{\sqrt{\sigma_g^2 \sigma_p^2}} = \frac{\sigma_g}{\sigma_p} = h \tag{3.25}$$

Where h is the square root of heritability.

Thus, from (3.22), $\hat{S} = i\, h\, \sigma_g$ (3.26)

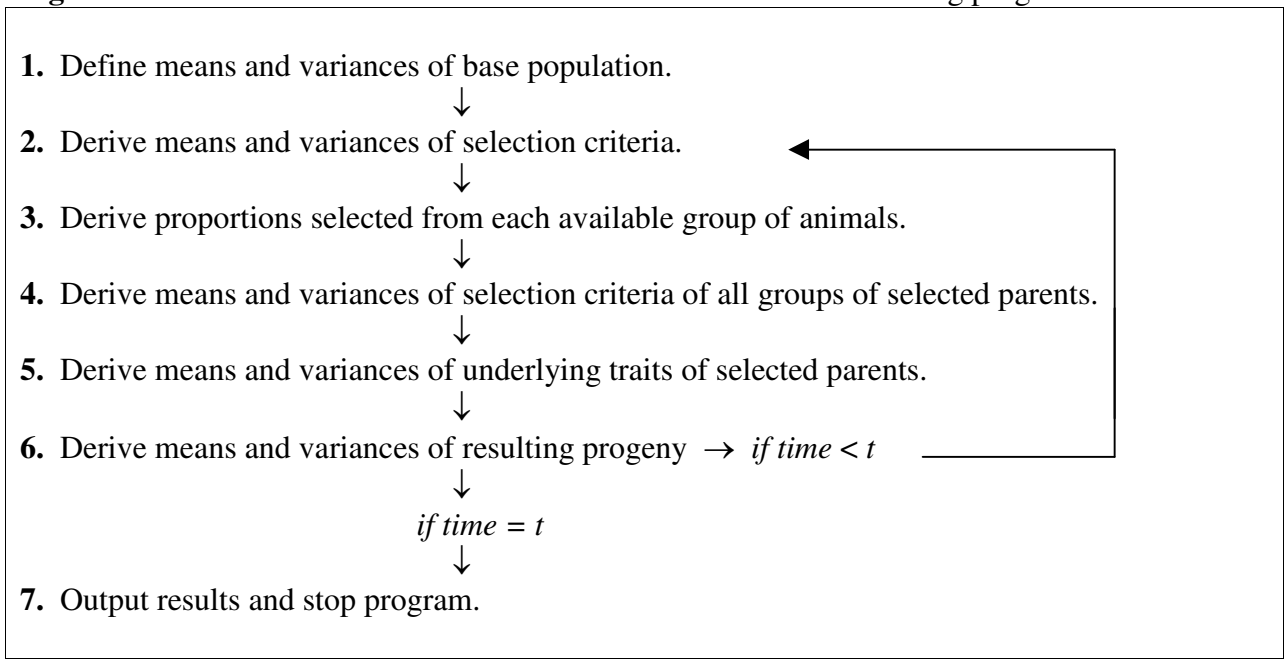
Recalling that heritability is $h^2 = \frac{\sigma_g^2}{\sigma_p^2}$, we get $\hat{S} = i\, h^2 \sigma_p$ (3.27)

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.



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 chose 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:

$$i = z/p \tag{3.28}$$

where *z* is the height of the normal distribution at the truncation point *x* and is given by

$$z = \frac{e^{-1/2x^2}}{\sqrt{2\pi}}$$ and π, 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:

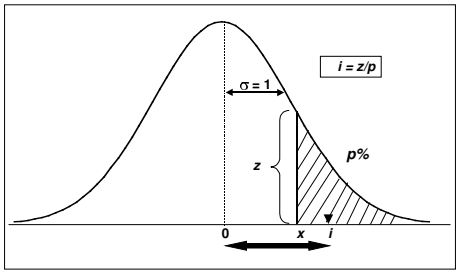
$$p^* = \frac{(s + {}^1/_2)}{n + \frac{s}{2n}} \tag{3.29}$$

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

$$i^* = \frac{z^*}{p^*} \tag{3.30}$$

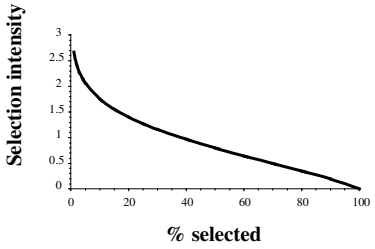
where z^* is the height of the normal distribution at the truncation point x^* corresponding to p^*.

Derivation of selection intensity from the standard Normal distribution



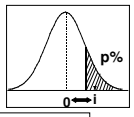
Effect of % Selected on Selection Intensity

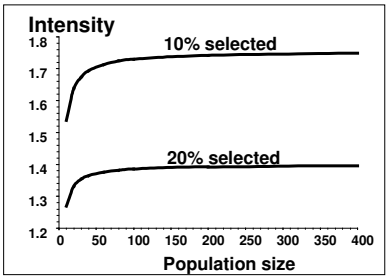
Large population size



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 ($\hat{g}_o = \frac{1}{2}\hat{g}_s + \frac{1}{2}\hat{g}_d$). 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:

$$i^* = \sqrt{1 - t_{av}}\ i \tag{3.31}$$

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:

$$t_{av} = t_{fs}\,\frac{n_w - 1}{n_w n_{fs} - 1} \tag{3.32}$$

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:

$$t_{av} = \frac{t_{fs}(n_w - 1) + t_{hs} n_w (n_{fs} - 1)}{n_w n_{fs} n_{hs} - 1} \tag{3.33}$$

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, of 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}, and w_{s3} ($\Sigma w_i = 1$). Fractions selected from each age group are p_{s1}, p_{s2}, and p_{s3}, for a total proportion selected of

$$P_s = p_{s1}\, w_{s1} + p_{s2}\, w_{s2} + p_{s3}\, w_{s3} \tag{3.34}$$

Let the genetic mean in age group i be denoted by $\bar{g}_{si}$ 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 σ_g. This assumption we be relaxed in later chapters.

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

$$\bar{g}^*_{si} = \bar{g}_{si} + S_{si} \tag{3.35}$$

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

$$\hat{S}_{si} = i_{si}\, r_{si}\, \sigma_g \tag{3.36}$$

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:

$$\begin{aligned}\bar{g}^*_s &= \frac{1}{P_s}\{p_{s1}\, w_{s1}\, \bar{g}^*_{s1} + p_{s2}\, w_{s2}\, \bar{g}^*_{s2} + p_{s3}\, w_{s3}\, \bar{g}^*_{s3}\} \\ &= \frac{1}{P_s}\sum p_{si}\, w_{si}\,(\bar{g}_{si}+S_{si})\end{aligned} \tag{3.37}$$

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

$$\bar{g}^*_d = \frac{1}{P_d}\sum p_{di}\, w_{di}\,(\bar{g}_{di}+S_{di}) \tag{3.38}$$

and the average genetic value of the progeny as

$$\begin{aligned}E(\bar{g}_o) &= \tfrac{1}{2}\bar{g}^*_s + \tfrac{1}{2}\bar{g}^*_d \\ &= \tfrac{1}{2}\frac{1}{P_s}\sum p_{si}\, w_{si}\,(\bar{g}_{si}+S_{si}) + \tfrac{1}{2}\frac{1}{P_d}\sum p_{di}\, w_{di}\,(\bar{g}_{di}+S_{di})\end{aligned} \tag{3.39}$$

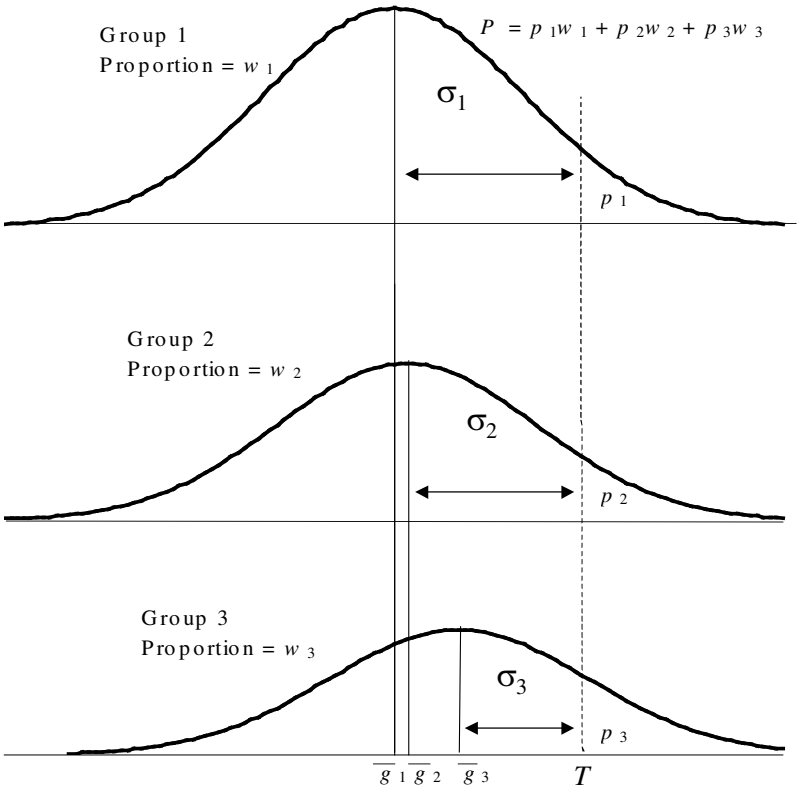
These equations allow for recursive prediction of the genetic mean of the population in successive time periods. In Chapter 8, we will formalize these recursive equations in the form of gene flow.

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.

Figure 3.2 Schematic representation of truncation selection of a total proportion P across multiple overlapping distributions



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_s. 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 ($T_i = \overline{g}_i + x_i\sigma_i$, where x_i is the standardized truncation point and σ_i the standard deviation of the i^{th} distribution ($\sigma_i = r_{si}\sigma_g$ for our case))

2. Choose the lowest T_i as a lower bound for $T \rightarrow T_l$
   Choose the highest T_i as a upper bound for $T \rightarrow T_u$. (T must lie between T_l and T_u.)

3. Compute the mean of the upper and lower bound $\rightarrow$ $T_m = ½ (T_u + T_l)$

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: $P_m = \Sigma p_i w_i$

6. If $|P_m - P| < \varepsilon$, where ε 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 <u>upper</u> bound $\rightarrow$ set $T_u = T_m$
   If $P_m > P$ then T_m becomes the new <u>lower</u> 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 $R = S = ir_{gI}\sigma_g$, response per year is obtained by dividing equation (3.22) by the generation interval, L, to get

$$R = \frac{ir_{g,I}\sigma_g}{L} \tag{3.40}$$

(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):

$$R = \frac{S_s + S_d}{L_s + L_d} \tag{3.41}$$

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

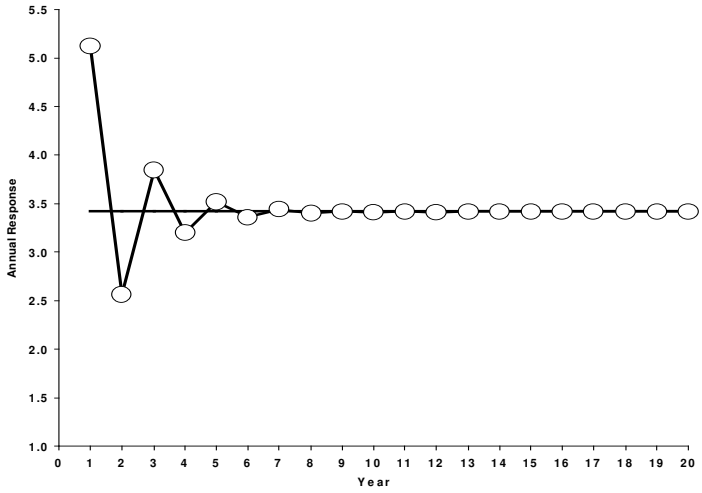
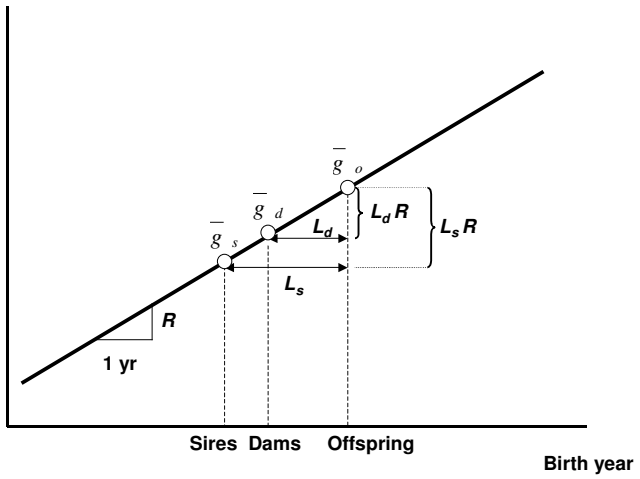


Figure 3.4 Asymptotic response to selection for breeding programs with overlapping generations



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):

$$\bar{g}_o = \tfrac{1}{2}\bar{g}^*_s + \tfrac{1}{2}\bar{g}^*_d = \tfrac{1}{2}(\bar{g}_s + S_s) + \tfrac{1}{2}(\bar{g}_d + S_d) \qquad (3.42)$$

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_sR 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:

$$\bar{g}_s = \bar{g}_o - L_sR$$

and similarly,

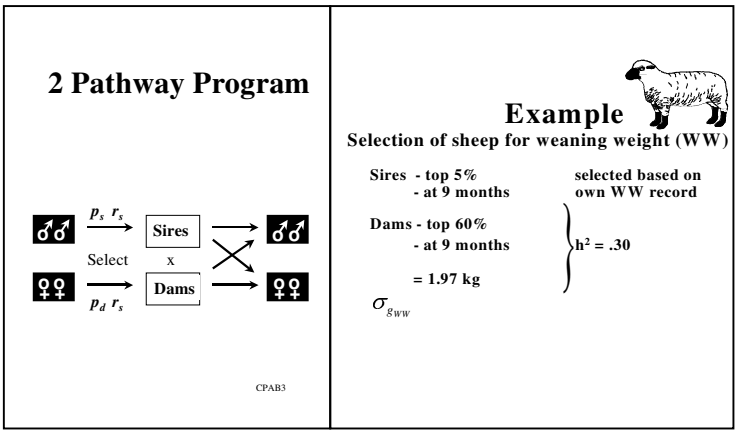
$$\bar{g}_d = \bar{g}_o - L_dR$$

Substituting into equation (3.42) we get:

$$\begin{aligned}\bar{g}_o &= \tfrac{1}{2}(\bar{g}_o - L_sR + S_s) + \tfrac{1}{2}(\bar{g}_o - L_dR + S_d)\\ &= \bar{g}_o - \tfrac{1}{2}R(L_s + L_d) + \tfrac{1}{2}(S_s + S_d)\end{aligned}$$

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.



Predicting Response in WW

Path	%	i	r = $\sqrt{h^2}$	Genetic Superiority	Gen. Interval
Sire	5	2.06	.55	2.23	1.17 yr
Dam	60	.64	.55	.69	1.17 yr
				2.92	2.34 yr

ΔG_{WW} = 2.92/2.34 = 1.25 kg/yr

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*).

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:

$$R = \frac{S_{sm} + S_{dm} + S_{sf} + S_{df}}{L_{sm} + L_{dm} + L_{sf} + L_{df}} = \frac{\sum_i S_i}{\sum_i L_i} \tag{3.42}$$

For each path, genetic superiorities can be derived as shown before as: $S_i = i_i r_i \sigma_g$

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:

$$S_s = \frac{1}{P_s} \{p_{s1}\, w_{s1}\, S_{s1} + p_{s2}\, w_{s2}\, S_{s2} + p_{s3}\, w_{s3}\, S_{s3}\} \tag{3.43}$$

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

$$L_s = \frac{1}{P_s} \{p_{s1}\, w_{s1}\, L_{s1} + p_{s2}\, w_{s2}\, L_{s2} + p_{s3}\, w_{s3}\, L_{s3}\} \tag{3.44}$$

Selection Across Age Groups

e.g. Selection of Bull Dams σ_g = 550 kg

Age Group	Age at Birth of Progeny	% of Bull Dams	% Selected	*i*	*r*	Genetic Superiority $ir\sigma_g$
Heifers	2 yr	50%	2.5%	2.34	.55	707.9
1st Lact.	3 yr	30%	1.5%	2.53	.68	946.2
2nd Lact.	4 yr	20%	1.5%	2.53	.72	1001.9

Selection Across Age Groups

- **Pooled Generation Interval**

  L_{dm} = 50% * 2
  + 30% * 3
  + 20% * 4

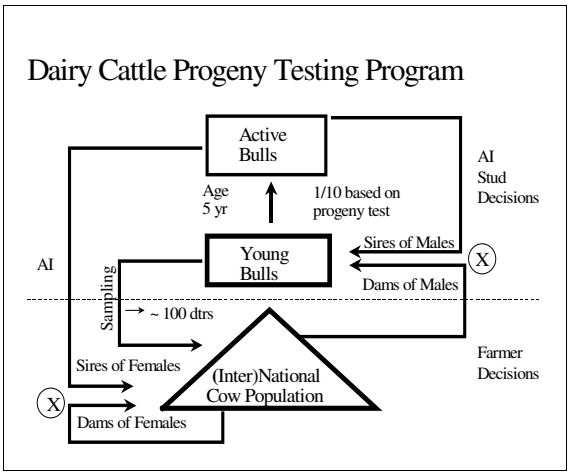
  = 2.7 yr

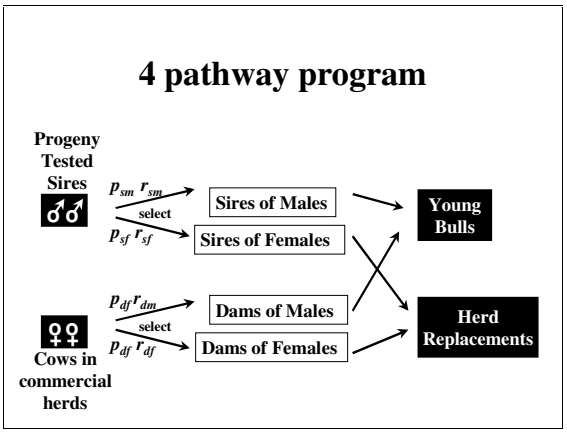
- **Pooled Genetic Superiority**

  S_{dm} = 50% * 707.9
  + 30% * 946.2
  + 20% * 1001.9

  = 838.2 kg

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 result 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.

Pathway	Proportion Selected (p_i)	Intensity (i_i)	Accuracy (r_i)	Genetic Superiority ($S_i = i\ r_i\sigma_g$)	Generation Interval (yr) (L_i)
Sires of males	2 %	2.42	0.90	2.178 σ_g	6
Sires of females	10 %	1.75	0.90	1.575 σ_g	7
Dams of males	0.5 %	2.89	0.60	1.743 σ_g	5
Dams of females	90 %	0.19	0.60	0.114 σ_g	6
TOTAL				$\Sigma S = 5.601\sigma_g$	$\Sigma L = 24$

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, σ_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 $\sigma_g = \sqrt{\sigma_g^2} = 0.5 \times 1080 = 540$ kg.

Hence $R = 0.233 \times 540 = 125.82$ kg per year

or, alternatively, $R = 125.82/6000 = 2.1\%$ 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 σ_g 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

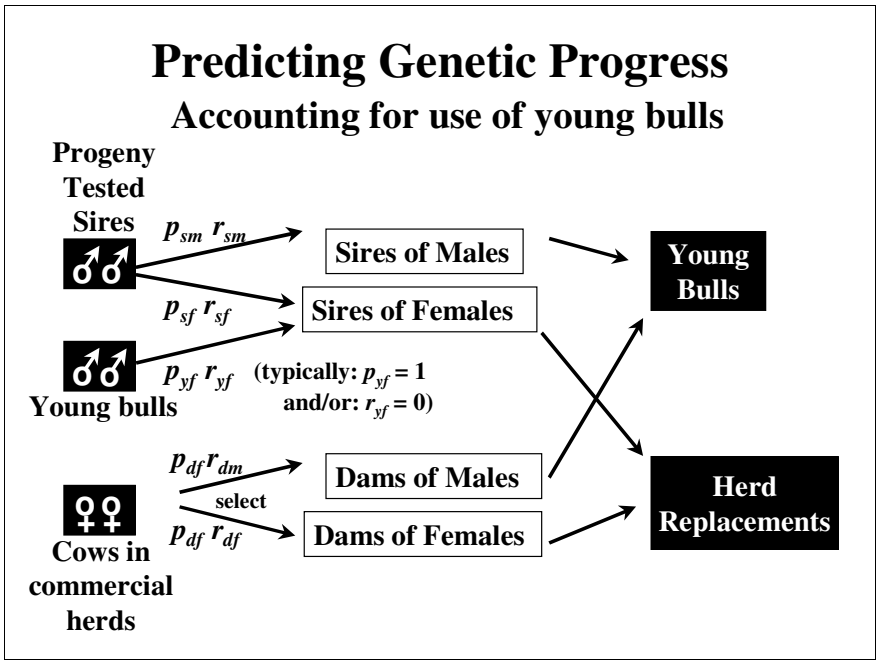
In the previous, the sire to female path only accounted for the use of progeny-tested sires to breed cows to produce herd replacements. However, young bulls also contribute to the next generation of females; in a practical breeding program, semen from young bulls can represent as much as 20% of all inseminations. 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:

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

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:

$$L_{sf} = y\, L_{yb,f} + (1\text{-}y)\, L_{pb,f}$$



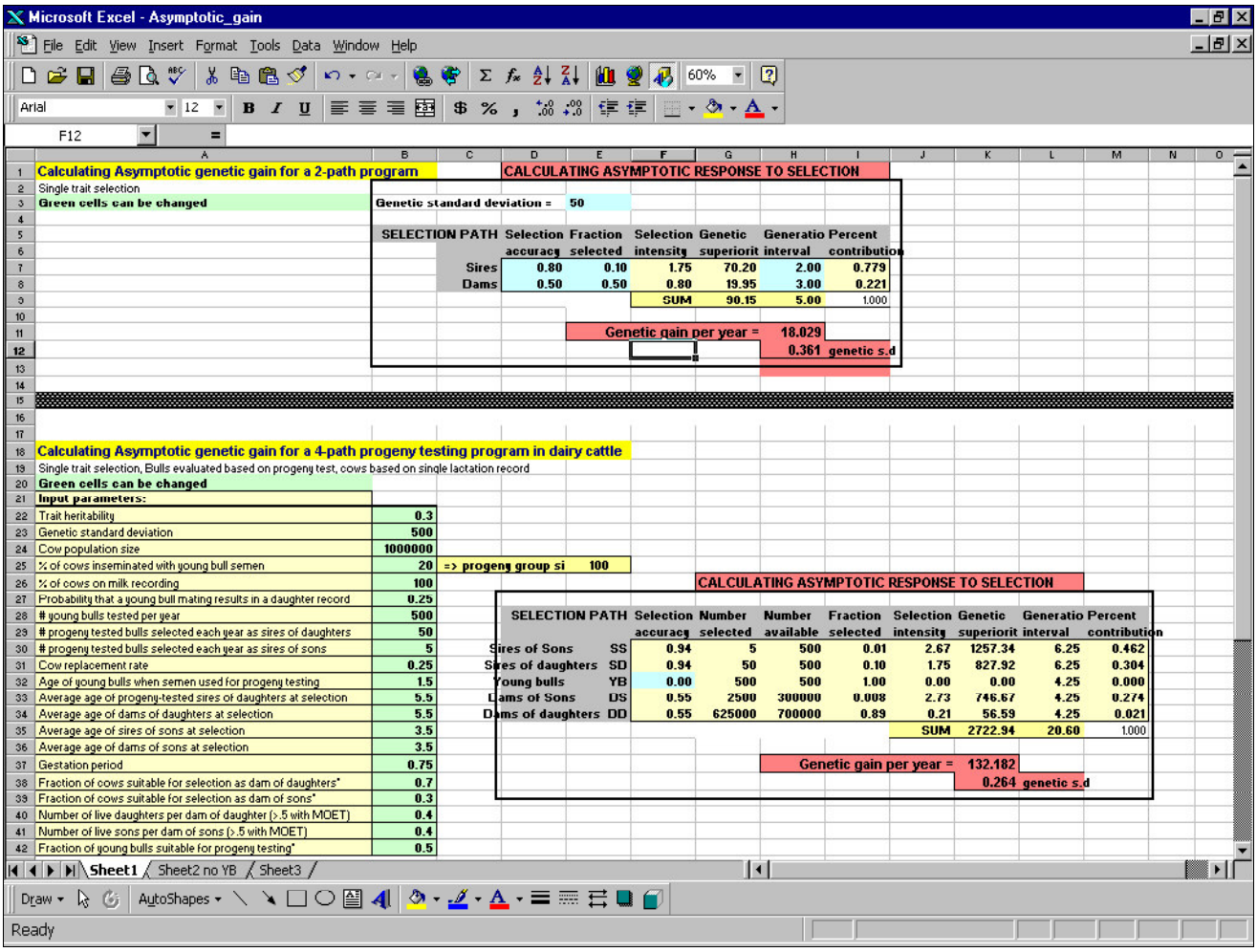
An example is given in Table 3.2, which assumes $y = 0.2$

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.

Pathway	Proportion Selected (p_i)	Intensity (i_i)	Accuracy (r_i)	Genetic Superiority ($S_i = i\, r_i \sigma_g$)		Generation Interval (yr) (L_i)	
Sires of males	2 %	2.42	0.90	$2.178\sigma_g$		6	
Sires of females - Young	100 %	0	0.50	0	$1.260\sigma_g$	2	6
Sires of females - Proven	10 %	1.75	0.90	1.575		7	
Dams of males	0.5 %	2.89	0.60	$1.734\sigma_g$		5	
Dams of females	90 %	0.19	0.60	$0.114\sigma_g$		6	
TOTAL				$\Sigma S = 5.268\sigma_g$		ΣL =23	

Now response per year becomes: $R = \frac{5.268}{23}\sigma_g = 0.230\ \sigma_g$ 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:

$$S_{2.1} = i\, r_{g_2 I_1} \sigma_{g_2} \tag{3.45}$$

Here $r_{g_2 I_1}$ 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: $r_{g_2 I_1} = r_{g_2 g_1} r_{g_1 I_1}$

Then:
$$S_{2.1} = i\, r_{g_2 g_1} r_{g_1 I_1}\, \sigma_{g_2} = r_{g_1 g_2} \frac{\sigma_{g_2}}{\sigma_{g_1}} i\, r_{g_1 I_1}\, \sigma_{g_1} = r_{g_1 g_2} \frac{\sigma_{g_2}}{\sigma_{g_1}} S_1 = b_{g_2 g_1} S_1 \qquad (3.46)$$

where $b_{g_2 g_1}$ 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:
$$R_{2.1} = b_{g_2 g_1} R_1 = = r_{g_1 g_2} \frac{\sigma_{g_2}}{\sigma_{g_1}} R_1 \qquad (3.47)$$

Where R_1 can be predicted using equation (3.41).

Predicting Correlated Response to Selection

- Select on trait 1
  - ↳ Response in trait 1
    - ⇨ Direct response to selection
  - ↳ Response in trait 2
    IFF $r_{g_{1,2}} \neq 0$
    - ⇨ Correlated response to selection
    - ⇨ Change in trait due to selection on other trait

Correlations between Traits

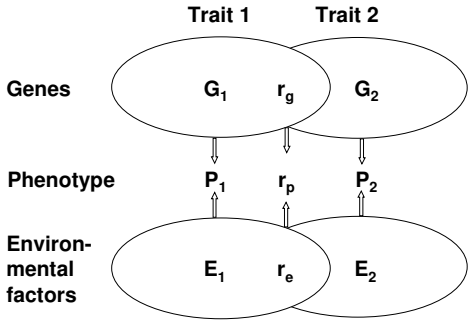
Phenotypic correlation (r_p) = correlation between phenotypes for traits 1 and 2 observed on the same individuals

Causes for existence of a phenotypic correlation:

1) Some genes can have effects on both traits = Pleiotropy
⇨ **genetic correlation (r_g)**

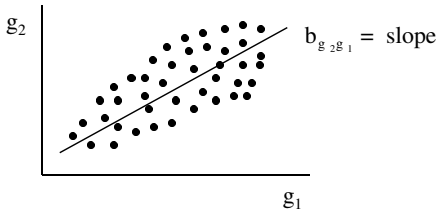
2) Some environmental factors can affect both traits
⇨ **environmental correlation (r_e)**

Correlations between Traits



Correlated Response to Selection

Regression of BV for trait 2 on BV for trait 1



Example

Selection of sheep on weaning weight

- Direct response = ΔG_{WW} = 1.25 kg/yr
- Correlated response in birth weight?

$\sigma_{g_{BW}} = .5$ kg
$r_{g_{WW,BW}} = +.3$
$\Delta G_{BW.WW} = b_{A_{BW}A_{WW}} \Delta G_{WW}$

Predicting Response in W W

Path	%	i	$r = \sqrt{h^2}$	Genetic Superiority	Gen. Interval
Sire	5	2.06	.55	2.23	1.17 yr
Dam	60	.64	.55	.69	1.17 yr
				2.92	2.34 yr

$\Delta G_{WW} = 2.92/2.34 = 1.25$ kg/yr

Prediction of Correlated Response

$$b_{A_{BW}.A_{WW}} = r_g \frac{\sigma_{g_{BW}}}{\sigma_{g_{WW}}} = (.3)\ \frac{.5}{1.97} = .076 \text{ kg/kg}$$

$$\Delta G_{BW.WW} = (.076)(1.25) = .095 \text{ kg/yr}$$

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 ♣L↓
  - Correlated trait has higher h^2.

Efficiency of Indirect vs Direct Selection

1 = correlated trait
2 = economic trait

- Direct selection:

$$\Delta G_2 = \frac{\Sigma i r_{\hat{A}_2} \sigma_{g_2}}{\Sigma L}$$

Efficiency of Indirect vs Direct Selection (cont'd)

- Indirect selection: (correlated response in trait 2 to selection on trait 1)

$$\Delta G_{2.1} = r_{g_{1,2}} \frac{\sigma_{g_2}}{\sigma_{g_1}} (\Delta G_1)$$

$$\uparrow \quad \Delta G_1 = \frac{i r_{\hat{A}_1} \sigma_{g_1}}{\Sigma L}$$

$$\text{Efficiency} = \frac{\Delta G_{2.1}}{\Delta G_2}$$

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 $R = \frac{(2.42 \text{ x } 0.9 + 1.75 \text{ x } 0.9 + 2.89 \text{ x } 0.64 + 0.19 \text{ x } 0.6}{6 + 7 + 6 + 6} \sigma_g = 0.229 \ \sigma_g$ per year

which is less than the predicted response of 0.233 σ_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.

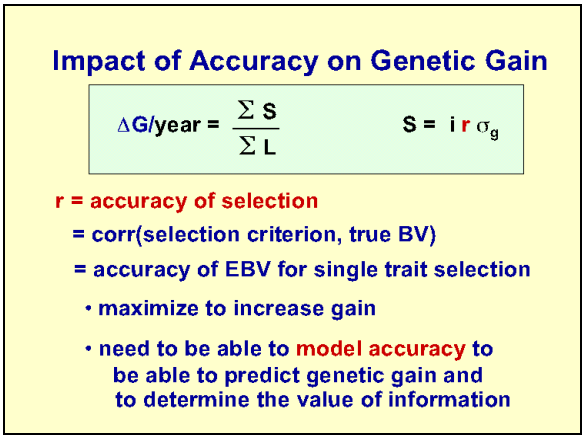
Chapter 4

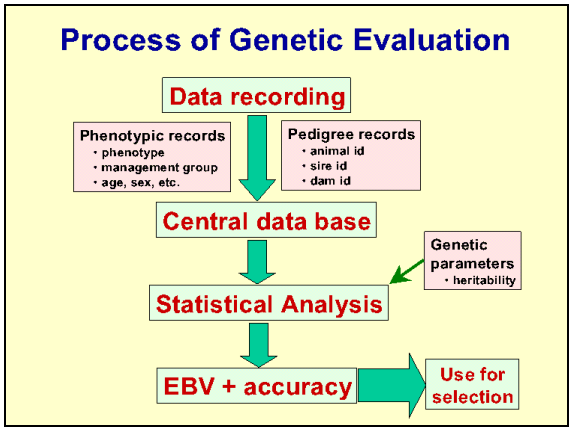
Deterministic Models for Estimated Breeding Values

The previous chapter established the main factors that affect response to selection, i.e. intensity of selection (i), accuracy of selection (r), genetic standard deviation (σ_g), and generation interval (L). The objective of this chapter is to develop methods to model and evaluate accuracy of selection, and to evaluate the main factors that determine this parameter. The latter will help us with the design of breeding programs.

Accuracy of selection is defined as the correlation between the criterion on which selection is based (I) and the objective of selection. For the moment, we will consider the breeding value of a single trait to be the selection objective but this will be extended to more complicated economic selection objectives in Chapter 6.

The previous chapter showed that when selection is on the individual's own phenotype, the accuracy of selection is equal to the correlation between phenotype and breeding value, which is equal to the square root of heritability (h). In practical animal breeding, selection is often not solely on own phenotype but on estimates of breeding values (EBV) that are derived from records on the animal itself and records on its relatives using Best Linear Unbiased Prediction (BLUP) for an animal model (Lynch and Walsh, 1998). An important property of EBV derived from an animal model is that all records that are available on the individual and its relatives are optimally used, while simultaneously adjusting for systematic environmental effects (e.g. herd-year-season), such that the accuracy of the EBV is maximized. Given the equation for predicting genetic superiority of selected animals, i.e. $S = ir\sigma_g$, it is clear that maximizing accuracy is crucial to maximizing genetic gain.





Stochastic simulation models of breeding programs can directly incorporate genetic evaluations based on animal models because the data that provide the input for such models are individually

simulated. This is not possible for deterministic models. Thus, when developing deterministic models for genetic improvement, other methods to model selection and accuracy of EBV from BLUP animal models must be used. In addition to allowing deterministic modeling of selection on EBV, these methods are also required to develop a basic understanding of factors that affect accuracy of selection, which are important for the design of breeding programs, including the contribution that different types of records make to accuracy of EBV.

In our development of methods to model accuracy of EBV, we will slowly build our methodology up using the following steps:

1. EBV from own records – simple regression
2. EBV from records on a single type of relatives – simple regression
3. EBV from multiple sources of information – multiple regression – selection index theory
4. EBV from BLUP animal models (module B)

As noted above, the common theme through these methods is the use of linear regression for the prediction of EBV from phenotypic records.

Before going into these developments, we will first describe some general properties of EBV. These properties hold regardless which of the methods listed above is used to estimate the EBV, provided the model used for evaluation is correct and systematic environmental factors are properly accounted for.

4.1 Some general properties of EBV

As indicated above, all methods for prediction of breeding values are based on the principles of linear regression: ***regression of breeding values on phenotypic records***. As a result, properties of linear regression can be used to derive general properties of EBV.

One important property of EBV is ***unbiasedness***. This means that the *expected* magnitude of the *true* breeding value of an animal is equal to its *estimated* breeding value:

$$E(g_i|\hat{g}_i) = \hat{g}_i$$

This implies that selection on $\hat{g}$ will maximize the expected value of g for the group of selected individuals. A related property is that the regression of true on estimated breeding values is equal to 1:

$$b_{g,\hat{g}} = 1$$

Given unbiasedness, the accuracy of EBV can be derived as the correlation between true and estimated BV as:

$$r = r_{g,\hat{g}} = b_{g,\hat{g}} \frac{\sigma_{\hat{g}}}{\sigma_g} = \frac{\sigma_{\hat{g}}}{\sigma_g} \qquad (4.1)$$

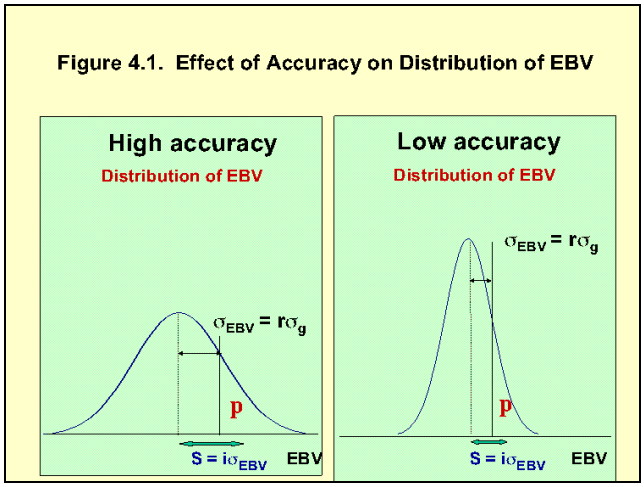
and the covariance between true and estimated BV as:

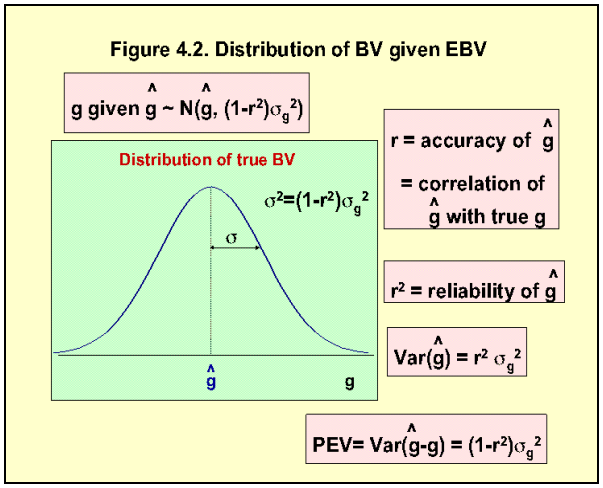
$$\sigma_{g,\hat{g}} = r_{g,\hat{g}} \ \sigma_g \ \sigma_{\hat{g}} = \sigma^2_{\hat{g}} \qquad (4.2)$$

The variance of EBV is then equal to:

$$\sigma^2_{\hat{g}} = r^2 \sigma^2_g \qquad (4.3)$$

Thus, the variance of EBV is equal to the square of accuracy (also referred to as 'reliability') multiplied by genetic variance. This shows the importance of accuracy: the larger the accuracy, the larger the variance and spread of EBV of animals in the population, the better we will able to distinguish between genetically superior and average or inferior animals, and the greater the genetic superiority of selected animals will be. This is illustrated in Figure 4.1.





Like any prediction, EBV also have a ***prediction error***, which is the deviation of true BV from the EBV: $\varepsilon_i = g_i - \hat{g}_i$

The variance of prediction errors (***prediction error variance***, PEV) can be derived as:

$$\sigma_\varepsilon^2 = \text{var}(g_i - \hat{g}_i) = \sigma_{\hat{g}}^2 + \sigma_g^2 - 2\sigma_{g,\hat{g}} = \sigma_{\hat{g}}^2 + \sigma_g^2 - 2\sigma_{\hat{g}}^2$$

$$= \sigma_g^2 - \sigma_{\hat{g}}^2 = \sigma_g^2 - r^2\sigma_g^2$$

$$= (1-r^2)\sigma_g^2 \qquad (4.4)$$

Note that $\sigma_g^2 = \sigma_{\hat{g}}^2 + \sigma_\varepsilon^2$

Thus, additive genetic variance is partitioned into variance that is explained by the EBV and unexplained error variance. The higher the accuracy is, the greater the proportion of genetic variance that is explained by the EBV. Also note that the covariance between EBV and prediction errors is equal to zero: $\sigma_{\hat{g},\varepsilon} = \sigma_{\hat{g},\hat{g}-g} = \sigma_{\hat{g}}^2 - \sigma_{g,\hat{g}} = \sigma_{\hat{g}}^2 - \sigma_{\hat{g}}^2 = 0$

This makes sense because a non-zero covariance would imply that the prediction error contains some information that can be used to improve the EBV.

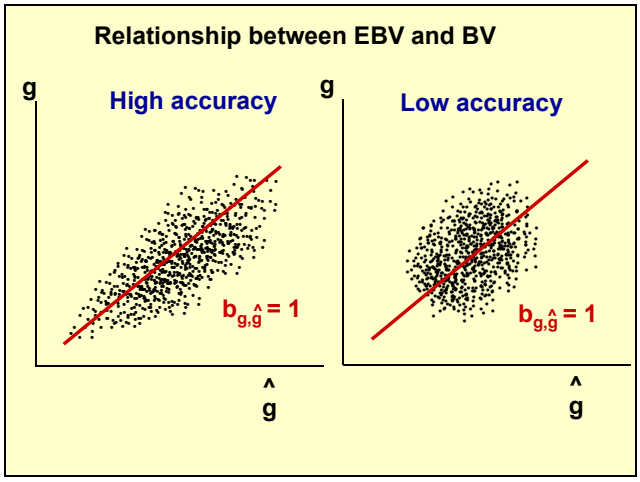
Given an animal's EBV and assuming normality, the animal's *true* BV is expected to follow a Normal distribution with mean equal to the EBV and variance equal to $(1-r^2)\sigma_g^2$:

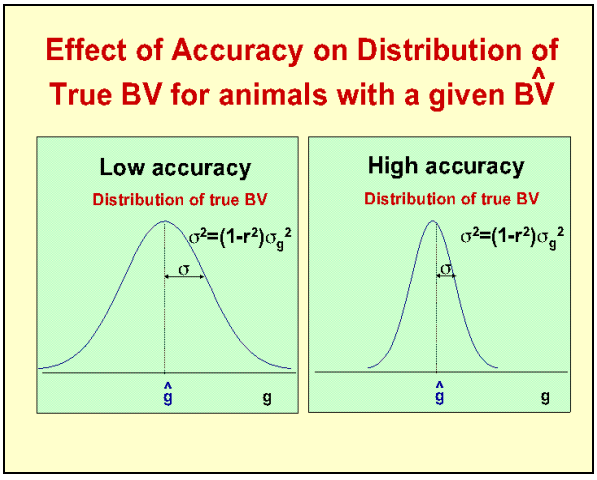
$$g_i | \hat{g}_i \sim N(\hat{g}_i, (1-r^2)\sigma_g^2) \qquad (4.5)$$

This distribution is illustrated in Figure 4.2.

Prediction errors are expected to follow a Normal distribution with mean zero:

$$\varepsilon_i \sim N(0,(1\text{-}r^2)\sigma_g^2) \qquad (4.6)$$





4.2 EBV from own records

In the derivations below, we will assume that phenotypic records, x_i, are adjusted for systematic environmental effects and deviated from the mean.

4.2.1 Phenotypic Selection

The simplest form of selection is based on EBV derived from a single record of the phenotype of the individual itself. In this case, the EBV can be derived from regression of BV on phenotype as:

$$\hat{g}_i = b_{g,x}\, x_i = b_{g,x}\,(\text{phenotype of individual}) \qquad (4.7)$$

The regression coefficient can be derived as:

$$b_{g,x} = \sigma_{x_i g_i}/\sigma_p^2 = \sigma_{g_i+e_i g_i}/\sigma_p^2 = \sigma_g^2/\sigma_p^2 = h^2 \qquad (4.8)$$

Thus the prediction of an individual's additive genetic value, expressed as a deviation from the population mean, is given by

$$\hat{g}_i = h^2 x_i \qquad (4.9)$$

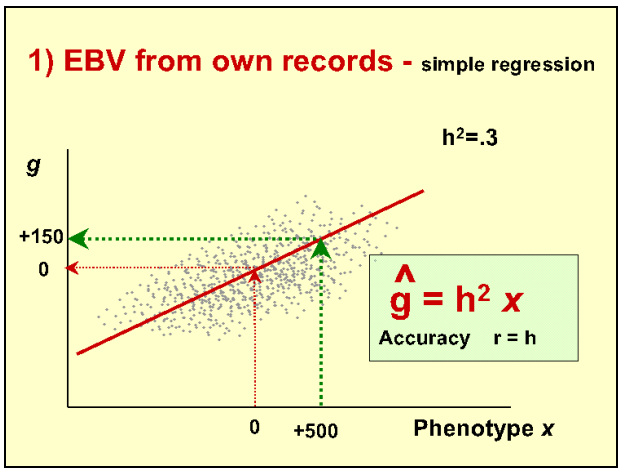
where x_i is the phenotype of individual i expressed as a deviation from the population mean.

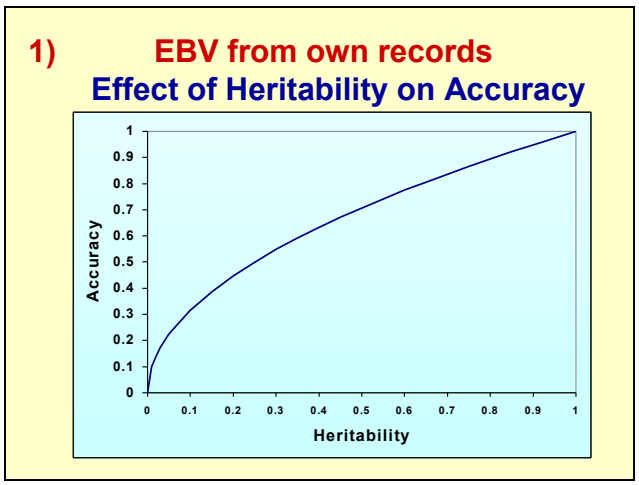
The accuracy of selection is: $r = r_{g,\hat{g}} = \sigma_{g_i,h2x_i}/\sigma_g\sigma_{h2x} = h^2\sigma_g^2/h\sigma_g^2 = h$ (4.10)

As an example, growth rate in pigs and cattle often has a heritability of around 0.5. Thus with phenotypic selection for growth rate, the EBV of individual i is: $\hat{g}_i = 0.5\, x_i$

and the accuracy of evaluation is: $r = \sqrt{0.5} = 0.707.$

Alternatively, if we were selecting on a single record for milk yield in cows with a heritability of 0.25, our EBV would be $\hat{g}_i = 0.25\, x_i$ and accuracy would be $r = 0.5$





4.2.2 Selection on the Mean of Two or more Phenotypic Records on a Single Trait

Definition of Repeatability

We can increase accuracy of selection by increasing the number of records collected on each individual. This can be done for traits that are expressed several times during the lifetime of an animal. For example, having two lactation records on a cow should give more information than having only one lactation. For traits with repeated observations, such as milk production, the environmental and/or non-additive genetic component of the phenotype can then be separated in a permanent component that affects the animal for its lifetime and a temporary component, which changes over time. Thus the phenotype for record j on animal i can be written as:

$$x_{ij} = g_i + pe_i + te_{ij} \qquad (4.11)$$

where pe_i is a permanent environment effect specific to animal i and te_{ij} a temporary environment effect that is specific to record j on animal i. The genetic and permanent environment effects are the same for all observations on the same individual. On the other hand, the temporary environment effects for different observations on the same individual are uncorrelated. This implies that all observations on the same individual are genetically the same trait. This leads to the concept of repeatability. Repeatability, t, is defined as the proportion of the total phenotypic variance which is due to permanent effects (environment and genetic)

associated with each animal. Thus, assuming no correlations between the genetic, permanent environment, and temporary environment effects, affecting a single observation,

$$t = \frac{\sigma_g^2 + \sigma_{pe}^2}{\sigma_p^2} \quad \text{or} \quad \frac{\sigma_g^2 + \sigma_{pe}^2}{\sigma_g^2 + \sigma_{pe}^2 + \sigma_{te}^2} \tag{4.12}$$

Imagine that a cow, i, has two lactation records, x_{i1} and x_{i2}, which can be denoted as

$$x_{i1} = g_i + pe_i + te_{i1}$$
$$x_{i2} = g_i + pe_i + te_{i2}$$

The correlation between two records on an individual is $r_{x_1x_2} = \dfrac{\sigma_{x_1x_2}}{\sqrt{\sigma_{x_1}^2 \sigma_{x_2}^2}}$

where
$$\sigma_{x_1x_2} = \sigma_{(g_i + pe_i + te_{i1},\, g_i + pe_i + te_{i2})}$$
$$= \sigma_g^2 + \sigma_{pe}^2$$

Hence, $$r_{x_1x_2} = \frac{\sigma_g^2 + \sigma_{pe}^2}{\sigma_p^2} = t$$

Thus, the repeatability of a trait is also the correlation between two records for that trait on the same individual; literally a measure of how "repeatable" that trait is over several records.

EBV from Repeated Records on a Single Trait

Imagine a situation where m records are collected on each individual and we wish to select on the mean of those m records. Then,

$$\hat{g}_i = b_{g\bar{x}}\, \bar{x}_i \tag{4.13}$$

where
$$\bar{x}_i = \sum_{j=1}^{m} x_{ij} / m \tag{4.14}$$

and x_{ij} is the j^{th} record for the chosen trait on individual i. Thus

$$\bar{x}_i = \sum_{j=1}^{m} (g_i + pe_i + te_{ij}) / m \tag{4.15}$$

Then, $$b_{g\bar{x}} = \sigma_{g\bar{x}} / \sigma_{\bar{x}}^2$$

The variance of $\bar{x}_i$ is: $$\sigma_{\bar{x}}^2 = \sigma_g^2 + \sigma_{pe}^2 + \frac{\sigma_{te}^2}{m} = t\sigma_p^2 + \frac{(1-t)\sigma_p^2}{m} = \frac{(mt + 1 - t)\sigma_p^2}{m}$$

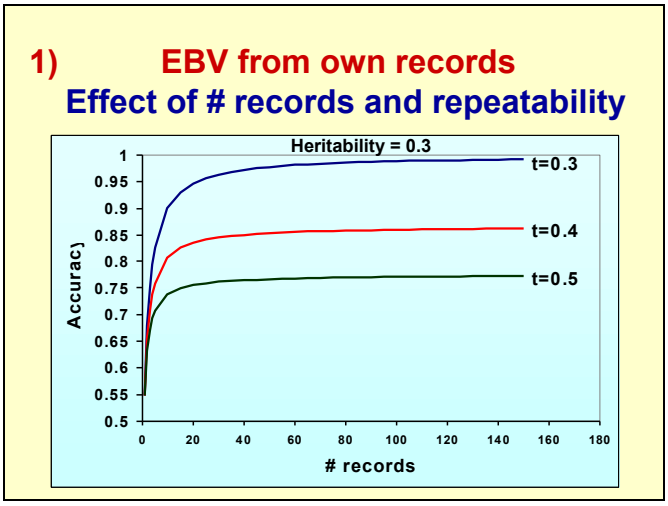
$$= \frac{((m-1)t + 1)\sigma_p^2}{m} \tag{4.16}$$

The covariance is: $$\sigma_{g\bar{x}} = \sigma_g^2 \tag{4.17}$$

Thus,
$$b_{g\bar{x}} = \frac{m\sigma_g^2}{\sigma_p^2\left((m-1)t+1\right)} = \frac{mh^2}{(m-1)t+1} \qquad (4.18)$$

And accuracy of selection is given by: $r = corr_{g\bar{x}} = \sqrt{\frac{mh^2}{(m-1)t+1}\frac{\sigma_g^2}{\sigma_g^2}} = \sqrt{\frac{mh^2}{(m-1)t+1}}$ (4.19)

Note that when t=1 there is no value in recording a trait more than once on an individual. Repeated measurements only add additional information when they allow separation of temporary and permanent effects acting on an observation.



Numerical Example of EBV Based on the Mean of Two or More Phenotypic Records

Consider selection for milk yield with a heritability of 0.25 and a repeatability of 0.5. Assume the observation is the mean of 1, 2, 5 or 10 lactation records. Substituting $h^2 = 0.25$, $t = 0.5$ and $m = 1, 2, 5$ or 10 into (4.18) and (4.19) we obtain regression coefficients of

$$b_{g\bar{x}} = 0.25, 0.333, 0.42 \text{ or } 0.45$$

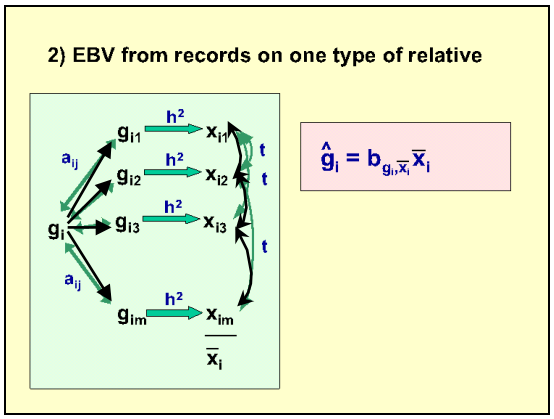
and accuracies of

$$r = 0.5\ , 0.58\ , 0.65\ \text{ or } 0.67.$$

4.3 EBV from One Type of Relatives' Records

The simple regression methods for estimation of BV described in the previous section for own records can be extended to one or more records on a single type of relatives.

Imagine a situation where 1 record is collected on each of m relatives of individual i for which we want to estimate the breeding value. Each relative j has the same additive genetic relationship a_{ij} with individual j. Also, the relatives have the same additive genetic relationship to each other, $a_{jj'}$.



Then, the BV of individual *i* can be predicted from the average of the records of its relatives based on:

$$\hat{g}_i = b_{g\bar{x}}\,\bar{x}_i$$

where

$$\bar{x}_i = \sum_{j=1}^{m} x_{ij} / m$$

and x_{ij} is the record on the j^{th} relative of *i*.

Then,

$$b_{g\bar{x}} = \sigma_{g\bar{x}} / \sigma^2_{\bar{x}}$$

To derive $\sigma_{g\bar{x}}$, let *t* be the (intra-class) correlation between phenotypic records on relatives *j* and *j'*:

$$\begin{aligned} t = r_{x_{ij}x_{ij'}} &= \sigma_{x_{ij}x_{ij'}} / \sigma^2_p = \sigma_{(g_{ij}+e_{ij},\, g_{ij'}+e_{ij'})} / \sigma^2_p \\ &= (a_{jj'}\sigma^2_g + c^2\sigma^2_p)/\sigma^2_p \\ &= a_{jj'}h^2 + c^2 \end{aligned} \qquad (4.20)$$

Here c^2 is the ***common environment correlation*** between records. This parameter quantifies the extent to which relatives are exposed to the same environment (e.g. litter mates):

$$c^2 = \sigma_{e_{ij}e_{ij'}} / \sigma^2_p \qquad (4.21)$$

As an aside, note that this equation for the intra-class correlation also holds for repeated own records. In that case, $a_{jj'}=1$, $c^2 = \sigma^2_{pe}/\sigma^2_p$, and thus $t = h^2 + \sigma^2_{pe}/\sigma^2_p = (\sigma^2_g + \sigma^2_{pe})/\sigma^2_p$, which is equal to repeatability (see equation 4.12).

The variance of the mean of *m* records with intra-class correlation *t* can be derived as:

$$\sigma^2_{\bar{x}} = \mathrm{Var}\left(\sum_{j=1}^{m} x_{ij} / m\right) = \frac{m\sigma^2_p + m(m-1)t\sigma^2_p}{m^2} = \frac{1+(m-1)t}{m}\sigma^2_p \qquad (4.22)$$

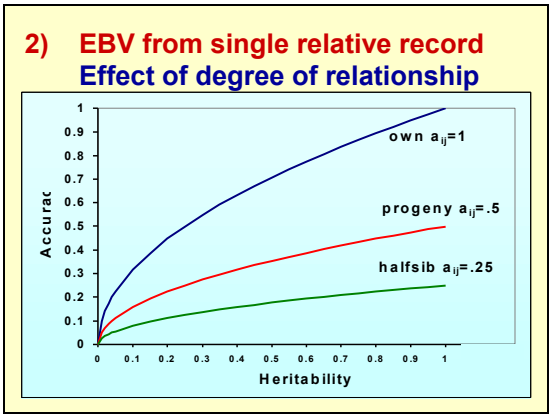
The covariance is:

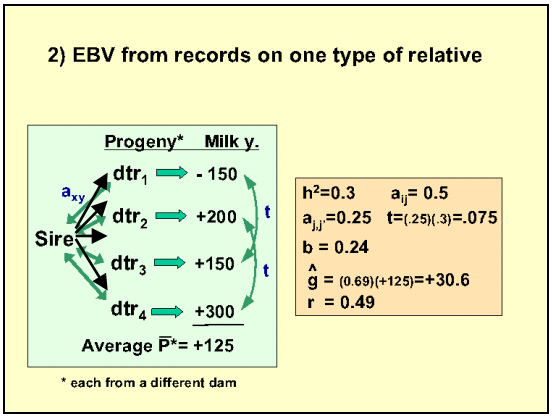
$$\sigma_{g\bar{x}} = a_{ij}\sigma^2_g \qquad (4.23)$$

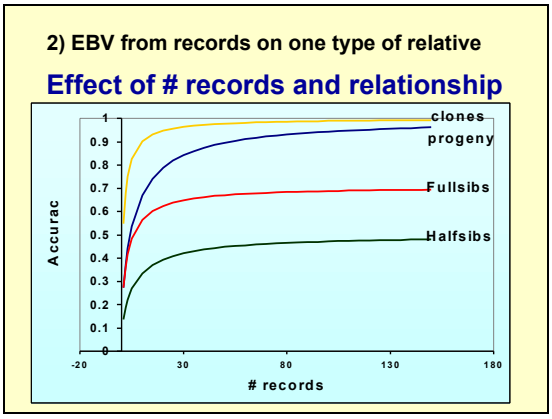
Thus,
$$b_{g\bar{x}} = \frac{ma_{ij}\sigma_g^2}{\sigma_p^2\left((m-1)t+1\right)} = a_{ij}\frac{mh^2}{(m-1)t+1} \qquad (4.24)$$

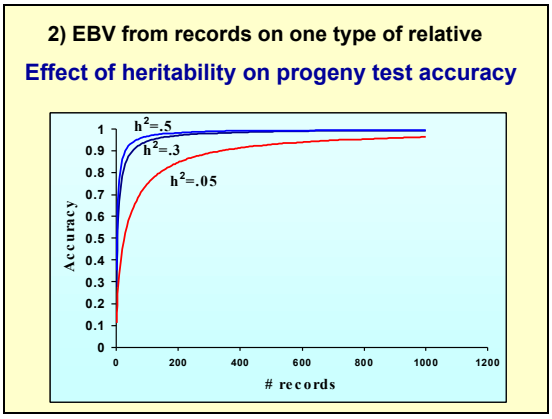
And, accuracy of selection is given by,
$$r = corr_{g\bar{x}} = a_{ij}\sqrt{\frac{mh^2}{(m-1)t+1}} \qquad (4.25)$$

Note that for repeated own records a_{ij}=1 and equations (4.24) and (4.25) simplify to equation (4.18) and (4.19).









4.4 EBV from Multiple Sources - Selection Index

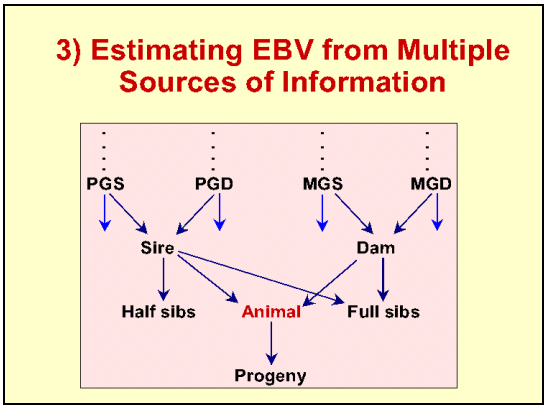
When records are available from multiple sources, e.g. records on the animal itself, its dam, half-sibs, progeny, etc., it will obviously be most beneficial to use all records to estimate the breeding value. This can be achieved by extending the simple regression methods described in the previous to a multiple regression setting:

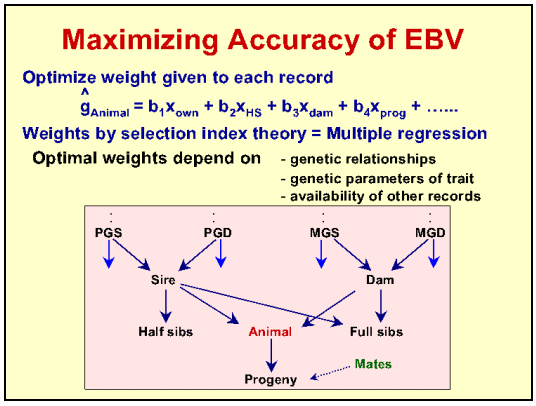
$$\hat{g}_i = b_1x_1 + b_2x_2 + \ldots + b_mx_m \qquad (4.26)$$

where x_i represents the i^{th} source of records, which could be an individual record or the mean of records on a given type of relative, and b_i are partial regression coefficients. Equation (4.26) is called a ***selection index*** and the coefficients b_i are called ***index weights***. The methodology that is used to derive the optimal index weights, i.e. those that maximize the accuracy of the EBV, is called selection index theory.

The selection index was first proposed by Smith (1936) for use in plant breeding for simultaneous selection on multiple traits, and seven years later, but apparently independently, by Hazel (1943) for animal breeding. In this Chapter we shall first discuss the basic problem, then go on to derive selection index equations, and then illustrate their use with some examples.

Selection index theory deals with the general problem of combining information from a variety of sources in such a way that the most accurate predictor of the overall genetic merit for a pre-defined combination of traits is obtained. Two separate types of selection indexes can be distinguished: 1) the ***economic selection index***, where information from several recorded traits is used to predict genetic merit for overall economic value, and 2) the ***family selection index***, where information from a single trait on various relatives is combined to predict the genetic merit of an individual for that trait.





The economic selection index and family selection index are special cases of the general selection index, where the selection index is defined as a linear function of a series of observations which when selected upon maximizes response of an aggregate genotype, which is a linear function of the additive genetic values of a defined set of traits. Although the focus in this Chapter is prediction of breeding values for a single trait, we will develop the theory of selection indexes within the context of the economic index because it is more general. We will then discuss the family index as a special case of the economic selection index and go into more detail into family indexes and their extension to modeling BLUP EBV. We will come back to various applications related to economic indexes in Chapter 5.

4.4.1 Selection Index theory

In economically oriented breeding programs, the trait that we want to improve could be called economic merit. The ***breeding objective*** of our program is then to maximize improvement of

economic merit. Economic merit might be defined in different ways, e.g. as profit per animal, profit per enterprise, economic efficiency, or something else. We will return to this problem in later Chapters. For the present, it is only necessary to recognize that the **breeding objective** is a general statement of the economic genetic goal of the breeding program.

For a given definition of the breeding objective, there will likely be several or many traits, which would contribute to the objective. The ***aggregate genotype*** is then defined as a function of the additive genetic values of the traits of interest of an individual, which if selected upon would achieve the breeding objective. The function need not necessarily be linear, but in many cases an approximate linear relationship can be found that adequately defines aggregate genotype over the range of genetic values encountered (see later chapters). If the function is a linear function, then the **aggregate genotype**, H, can be written as

$$H = v_1 g_1 + v_2 g_2 + \ldots + v_n g_n \quad = \mathbf{v'g} \tag{4.27}$$

where g_i is the additive genetic value of trait i, expressed as a deviation from the population mean, and v_i is a weighting factor (usually, but not necessarily, an economic weight) for trait i. In vector notation, $\mathbf{v'} = [\, v_1 \,,\, v_2 \,,\, \ldots \,,\, v_n]$ and $\mathbf{g'} = [\, g_1 \,,\, g_2 \,,\, \ldots \,,\, g_n]$.

In practice, the additive genetic values (i.e. true BV) of the various traits for an individual are not known. However we can record each individual's performance for a number of traits. The observations on these traits can then be combined into a ***selection index***, I of the form,

$$I = b_1 x_1 + b_2 x_2 + \ldots + b_m x_m \quad = \mathbf{b'x} \tag{4.28}$$

where x_j is the j^{th} phenotypic observation, as a deviation from the population mean, and b_i is a selection index coefficient (weight) for that observation. In vector notation, $\mathbf{b'} = [\, b_1 \,,\, b_2 \,, \ldots \,,\, b_m]$ and $\mathbf{x'} = [\, x_1 \,,\, x_2 \,,\, \ldots \,,\, x_m]$. In principle, observations x_j do not necessarily have to be on the traits that are in the aggregate genotype or on the animal that is being evaluated; observations can be on any trait and from the animal itself or its relatives.

The problem is then to estimate the selection index weights, b_i, such that selection of individuals on their **selection index** value, I, maximizes response in the **aggregate genotype**, H. Equivalently, we want to find b_i such that the correlation between I and H is maximized, or that the variance of prediction errors (Var(H-I)) is minimized.

With family selection indexes, the problem is to combine information from different types of relatives to provide the most accurate estimate of the additive genetic value of a given trait (g) for a given individual. In this case, the aggregate genotype is given by $H = g$ and, thus $\mathbf{v} = [1]$. In this case the selection index is equal to the EBV for the trait evaluated:

$$I = \hat{g} = b_1 x_1 + b_2 x_2 + \ldots + b_m x_m \tag{4.29}$$

Similar to an economic index, a family index can include information on the animal itself and its relatives for the trait being evaluated, as well as records on other traits. Thus, the derivations that follow for an economic index also apply to family indexes by setting $H = g$ and $\mathbf{v} = [1]$.

4.4.1.1 Derivation of index coefficients

We wish to define I such that selection of animals on I maximizes response in H. From standard regression theory (see also Chapter 3) expected response (genetic superiority) of selected individuals in H, S_H, is given by

$$S_H = b_{H.I}(I - \bar{I}) \tag{4.30}$$

where b is the regression of aggregate genotype on index values, I is the index value of the selected animal or group of animals, and $\bar{I}$ is the mean index value of all selection candidates. Since $I - \bar{I}$ can be written as $i\sigma_I$, where i is the intensity of selection (see Chapter 3),

$$S_H = b_{HI}\, i\sigma_I = \frac{\sigma_{HI}}{\sigma_I^2} i\sigma_I = i\sigma_{HI}/\sigma_I \tag{4.31}$$

Thus for any given intensity of selection, i, response in H is maximized when σ_{HI}/σ_I is maximized.

Apart from maximizing response in H to selection on I, it would also be useful if the index value, I, was an unbiased predictor of the aggregate genotypic value H. This means that the true aggregate genotype of an individual is, on average, no more likely to be greater than its index value than it is to be less than its index value, or

$$E(H - \bar{H}) = I - \bar{I} \tag{4.32}$$

Under the assumption of multivariate normality, this is achieved when the regression of H on I, $b_{HI} = 1$. Thus we wish to find the index coefficients $b_1, b_2 \ldots b_n$ that maximize σ_{HI}/σ_I, subject to $b_{HI} = 1$.

Considering first the maximization of σ_{HI}/σ_I. Let $\sigma_{g_{ki}}$ be the genetic covariance between the k^{th} observation in the index and the i^{th} trait in the aggregate genotype. Similarly, let $\sigma_{p_{ki}}$ be the phenotypic covariance between the k^{th} and l^{th} observations in the selection index. Recalling the definition of I given by equation (4.28), it follows that

$$\sigma_I^2 = b_1^2\sigma_{p11} + b_2^2\sigma_{p22} + \ldots + 2b_1b_2\sigma_{p12} + 2b_1b_3\sigma_{p13} \ldots = \sum_{k=1}^{m} \sum_{l=1}^{m} b_k\, b_l \sigma_{p_{ki}} \tag{4.33}$$

Similarly, the covariance between H and I, recalling the definitions given at (4.27) and (4.28), is

$$\sigma_{HI} = b_1v_1\sigma_{g11} + b_1v_2\sigma_{g12} + \ldots + b_mv_n\sigma_{g_{mn}} = \sum_{k=1}^{m} \sum_{l=1}^{m} b_k\, v_i \sigma_{g_{ki}} \tag{4.34}$$

If we write the term to be maximized as, $M = \sigma_{HI}/\sigma_I$

then $\log M = \log \sigma_{HI} - \log \sigma_I$

or $\log M = \log \sigma_{HI} - \tfrac{1}{2} \log \sigma_I^2$

and substituting from (4.33) and (4.34):

$$\log M = \log(\sum\sum b_k v_i \sigma_{g_{ki}}) - \tfrac{1}{2}\log(\sum\sum b_k b_l \sigma_{p_{ki}}) \qquad (4.35)$$

Since M will be maximal when log$M$ is maximal, we can maximize M by differentiating log$M$ with respect to each of the b in turn and setting each partial differential to zero:

$$\frac{\delta \log M}{\delta b_k} = 0 \qquad \text{for } k = 1 \text{ to } m.$$

From standard differential algebra, with log$M$ defined at (4.35), it follows that

$$\frac{\delta \log M}{\delta b_k} = \frac{\sum_{i=1}^{n} v_i \sigma_{g_{ki}}}{\sigma_{HI}} - \frac{\sum_{l=1}^{m} b_l \sigma_{p_{kl}}}{\sigma_I^2}$$

Hence, M is maximal when $\sum_{l=1}^{m} b_l \sigma_{p_{kl}} = \frac{\sigma_I^2}{\sigma_{HI}} \sum_{i=1}^{n} v_i \sigma_{g_{ki}}$ (4.36)

But from standard regression theory: $\frac{\sigma_I^2}{\sigma_{HI}} = \frac{1}{b_{HI}}$

and if the index I is to give unbiased estimates of the aggregate genotype H, we recall that b_{HI} must equal 1. Hence (4.36) becomes,

$$\sum_{l=1}^{m} b_l \sigma_{p_{kl}} = \sum_{i=1}^{n} v_i \sigma_{g_{ki}} \qquad (4.37)$$

Since there are m observations in the index, there are m equations of the general form of (4.37),

i.e.

$$\sum_{l=1}^{m} b_l \sigma_{p_{1l}} = \sum_{i=1}^{n} v_i \sigma_{g_{1i}}$$

$$\sum_{l=1}^{m} b_l \sigma_{p_{2l}} = \sum_{i=1}^{n} v_i \sigma_{g_{2i}}$$

$$\vdots \qquad \vdots$$

$$\sum_{l=1}^{m} b_l \sigma_{p_{ml}} = \sum_{i=1}^{n} v_i \sigma_{g_{mi}}$$

If we write these equations in their expanded form, i.e.

$$\begin{aligned}
b_1\sigma_{p_{1.1}} + b_2\sigma_{p_{1.2}} \ \ldots \ + b_m\sigma_{p_{1.m}} &= v_1\sigma_{g_{1.1}} + v_2\sigma_{g_{1.2}} \ \ldots \ + v_n\sigma_{g_{1.n}} \\
b_1\sigma_{p_{2.1}} + b_2\sigma_{p_{2.2}} \ \ldots \ + b_m\sigma_{p_{2.m}} &= v_1\sigma_{g_{2.1}} + v_2\sigma_{g_{2.2}} \ \ldots \ + v_n\sigma_{g_{2.n}} \\
&\ \ \vdots \qquad\qquad\qquad\quad \vdots \\
b_1\sigma_{p_{m.1}} + b_2\sigma_{p_{m.2}} \ \ldots \ + b_m\sigma_{p_{m.m}} &= v_1\sigma_{g_{m.1}} + v_2\sigma_{g_{m.2}} \ \ldots \ + v_n\sigma_{g_{m.n}}
\end{aligned}$$

it is clear that they can be written in matrix notation as:

$$\mathbf{Pb = Gv} \tag{4.38}$$

where: **b** = column vector of *m* selection index coefficients
P = *m* x *m* matrix of phenotypic covariances among the observations in the index,
G = *m* x *n* matrix of genetic covariances among the *m* index observations and the *n* traits in the aggregate genotype
v = column vector of economic weights of the *n* traits in the aggregate genotype.

Recalling that pre-multiplying a matrix by itself yields an identity matrix, i.e. that, $\mathbf{P}^{-1}\,\mathbf{P} = \mathbf{I}$, the solution to obtaining **b** can be obtained by pre-multiplying both sides of (4.38) by $\mathbf{P}^{-1}$ to obtain,

$$\mathbf{b = P^{-1}Gv} \tag{4.39}$$

These are the so-called ***selection index equations*** that must be solved to find the optimal index weights.

4.4.1.2 Alternative derivation using matrix notation

The object is to minimize the variance of the difference between the predicted value, *I*, and the true value, *H*, i.e. minimize Var(*H-I*). Thus we wish to minimize

$$\begin{aligned}
E(H - I)^2 &= E[I - H)'\,(I - H)] \\
&= E[I - H)'\,(I - H)'] \\
&= E[(\mathbf{b'x - v'g})(\mathbf{x'b - g'v})] \\
&= E[(\mathbf{b'xx'b - b'xg'v - v'gx'b + v'gg'v}]
\end{aligned}$$

where **x** = column vector of observations and **g** = column vector of genetic values. Each of the terms in the above equality can be found as:

$$\begin{aligned}
E(\mathbf{b'xx'b}) &= \mathbf{b'}E(\mathbf{xx'})\mathbf{b} = \mathbf{b'Pb}, \\
E(\mathbf{b'xg'v}) &= \mathbf{b'}E(\mathbf{xg'})\mathbf{v} = \mathbf{b'Gv}, \\
E(\mathbf{v'gx'b}) &= \mathbf{v'G'b} = \mathbf{b'Gv} \qquad \text{since } \mathbf{v'G'b} \text{ is a scalar}
\end{aligned}$$

and $\quad E(\mathbf{v'gg'v}) = \mathbf{v'}E(\mathbf{gg'})\mathbf{v} = \mathbf{v'Cv}$

Therefore, to minimize $M = \mathbf{b'Pb - 2b'Gv + v'Cv}$

we must find the values which correspond to $\frac{\delta M}{\delta \mathbf{b}} = 0 = 2\mathbf{Pb} - 2\mathbf{Gv} + 0$

Therefore $\mathbf{Pb = Gv}$

Hence, $\mathbf{b = P^{-1}Gv}$ which is identical to equation (4.39).

4.4.1.2 Accuracy of the index

The accuracy of the selection index can be computed as the correlation between *I* and *H*:

$$r_{HI} = \frac{\sigma_{HI}}{\sigma_I \sigma_H} \tag{4.40}$$

The variance of the index, σ_I^2, is easily found as

$$\begin{aligned}\sigma_I^2 &= Var(b_1x_1 + b_2x_2 \ldots b_mx_m) \\ &= b_1^2\sigma_{p_1}^2 + b_2^2\sigma_{p2}^2 + \ldots + 2b_1b_2\sigma_{p12} + 2b_1b_3\sigma_{p13}\end{aligned}$$

or in matrix notation: $\sigma_I^2 = Var(\mathbf{b'x}) = \mathbf{b'}\ Var(\mathbf{x})\mathbf{b} = \mathbf{b'Pb}$ (4.41)

Following the same argument as for σ_I^2, $\sigma_H^2\ Var(\mathbf{v'g}) = \mathbf{v'}\ Var(\mathbf{g})\mathbf{v} = = \mathbf{v'Cv}$ (4.42)

where **C** is an *n* x *n* matrix of genetic covariances among the traits in the aggregate genotype.

Similarly, it follows that $\sigma_{HI} = Cov(\mathbf{b'x}, \mathbf{v'g}) = \mathbf{b'}\ Cov(\mathbf{x,g})\mathbf{v} = \mathbf{b'Gv}$ (4.43)

Hence, $$r_{HI} = \frac{\sigma_{HI}}{\sigma_I \sigma_H} = \frac{\mathbf{b'Gv}}{\sqrt{\mathbf{b'Pb\ v'Cv}}} \tag{4.44}$$

Note that because the index was constrained such that $b_{HI} = 1$ and $b_{HI} = \sigma_{HI}/\sigma_I^2$, thus $\sigma_{HI} = \sigma_I^2$

and from equations (4.41) and (4.43), $\mathbf{b'Pb = b'Gv}$ (4.45)

Thus, for the optimal index, equation (4.44) for accuracy simplifies to:

$$r_{HI} = \frac{\sigma_I}{\sigma_H} = \sqrt{\frac{\mathbf{b'Pb}}{\mathbf{v'Cv}}} = \sqrt{\frac{\mathbf{b'Gv}}{\mathbf{v'Cv}}} \tag{4.46}$$

Note, however, that equations (4.45) and (4.46) only hold for the optimal index, whereas equation (4.44) holds for any arbitrary index.

4.4.2 Family Selection Indexes

With family selection indexes, the problem is to combine information from different types of relatives to provide the most accurate estimate of the additive genetic value of a given trait (g) for a given individual. As indicated previously, in this case $H = g$, $\mathbf{v} = [1]$, and $\sigma_H^2 = \sigma_g^2$. This simplifies derivations to:

from equation (4.39) $$\mathbf{b} = \mathbf{P}^{-1}\mathbf{G} \qquad (4.47)$$

and from equation (4.46) $$r_{HI} = r_{g,\hat{g}} = \sqrt{\frac{\mathbf{b'G}}{\sigma_\mathbf{g}^2}} \qquad (4.48)$$

4.4.2.1 Examples of family selection indexes

Single source of information

The simplest form of a family index are the cases discussed in sections 4.2 and 4.3, where only a single source of observations is used, i.e. a single record or the mean of m records of the same type. The simplest case is a single record of the phenotype of the individual itself. In this case, the selection index is $I = \hat{g} = b_1 x_1$ and the aggregate genotype is $H = g$

where x_1 and g are both expressed as deviations from their population mean.

In this case, $\mathbf{P} = \sigma_{\bar{x}}^2$ and $\mathbf{G} = \sigma_{g\bar{x}}$

Hence, $\mathbf{b} = b = \mathbf{P}^{-1}\mathbf{G} = (\sigma_{\bar{x}}^2)^{-1}\sigma_{g\bar{x}} = \sigma_{g\bar{x}} / \sigma_{\bar{x}}^2$

The accuracy of selection, given by (4.48), is $$r_{HI} = r_{g,\hat{g}} = \sqrt{\frac{\mathbf{b'G}}{\sigma_\mathbf{g}^2}} = \sqrt{\frac{b\mathbf{G}}{\mathbf{C}}} = \frac{\sigma_{g\bar{x}}}{\sigma_{\bar{x}}\sigma_g}$$

These results are equivalent to those obtained in section 4.4.2.

More than one observation in the index

For the previous example, when there was only one source of information in the index, algebraic expectations for b and r_{HI} were derived directly in terms of basic population parameters. Appropriate formulae can be derived for a wide range of situations, including some situations with two or more sources for a single trait. A few more examples are given in Table 4.1, and a more extensive list is given by Van Vleck, 1993. Once there is more than one source of information in the index, it is often more useful to derive the expectations for the elements of **P** and **G** and then solve for b, b_{HI}, etc. using a computer package for matrix programming, rather than attempting to derive an algebraic solution directly.

Table 4.1 Selection index coefficients, b, and accuracies, r_{HI}, for some common sources of information in family indexes to predict additive genetic value for a single trait.

Information Source	**b**	$r_{HI} = r_{g,\hat{g}}$
Single record on individual	h^2	$\sqrt{h^2}$
m records on individual	$\frac{mh^2}{(m-1)t+1}$	$\sqrt{\frac{mh^2}{(m-1)t+1}}$
Single record on one parent	$\frac{1}{2} h^2$	$\frac{1}{2}\sqrt{h^2}$
m records on one parent	$\frac{mh^2}{2((m-1)t+1)}$	$\frac{1}{2}\sqrt{\frac{mh^2}{((m-1)t+1)}}$
Single record on both parents	$\frac{1}{2} h^2, \frac{1}{2} h^2$	$0.71\ \sqrt{h^2}$
m records on both parents	$\frac{mh^2}{2((m-1)t+1)}, \frac{mh^2}{2((m-1)t+1)}$	$0.71\ \sqrt{\frac{mh^2}{((m-1)t+1)}}$
Mean of n half-sib progeny with one record	$\frac{2nh^2}{((n-1)h^2+4)}$	$\sqrt{\frac{nh^2}{(n-1)h^2+4}}$

4.4.2.2 General equations to derive elements of selection index matrices

This section describes general equations that can be used to derive elements of the **P, G**, and **C** matrices that are needed for selection index calculations. Possible sources of information in the index are individual records and the mean of m records on a group of individuals or of m own records. Records on different traits can be included in the index and the aggregate genotype can consist of a single trait or of multiple traits.

It must be noted that these equations assume no selection or inbreeding. The impact of selection and inbreeding on index derivations will be discussed in a later chapter.

Notation:

- m = number of records within a group
- c^2 = common environment component within a group of individuals that contribute to a mean
- σ_{p_k} = phenotypic standard deviation of trait k
- σ_{g_k} = additive genetic standard deviation of trait k
- $r_{p_{kl}}$ = phenotypic correlation between traits k and l
- $r_{g_{kl}}$ = genetic correlation between traits k and l
- a = additive genetic relationship within a group
- a_{ij} = additive genetic relationship between individual(s) in groups i and j
- a_{hj} = additive genetic relationship between the individual in the breeding goal (h) and individuals in group j

P-matrix

diagonal:

• Variance of *m* records of a given type

$$\frac{1+(m-1)t}{m}\sigma_p^2 \qquad (=\sigma_p^2 \text{ for } m=1) \tag{4.49}$$

with t = repeatability for repeated records

$t = ah^2 + c^2$ for multiple individuals

off-diagonal:

• Covariance between mean of *m* records on different traits (*k* and *l*) for the same group:

$$\frac{r_{p_{kl}}\sigma_{p_k}\sigma_{p_l}+(m-1)a\,r_{g_{kl}}\sigma_{g_k}\sigma_{g_l}}{m} \qquad (= r_{p_{kl}}\sigma_{p_k}\sigma_{p_l} \text{ for } m=1) \tag{4.50}$$

• Covariance between (mean of) record(s) on same trait *k* for different groups (*i* and *j*):

$$(a_{ij}\,h_k^2 + c_k^2)\sigma_{p_k}^2 \tag{4.51}$$

• Between records on different traits (*k* and *l*) in different groups (*i* and *j*):

$$a_{ij}r_{g_{kl}}\sigma_{g_k}\sigma_{g_l} \tag{4.52}$$

G-matrix

• Covariance of the genetic value for trait *k* on the breeding goal animal (*h*) with records on trait *l* for group *j*

$$a_{hj}r_{g_{kl}}\sigma_{g_k}\sigma_{g_l} \qquad (=a_{hj}\sigma_{g_k}^2 \text{ if } k=l) \tag{4.53}$$

C-matrix

Diagonal:

• Variance of genetic value for trait *k*

$$\sigma_{g_k}^2 \tag{4.54}$$

Off-diagonal:

• Covariance between genetic values for traits *k* and *l* on breeding goal animal

$$r_{g_{kl}}\sigma_{g_k}\sigma_{g_l} \tag{4.55}$$

4.4.2.2.1 Example Index of individual record and full-sib mean performance

Imagine a situation where we have an observation on the individual's performance plus the mean performance of that individual's *m* full sibs, and we wish to predict the individual's breeding value. The index will then take the form,

$$I = \hat{g} = b_1x_1 + b_2x_2$$

where x_1 is the individual's phenotype and x_2 is the full-sib mean phenotype, both expressed as deviations from the population mean.

Then **P** and **G** will take the form,

$$\mathbf{P} = \begin{bmatrix} \sigma^2_{x_1} & \sigma_{x_1x_2} \\ \sigma_{x_1x_2} & \sigma^2_{x_2} \end{bmatrix}, \qquad \mathbf{G} = \begin{bmatrix} \sigma_{x_1 g} \\ \sigma_{x_2 g} \end{bmatrix} \tag{4.56}$$

Elements of **P** and **G** can be derived using the equations developed in the previous section. As an example, consider a selection index based on individual phenotype and the mean performance of 5 full sibs for animals in a population recorded for growth rate with a heritability of 0.5. We will assume there is no common environmental component.

Then:
$$\mathbf{P} = \begin{bmatrix} 1 & \tfrac{1}{2}h^2 \\ \tfrac{1}{2}h^2 & \dfrac{1+(m-1)\tfrac{1}{2}h^2}{m} \end{bmatrix} \sigma^2_p = \begin{bmatrix} 1 & .25 \\ .25 & .4 \end{bmatrix} \sigma^2_p \tag{4.57}$$

and:
$$\mathbf{G} = \begin{bmatrix} h^2 \\ \tfrac{1}{2}h^2 \end{bmatrix} \sigma^2_p = \begin{bmatrix} .5 \\ .25 \end{bmatrix} \sigma^2_p \tag{4.58}$$

Selection index coefficients are given by $\mathbf{b} = \mathbf{P}^{-1}\mathbf{G}$ which, since σ^2_p cancels out, gives

$$\mathbf{b} = \begin{bmatrix} 1 & .25 \\ .25 & .4 \end{bmatrix}^{-1} \begin{bmatrix} .5 \\ .25 \end{bmatrix} = \begin{bmatrix} .4074 \\ .3704 \end{bmatrix}$$

Hence, the selection index would be

$$I = \hat{g} = 0.4074\, x_1 + 0.3704\, x_2$$

The accuracy of this index or EBV is given by

$$r_{HI} = r_{g,\hat{g}} = \sqrt{\frac{\mathbf{b'G}}{\sigma^2_{\mathbf{g}}}} = \sqrt{\frac{\begin{bmatrix} .4074 \\ .3704 \end{bmatrix}' \begin{bmatrix} .5 \\ .25 \end{bmatrix} \sigma^2_p}{0.5\,\sigma^2_p}} = 0.77 \tag{4.59}$$

We can compare this accuracy with the accuracy of 0.707 for phenotypic selection on the same trait as shown in Section 2.8.1. By adding information on the mean performance of 5 full sibs, the accuracy of evaluation is increased from 0.71 to 0.77, i.e. by 8.9%. And, since $S = i r_{g,\hat{g}}\, \sigma_g$, and i and σ_g are not affected by the addition of extra information to the index, expected response will also increase by 8.9%.

4.5 Selection Index and Animal Model BLUP

An assumption in the use of selection indexes to estimate breeding values is either that there are no fixed effects in the data used, or that fixed effects are known without error. This may be true in some situations. An example are some forms of selection in egg-laying poultry where all birds are hatched in one or two very large groups and reared and recorded together in single locations. But in most cases, fixed effects are important and not known without error. For example, with pigs, different litters are born at different times of the year, often in several different locations. In progeny testing schemes in dairy cattle, cows are born continuously, begin milking at different times of year and in a very large number of different herds.

For this reason (and others) genetic evaluation in practice is often based on methods of Best Linear Unbiased Prediction, BLUP, which is a linear mixed model methodology which simultaneously estimates random genetic effects while accounting for fixed effects in the data in an optimum way. Relationships among animals can be included in the model. A sire model would account for relationships through the sire, i.e. half-sibships. A sire and dam model accounts for relationships through both the sire and the dam, i.e. full and half-sibships. An animal model accounts for all relationships among all animals in the data set. A description of the theory and application of BLUP, and animal model BLUP in particular, can be found in Schmidt (1988), Mrode (1996), and Lynch and Walsh (1998).

When relationships are included in a BLUP procedure, the method is equivalent to a selection index with the additional ability to efficiently estimate and correct the data for fixed effects. In the absence of fixed effects, BLUP with relationships is identical to a selection index. For example, a BLUP sire and dam model without records on the sire and dam would be the same as a selection index based on individual, full sib and half-sib records. An animal model BLUP would be equivalent to a selection index based on all related individuals, including ancestors, with records.

These equivalences are important for the design of breeding programs, because it means that in many situations, many aspects of selection programs with BLUP evaluation can be effectively studied with simulations based on equivalent selection indexes. There are two approaches to modeling Animal model BLUP EBV using selection index:

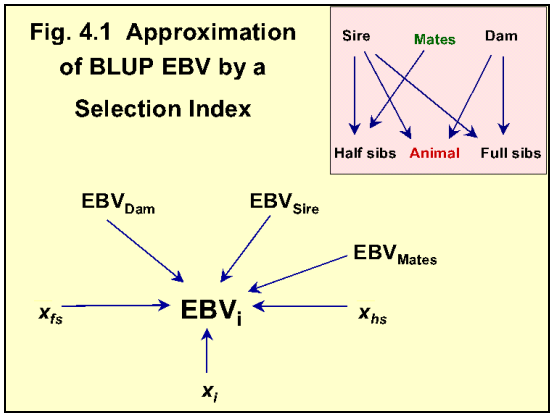
1) Develop a selection index based only on those relatives providing the greatest amount of information, rather than all possible relatives as in the animal model. For example, when records on parents, full and half sibs, and progeny are accounted for, information on more distant relatives may only provide a trivial increase in accuracy of selection.

2) Develop a selection index that includes parental EBV as sources of information, along with records on the individual itself, collateral relatives, and progeny, if available. In such an index, the parental EBV account for all ancestral information.

Development of the first type of index follows from the previous sections. We will describe the development of the second type of index in more detail in the following.

Consider the following information sources to estimate the BV of individual i for a hierarchical breeding design in which each sire is mated to m dams and each dam has n progeny (Figure 4.1):

- x_i = the animal's own record,
- x_{fs} = the average of single records on the individual's n-1 full sibs
- x_{hs} = the average of single records on the individual's $(m\text{-}1)n$ half sibs
- $\hat{g}_s$ = the EBV of the individual's sire, excluding x_i, x_{fs}, and x_{hs}
- $\hat{g}_d$ = the EBV of the individual's dam, excluding x_i, x_{fs}, and x_{hs}
- $\bar{\hat{g}}_m$ = the mean EBV of the $(m\text{-}1)$ mates of the sire that produced the individual's half sibs



Based on this information, the selection index to estimate the individual's BV can be formulated as:

$$I_i = \hat{g}_i = b_1 x_i + b_2 x_{fs} + b_3 x_{hs} + b_4 \hat{g}_s + b_5 \hat{g}_d + b_6 \bar{\hat{g}}_m \tag{4.60}$$

$$\mathbf{P} = \begin{bmatrix} \sigma^2_{x_i} & \sigma_{x_i x_{fs}} & \sigma_{x_i x_{hs}} & \sigma_{x_i \hat{g}_s} & \sigma_{x_i \hat{g}_d} & \sigma_{x_i \bar{\hat{g}}_m} \\ & \sigma^2_{x_{fs}} & \sigma_{x_{fs} x_{hs}} & \sigma_{x_{fs} \hat{g}_s} & \sigma_{x_{fs} \hat{g}_d} & \sigma_{x_{js} \bar{\hat{g}}_m} \\ & & \sigma^2_{x_{hs}} & \sigma_{x_{hs} \hat{g}_s} & \sigma_{x_{hs} \hat{g}_d} & \sigma_{x_{hs} \bar{\hat{g}}_m} \\ & & & \sigma^2_{\hat{g}_s} & \sigma_{\hat{g}_s \hat{g}_d} & \sigma_{\hat{g}_s \bar{\hat{g}}_m} \\ & & & & \sigma^2_{\hat{g}_d} & \sigma_{\hat{g}_d \bar{\hat{g}}_m} \\ & & & & & \sigma^2_{\bar{\hat{g}}_m} \end{bmatrix} \tag{4.61}$$

$$\mathbf{G} = \begin{bmatrix} \sigma_{g_i x_i} & \sigma_{g_i x_{fs}} & \sigma_{g_i x_{hs}} & \sigma_{g_i \hat{g}_s} & \sigma_{g_i \hat{g}_d} & \sigma_{g_i \bar{\hat{g}}_m} \end{bmatrix} \tag{4.62}$$

$$\mathbf{P} = \begin{bmatrix} 1 & {}^1/_2\,h^2 + c^2 & {}^1/_4\,h^2 & {}^1/_2\,r_s^2 h^2 & {}^1/_2\,r_d^2 h^2 & 0 \\ & \dfrac{1+(n-2)({}^1/_2\,h^2 + c^2)}{n-1} & {}^1/_4\,h^2 & {}^1/_2\,r_s^2 h^2 & {}^1/_2\,r_d^2 h^2 & 0 \\ & & {}^1/_4\,h^2 + \dfrac{{}^1/_4\,h^2 + c^2}{m-1} + \dfrac{1-{}^1/_2\,h^2 - c^2}{n(m-1)} & {}^1/_2\,r_s^2 h^2 & 0 & \dfrac{{}^1/_2\,r_m^2 h^2}{m-1} \\ & & & r_s^2 h^2 & 0 & 0 \\ & & & & r_d^2 h^2 & 0 \\ & & & & & \dfrac{r_m^2 h^2}{m-1} \end{bmatrix} \sigma_p^2 \quad (4.63)$$

$$\mathbf{G} = \begin{bmatrix} h^2 & {}^1/_2\,h^2 & {}^1/_4\,h^2 & {}^1/_2\,r_s^2\,h^2 & {}^1/_2\,r_d^2\,h^2 & 0 \end{bmatrix} \sigma_p^2 \quad (4.64)$$

With
$$x_{hs} = \left(\sum_{k=1}^{m-1} \sum_{l=1}^{n} \frac{x_{kl}}{n} \right) / (m\text{-}1) \quad (4.65)$$

Where
$$x_{kl} = {}^1/_2\,g_s + {}^1/_2\,g_{d_k} + g_{ms_{kl}} + c_{kl} + e_{kl} \quad (4.66)$$

Thus
$$x_{hs} = {}^1/_2\,g_s + \frac{\sum_{k=1}^{m-1} \left({}^1/_2\,g_{d_k} + c_k \right)}{m-1} + \frac{\sum_{k=1}^{m-1} \sum_{l=1}^{n} \left(g_{ms_{kl}} + e_{kl} \right)}{n(m-1)} \quad (4.67)$$

And
$$\sigma^2_{x_{hs}} = {}^1/_4\sigma_g^2 + \frac{{}^1/_4\sigma_g^2 + c^2\sigma_p^2}{m-1} + \frac{{}^1/_2\sigma_g^2 + \sigma_e^2}{n(m-1)} \quad (4.68)$$

Also,
$$\sigma^2_{\hat{g}} = r^2_{g,\hat{g}}\,\sigma_g^2 \quad (4.69)$$

And
$$\sigma_{x_i\,\hat{g}_s} = \sigma_{({}^1/_2\,g_s + {}^1/_2\,g_d + g_{m_i} + e_i, \hat{g}_s)} = \sigma_{({}^1/_2 g_s, \hat{g}_s)} = {}^1/_2\sigma_{g_s,\hat{g}_s} = {}^1/_2\,r_s^2\sigma_g^2 \quad (4.70)$$

As before, index weights can be derived as: $\mathbf{b} = \mathbf{P}^{-1}\mathbf{G}$

And accuracy as: $r_{g,\hat{g}} = \sqrt{\mathbf{b}'\mathbf{P}\mathbf{b}/\sigma_g^2}$

Because elements of the **P** and **G** matrices depend on accuracy of EBV of the sire and dam, which in turn depend on the EBV of their parents, iteration must be used to derive the final index and its accuracy. This can be done by using some starting value for accuracy of parental EBV, e.g. $r_s = r_d = h$, deriving the index and its accuracy, and then using the resulting accuracy as the new accuracy for r_s and r_d, resolving the index, etc.. This process of iteration is akin to building pedigree information; in each iteration, an additional ancestral generation with data is added, which increases accuracy but at a diminishing rate, until accuracy asymptotes (see example).

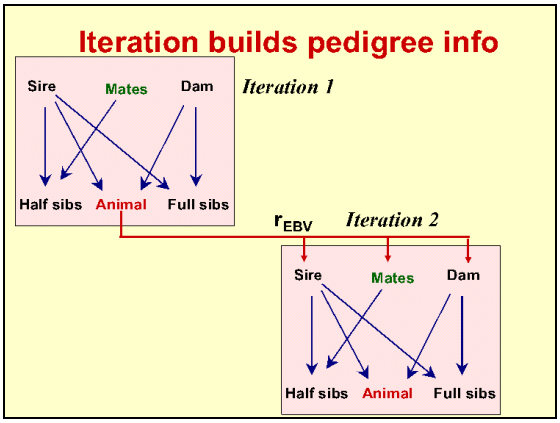
Iterative Procedure to Compute Accuracy of EBV
Building-up pedigree information

$EBV_i = b' [x_i , x_{fs} , x_{hs} , \hat{g}_{dam} , \hat{g}_{sire} , \hat{g}_{mates}]'$

$b = P^{-1}G \qquad r_{\hat{g},g} = \sqrt{b'Pb}/\sigma_g$

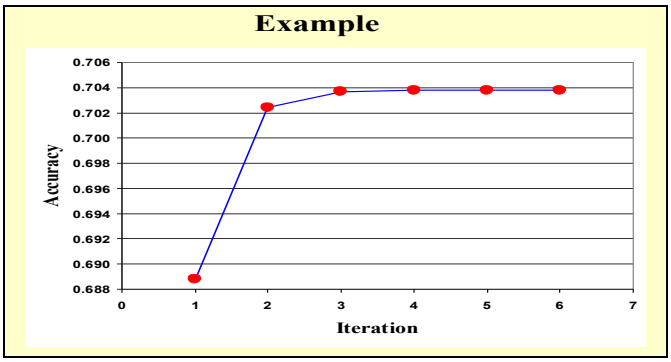
1) Set accuracy of $\hat{g}_{sire}=\hat{g}_{dam}=\hat{g}_{mates}= h$ (own record)
2) Set up index (P, G) and derive accuracy $r_{\hat{g},g}=\sqrt{b'Pb}/\sigma_g$
3) Set accuracy of $\hat{g}_{sire}=\hat{g}_{dam}=\hat{g}_{mates}$ equal to $r_{\hat{g},g}$
4) Repeat steps 2 and 3 until accuracy converges

Needs adaptation if selection after progeny



Example

Iteration 1		σ_p^2= 100		n= 10	n offspring per dam --> n-1 full sibs of individual					
		h^2= 0.25		m= 20	m dams per sire with n offspring each					
		c^2= 0			---> (m-1)n halfsibs consisting of (m-1) fullsib families					
STARTING VALUE		σ_g^2= 25								
FOR ACCURACY OF		σ_e^2= 75								
PARENTAL EBV										
$r_s=r_d=r_m$= 0.5000		P =	x_i	x_{fs}	x_{hs}	g_{sire}	g_{dam}	g_{mates}		
(start with r=h)		x_i	***100.00***	12.50	6.25	3.13	3.13	0.00	b=P⁻¹G=	0.175
		x_{fs}	12.50	***22.22***	6.25	3.13	3.13	0.00		0.291
		x_{hs}	6.25	6.25	***7.04***	3.13	0.00	0.16		0.464
		g_{sire}	3.13	3.13	3.13	***6.25***	0.00	0.00		0.035
		g_{dam}	3.13	3.13	0.00	0.00	***6.25***	0.00		0.267
		g_{mates}	0.00	0.00	0.16	0.00	0.00	***0.33***		-0.232
										0.6888
		G =	25.00	12.50	6.25	3.13	3.13	0.00	Acc =	
Iteration 2										
$r_s=r_d=r_m$= 0.6888		P =	x_i	x_{fs}	x_{hs}	g_{sire}	g_{dam}	g_{mates}		
from previous iteratior		x_i	***100.00***	12.50	6.25	5.93	5.93	0.00	b=P⁻¹G=	0.169
		x_{fs}	12.50	***22.22***	6.25	5.93	5.93	0.00		0.234
		x_{hs}	6.25	6.25	***7.04***	5.93	0.00	0.31		0.500
		g_{sire}	5.93	5.93	5.93	***11.86***	0.00	0.00		0.048
		g_{dam}	5.93	5.93	0.00	0.00	***11.86***	0.00		0.298
		g_{mates}	0.00	0.00	0.31	0.00	0.00	***0.62***		-0.250
									Acc =	**0.7024**
		G =	25.00	12.50	6.25	5.93	5.93	0.00		



In the previous selection indexes were used to provide genetic evaluations for a single trait based on records of that trait on the individual and/or other relatives. This is known as single-trait evaluation. It should be clear from selection index theory, that information on other traits could also be included in the index, to give a multi-trait evaluation (see Villanueva et al. 1993).

Chapter 5

Selection-Induced Gametic Phase Disequilibrium The Bulmer Effect

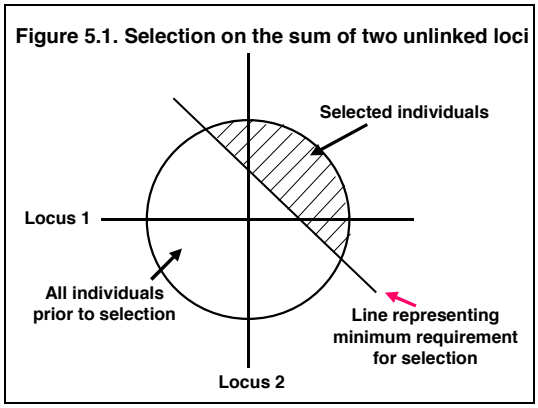
In the previous chapters, genetic variance was assumed constant over generations. Selection, however, has an impact not only on the mean of the population but also on genetic variance. Changes in genetic variance affect the amount of change that can be made in future generations.

The objective of this section is to model the effect of selection on genetic variance and to incorporate its effects in derivation of selection indexes and response to selection. As in the previous chapters, the basis of these models will be the infinitesimal genetic model, in which the trait is assumed to be affected by a large number of unlinked loci with small effect.

5.1 Effects of Selection on Genetic Variance

Pearson (1903), in his discussions on conditional variances early this century, noted that truncating a distribution affected both the mean and variance of the population. Anecdotally, founding animal breeders such as Lush, Falconer, and Henderson are said to have recognized that this could have implications for animal breeding since truncation selection could reduce genetic variance among parents from that observed before selection. Bulmer (1971, 1976, 1981) was the first to publish an examination of this effect of selection on genetic variance, and, consequently, the effect is often referred to as the "Bulmer Effect" or the effect of linkage disequilibrium. A more appropriate term is Falconer's "gametic phase disequilibrium" (See Falconer and Mackay, 1996, for an explanation of this term.)

To explain gametic phase disequilibrium, we will look at a situation where two unlinked genes affect a trait in an additive manner and each gene has a large number of alleles and both genes make the same contribution to genetic variance in the trait. Animals from a previously unselected population are selected on the sum of the genetic effects at both loci as illustrated in Figure 5.1.



When an animal has a high value at locus 1, it has a high chance of being selected, irrespective of the value at locus 2. Similarly, an animal with a high value at locus 2 will have a high value of being selected irrespective of the value at locus 1. However, animals with a moderately high value at locus 1 will only be selected when the value at locus 2 is also at least moderately high. As a consequence of this selection, effects at the two loci are negatively correlated in the selected individuals. In other words, the effects at the two loci in the selected individuals are no longer uncorrelated, i.e. selection has introduced ***gametic phase disequilibrium***.

Genetic variance in the trait is equal to variance of the sum of the gene effects at the two loci:

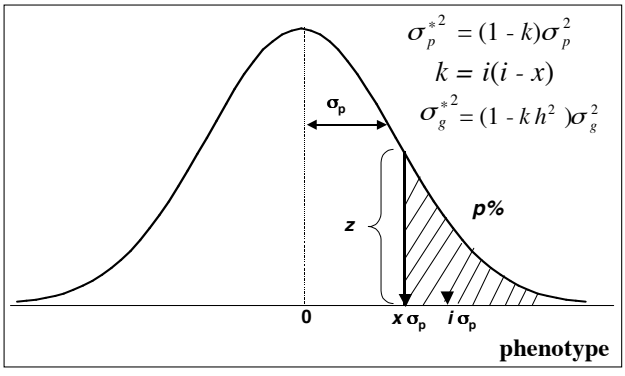
$$\sigma_g^2 = \sigma_{g_1}^2 + \sigma_{g_2}^2 + \sigma_{g_1 g_2} \qquad (5.1)$$

where $\sigma_{g_i}^2$ is the variance due to effects at locus *i* and $\sigma_{g_1 g_2}$ is the covariance between the effects at the two loci. Prior to selection, $\sigma_{g_1 g_2}$ is equal to zero, which reflects that the genes are in (linkage) equilibrium. Selection introduced a negative covariance and, as can be seen from equation (5.1), it is this negative covariance or disequilibrium between the two loci that reduces the genetic variance in the group of selected individuals.

It is important to recall that in the infinitesimal genetic model, individual genes are not recognized. The reduction in variance among selected individuals can also be derived from normal distribution theory, an approach that we will follow from here on. Nevertheless, it is important to keep in mind that the underlying mechanism that creates the reduction in genetic variance is the negative disequilibrium that is created between loci.

The effects of selection on genetic variance will first be described for a situation where animals are selected on their phenotype; a situation which is often referred to as "mass selection". The distribution of phenotypes prior to selection will have a standard deviation of σ_p, but as is clear from Figure 5.2, the standard deviation will be considerably less than among the proportion, p, of animals that are selected for breeding.

Figure 5.2 Mass Selection



The group of selected animals represents one tail of the distribution of the phenotypic distribution. If σ_p^2 is the phenotypic variance in the population before selection, *k* the factor by which the variance is reduced, and a subscript * is used to denote parameters after selection, the variance, σ_p^{*2}, in the selected individuals is:

$$\sigma_p^{*2} = (1 - k)\,\sigma_p^2 \qquad (5.2)$$

Factor *k* depends on intensity of selection (Pearson, 1903). When selection is by truncation of a normal distribution, then:

$$k = i(i - x) \qquad (5.3)$$

where *i* is the selection intensity and *x* is the standardized truncation point to the normal distribution corresponding to *i*, expressed in standard deviation units.

For genetic improvement, the question is what effect does selection on phenotype have on genetic variance of the trait. Again, from standard normal distribution theory it follows that with truncation selection on trait *y* the variance of a correlated trait *x* in the selected group, σ_x^{*2}, is given by

$$\sigma_x^{*2} = (1 - k\,r_{xy}^2)\,\sigma_x^2 \qquad (5.4)$$

where r_{xy} is the correlation between traits *x* and *y*.

Covariances between variables are similarly affected by selection. For example, the genetic covariance between *w* and *x* after selection on *y* is

$$\sigma_{wx}^* = \sigma_{wx} - k\frac{\sigma_{wy}\sigma_{xy}}{\sigma_y^2} \qquad (5.5)$$

Note that equation (5.4) for genetic variance is just a special case of (5.5) when $w = x$.

For mass selection, genetic variance among the selected individuals can be deduced as follows:

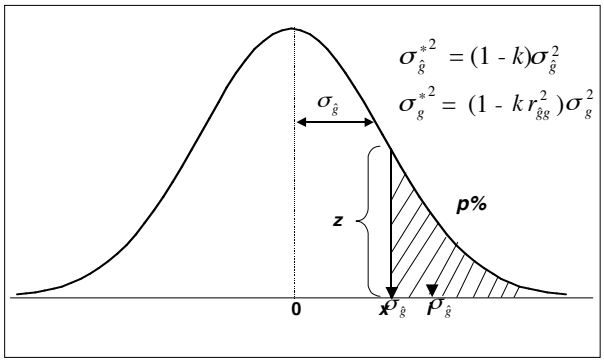
$$\begin{aligned}\sigma_g^{*2} &= (1 - k\,r_{gy}^2)\,\sigma_g^2 \qquad (5.6)\\ &= (1 - k\,h^2)\,\sigma_g^2\end{aligned}$$

where σ_g^2 is the genetic variance before selection. The correlation between additive genetic value *g* and phenotypic value, *y*, is *h*, the square of the heritability. The phenotypic variance is reduced by a factor *k* and the proportion h^2 of σ_g^2 is reduced by that same factor.

The formulae to calculate the reduction in genetic variance will now be generalized to a situation where selection is based on estimated breeding value $\hat{g}$ (Figure 5.3). Genetic variance among selected individuals can be derived using the correlation between the true genetic value *g* and the EBV, $\hat{g}$, which is equal to $r_{g,\hat{g}}$. When selection is on $\hat{g}$, the variance in $\hat{g}$ among the selected animals is $\sigma_{\hat{g}}^{*2} = (1 - k)\,\sigma_{\hat{g}}^2$ and from (5.4) it follows that the genetic variance among the selected animals is

$$\sigma_g^{*2} = (1 - k\,r_{\hat{g}g}^2)\,\sigma_g^2 \qquad (5.7)$$

Figure 5.3 EBV Selection



Referring back to Chapter 2, genetic variance of a population prior to selection can be partitioned into the parental and Mendelian sampling components as:

$$\sigma^2_g = ¼\,\sigma^2_{g_s} + ¼\,\sigma^2_{g_d} + \sigma^2_{g_m} \qquad (5.8)$$

This can now be modified to give genetic variance after selection among sires and dams. Using (5.6), genetic variance in the selected sires and dams can be calculated, where $\sigma^{*2}_{g_s}$ is the genetic variance among the selected sires and $\sigma^{*2}_{g_d}$ is the genetic variance among the selected dams. This leads to:

$$\sigma^2_g = ¼\,\sigma^{*2}_{g_s} + ¼\,\sigma^{*2}_{g_d} + \sigma^2_{g_m} \qquad (5.9)$$

This can be generalized to predict the genetic variance in generation t+1 from the variance among the parents selected in generation t

$$\sigma^2_{g_{(t+1)}} = ¼\,\sigma^{*2}_{g_{s(t)}} + ¼\,\sigma^{*2}_{g_{d(t)}} + \sigma^2_{g_m} \qquad (5.10)$$

Note that only the parental contributions to variation are affected by selection. The variance generated by Mendelian sampling, $\sigma^2_{g_m}$, is unaffected by selection and is equal to $½\,\sigma^2_{g_{(o)}}$ where $\sigma^2_{g_{(o)}}$ is the genetic variance in the unselected and non-inbred base population. The intuitive reasoning for this is that Mendelian sampling variance represents variation created by sampling one of a pair of parental alleles at each locus. This sampling process is unaffected by selection. Mendelian sampling variance is, however, affected by inbreeding, which will be discussed later.

Based on this, the following general recursive equation can be developed to predict genetic variance among progeny:

$$\sigma^2_{g_{(t+1)}} = ¼(1 - k_s\, r^2_{s_{(t)}})\,\sigma^2_{g_{(t)}} + ¼(1 - k_d\, r^2_{d_{(t)}})\,\sigma^2_{g_{(t)}} + ½\,\sigma^2_{g_{(o)}} \qquad (5.11)$$

where k_s and k_d are based on selection intensities among males and females, and $r_{s_{(t)}}$ and $r_{d_{(t)}}$ are the respective accuracies of selection in generation t.

5.2 Prediction of Genetic Variance and Response for Mass Selection

In Table 5.1, the genetic variance is given for different (4) generations of mass selection in males and females. Generation 0 is assumed to be unselected, h^2=½, and $\sigma^2_e = \sigma^2_{g_{(o)}}$ =100. Truncation selection is used in both males and females and 5% of animals with highest phenotype are selected. In that case: i = 2.063 and x = 1.645, which based on equation (5.3) results in $k = i(i-x)$ = 2.063(2.063-1.645) = 0.862. Using equation (5.6), genetic variance among selected parents (sires and dams) is (1-0.862x½)100 = 56.9.

From equation (5.11) it follows that genetic variance in generation 1 is equal to ¼x56.9 + ¼x56.9 + ½x50 = 78.45. Selection reduced genetic variance to 78.45. In the base population, σ^2_e was 100 and the level of this variance is not affected by selection. Heritability in generation 1 is now 78.45/(100+78.45)=0.44. With this new level of h^2, variance among parents selected in generation 1 can be calculated using (5.6) and variance in generation 2 using (5.11).

Table 5.1 Effect of truncation selection with p=5% in males and females (i=2.063, x=1.645) during 5 generations (t = 0 to 4) on additive genetic variance $\sigma^2_{g_{(t)}}$ and average additive genetic merit of individuals ($\bar{g}_{(t)}$). Heritability in generation 0 was ½ (no inbreeding).

t	$\sigma^2_{g_{(t)}}$	$h^2_{(t)}$	$\bar{g}_{(t)}$	$\bar{g}_{(t)} - \bar{g}_{(t-1)}$
0	100	0.50	50.0	0
1	78	0.43	64.6	14.6
2	74	0.43	76.7	12.1
3	74	0.42	88.3	11.6
4	73	0.42	99.8	11.5
5	73	0.42	111.3	11.5
Selection stopped (random selection from here on)				
6	87	0.47	111.3	0
7	93	0.48	111.3	0
8	97	0.49	111.3	0
9	98	0.49	111.3	0
10	99	0.49	111.3	0

From Table 5.1 it can be seen that genetic variance reaches an equilibrium after three generations of selection. Genetic variance is equal to 74 and does not decrease further although selection is continued. This is referred to as the asymptotic genetic variance. When this is reached, the amount of gametic phase disequilibrium created by selection of individuals is equal to the amount of gametic phase disequilibrium which is broken down during meiosis (Mendelian sampling). When selection is stopped after four generations, no new gametic phase disequilibrium is created in the parents and the variance reduction is halved each generation as a result of Mendelian sampling. After 10 generations, genetic variance is back to its original level.

Response to selection from one generation to the next can be predicted as derived in chapter 1, but using parameters that apply to the parental generation:

$$\bar{g}_{(t+1)} = \bar{g}_{(t)} + ih_{(t)}\sigma_{g_{(t)}} \quad (5.12)$$

The mean of the population changes as a result of selection. After five generations of selection the population level has increased by 111.3 units (Table 5.10. The greatest genetic gain was realized in generation 1 because this was the generation with the highest h^2 and genetic variance. Response in subsequent generations is reduced both because of a reduction in genetic variance, as well as a result of a reduction in accuracy of selection. The population remains at the same level after selection has stopped.

Genetic variance in the population is reduced by 26% after one round of selection but this is the result of a much larger, i.e. 52%, reduction in variance among selected sires and dams. This results from the fact that variance due to Mendelian sampling is not affected by selection and consequently remains 50. Another way to look at this is to consider variation within and between full sib families. Without selection, between and within family variances are both equal to 50. With selection, variation between full sib families is equal to $\frac{1}{4}\sigma^{*2}_{g_{s(t)}} + \frac{1}{4}\sigma^{*2}_{g_{d(t)}}$, while the within full sib family genetic variance is equal to $\sigma^2_{g_m} = \frac{1}{2}\sigma^2_{g_{(o)}}$. In generation 1, the between full-sib genetic variance is equal to ¼x56.9+¼x56.9=28.45, while the within full-sib variance remains equal to 50. This demonstrates that selection has changed the ratio of within and between family genetic variance. An implication of this is that using a reduced heritability in deriving selection index weights is not the correct way to deal with changes in genetic variance resulting from selection because this assumes that all components of genetic variance are affected in the same way, which is not true, as we have seen in equation (5.11). Mass selection is a special case in which we only use the observations on the individual and forms an exception to this rule.

5.2.1 Asymptotic Genetic Variance and Response to Selection

The previous enables recursive prediction of changes in variance and response to selection. Both variance and response reach steady state or asymptotic values after a number of generations. For the case of mass selection (and BLUP selection as we will see later), these steady state parameters can also be derived directly, as will be demonstrated below.

Starting with recursive equation (5.11): $\sigma^2_{g_{(t+1)}} = \frac{1}{4}(1-k_s r^2_{s_{(t)}})\sigma^2_{g_{(t)}} + \frac{1}{4}(1-k_d r^2_{d_{(t)}})\sigma^2_{g_{(t)}} + \frac{1}{2}\sigma^2_{g_{(o)}}$, steady state parameters (denoted by subscript $_{(L)}$) can be derived by setting $\sigma^2_{g_{(L)}} = \sigma^2_{g_{(t+1)}} = \sigma^2_{g_{(t)}}$, $r_{s_{(L)}} = r_{s_{(t)}}$, and $r_{d_{(L)}} = r_{d_{(t)}}$, which results in the following steady-state equation:

$$\sigma^2_{g_{(L)}} = \frac{1}{4}(1-k_s r^2_{s_{(L)}})\sigma^2_{g_{(L)}} + \frac{1}{4}(1-k_d r^2_{d_{(L)}})\sigma^2_{g_{(L)}} + \frac{1}{2}\sigma^2_{g_{(o)}} \quad (5.12)$$

This equation can be solved if an equation can be developed that expresses accuracy of selection at the limit, $r_{s_{(L)}}$ and $r_{d_{(L)}}$, in terms of $\sigma^2_{g_{(L)}}$ and base population parameters. This is possible for mass selection and, as will be shown later, also for selection on BLUP EBV but not in general for selection on other types of selection indexes.

For mass selection, $r_{s_{(t)}} = r_{d_{(t)}} = h_{(t)}$, and assuming for simplicity equal selected fractions in both sexes, thus $k_s=k_d=k$, equation (5.11) simplifies to the following the recursive equation:

$$\sigma^2_{g_{(t+1)}} = \tfrac{1}{2}(1 - k\, h^2_{(t)})\,\sigma^2_{g_{(t)}} + \tfrac{1}{2}\,\sigma^2_{g_{(o)}} \qquad (5.13)$$

and at the limit, from (5.12): $\sigma^2_{g_{(L)}} = \tfrac{1}{2}(1 - k\, h^2_{(L)})\,\sigma^2_{g_{(L)}} + \tfrac{1}{2}\,\sigma^2_{g_{(o)}}$ (5.14)

with $h^2_{(L)} = \sigma^2_{g_{(L)}} / (\sigma^2_{g_{(L)}} + \sigma^2_e)$ (5.15)

Using $\sigma^2_e = \dfrac{1 - h^2_{(0)}}{h^2_0}\,\sigma^2_{g_{(o)}}$ (5.16)

steady state heritability can be solved in terms of the base population heritability as:

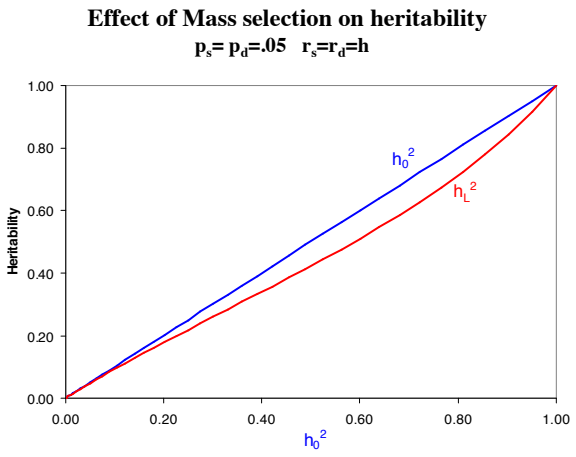
$$h^2_{(L)} = \frac{h^2_{(0)}}{1 + (1 - h^2_{(0)})\,k h^2_{(L)}} = \frac{-1 + \sqrt{1 + 4h^2_{(0)}k(1 - h^2_{(0)})}}{2k(1 - h^2_{(0)})} \qquad (5.17)$$

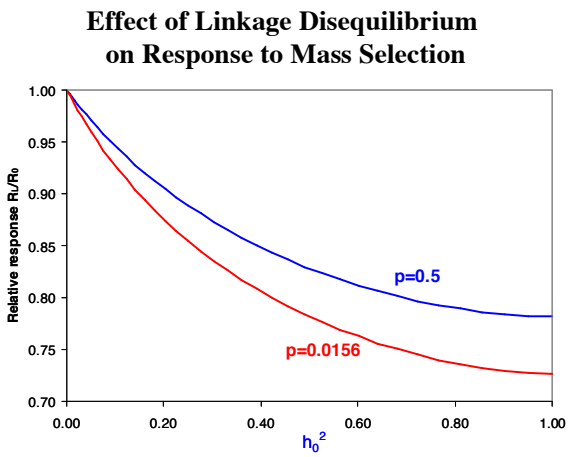
Substituting into equation (5.14) gives the following expression for the steady state genetic variance in terms of base population parameters:

$$\sigma^2_{g_{(L)}} = \frac{2\sigma^2_{g_{(0)}}(1 - h^2_{(0)})}{1 - 2h^2_{(0)} + \sqrt{1 + 4h^2_{(0)}k(1 - h^2_{(0)})}} \qquad (5.18)$$

An expression for the response to mass selection in the limit relative to response in the initial generation is:

$$R_{(L)}/R_{(1)} = \frac{i h_{(L)} \sigma_{g_{(L)}}}{i h_{(0)} \sigma_{g_{(0)}}} = \sqrt{\frac{h^2_{(L)}}{h^2_{(0)}(1 + k\, h^2_{(L)})}} \qquad (5.19)$$





5.3. Incorporating Gametic Phase Disequilibrium in the Selection Index

Because selection affects genetic variances and co-variances, it also affects elements of the **P** and **G** matrices that are needed to derive the optimal weights for selection indexes. In this section we will illustrate how changes in genetic parameters can be incorporated in selection index derivations and will evaluate their impact on accuracy of the index.

In general, because selection affects between and within-family variances differentially, derivation of elements of the **P** and **G** matrices must be based on the partitioning of the genetic value of individuals into parental and Mendelian sampling components:

$$g_{offspring} = \tfrac{1}{2}\, g_s + \tfrac{1}{2}\, g_d + g_m \tag{5.20}$$

and, from equation (5.10), genetic variance in the offspring generation, t, must be partitioned into:

$$\sigma^2_{g_{(t)}} = \tfrac{1}{4}\,\sigma^{*2}_{g_{s(t-1)}} + \tfrac{1}{4}\,\sigma^{*2}_{g_{d(t-1)}} + \sigma^2_{g_m} \tag{5.21}$$

with from (5.7)

$$\sigma^{*2}_{g_{s(t-1)}} = (1 - k_s\, r^2_{s_{(t-1)}})\,\sigma^2_{g_{(t-1)}}$$

$$\sigma^{*2}_{g_{d(t-1)}} = (1 - k_d\, r^2_{d_{(t-1)}})\,\sigma^2_{g_{(t-1)}}$$

$$\sigma^2_{g_m} = \tfrac{1}{2}\,\sigma^2_{g_{(o)}}$$

As an example consider the situation where selection of sires and dams is on an index using phenotype of the individual and the mean performance of m full sibs. The index for selection in generation t will then take the form,

$$\hat{g}_{(t)} = b_{1_{(t)}} x_1 + b_{2_{(t)}} x_2 \tag{5.22}$$

where x_1 is the individual's phenotype and x_2 is the full-sib mean phenotype, both expressed as deviations from the population mean. Then the matrices needed to derive the index for generation t, $\mathbf{P}_{(t)}$ and $\mathbf{G}_{(t)}$ will take the form,

$$\mathbf{P}_{(t)} = \begin{bmatrix} \sigma^2_{x_1} & \sigma_{x_1 x_2} \\ \sigma_{x_1 x_2} & \sigma^2_{x_2} \end{bmatrix}, \qquad \mathbf{G}_{(t)} = \begin{bmatrix} \sigma_{x_1 g} \\ \sigma_{x_2 g} \end{bmatrix} \tag{5.23}$$

Elements can be derived as follows:

From equation (5.21):

$$\sigma^2_{x_1} = \tfrac{1}{4}\,\sigma^{*2}_{g_{s(t-1)}} + \tfrac{1}{4}\,\sigma^{*2}_{g_{d(t-1)}} + \sigma^2_{g_m} + \sigma^2_e \tag{5.24}$$

$$\sigma^2_{x_2} = \tfrac{1}{4}\,\sigma^{*2}_{g_{s(t-1)}} + \tfrac{1}{4}\,\sigma^{*2}_{g_{d(t-1)}} + (\sigma^2_{g_m} + \sigma^2_e)/m \tag{5.25}$$

$$\sigma_{x_1 x_2} = \tfrac{1}{4}\,\sigma^{*2}_{g_{s(t-1)}} + \tfrac{1}{4}\,\sigma^{*2}_{g_{d(t-1)}} \tag{5.26}$$

$$\sigma_{x_1 g} = \tfrac{1}{4}\,\sigma^{*2}_{g_{s(t-1)}} + \tfrac{1}{4}\,\sigma^{*2}_{g_{d(t-1)}} + \sigma^2_{g_m} \tag{5.27}$$

and

$$\sigma_{x_2 g} = \tfrac{1}{4}\,\sigma^{*2}_{g_{s(t-1)}} + \tfrac{1}{4}\,\sigma^{*2}_{g_{d(t-1)}} \tag{5.28}$$

In generation 0, prior to selection, the above equations simplify to those derived in section 4.4.2.2.1.

For a trait with $h^2 = 0.5$, $\sigma^2_{g_{(0)}} = 25$, $\sigma^2_{p_{(0)}} = 50$, and m=5 full-sibs, we get the following:

$$\mathbf{P}_{(0)} = \begin{bmatrix} 50 & 12.5 \\ 12.5 & 20 \end{bmatrix} \qquad \mathbf{G}_{(0)} = \begin{bmatrix} 25 \\ 12.5 \end{bmatrix} \quad \text{and} \quad \mathbf{b}_{(0)} = \mathbf{P}_{(0)}^{-1}\mathbf{G}_{(0)} = \begin{bmatrix} .4074 \\ .3704 \end{bmatrix}$$

Accuracy is $$r_{(0)} = \sqrt{\frac{\mathbf{b}'_{(0)}\mathbf{G}_{(0)}}{\sigma^2_{g_{(0)}}}} = 0.77$$

When in generation 0 only the 5% of sires and dams with the highest EBV are used to produce offspring, $k = 0.863$ and

$$\sigma^{*2}_{g_{s(1)}} = \sigma^{*2}_{g_{d(1)}} = (1 - k\, r^2_{(0)})\,\sigma^2_{g_{(0)}} = (1 - 0.863 \times 0.77^2)25 = 12.21$$

and $$\sigma^2_{g_{(t)}} = \tfrac{1}{4}\,\sigma^{*2}_{g_{s(t-1)}} + \tfrac{1}{4}\,\sigma^{*2}_{g_{d(t-1)}} + \sigma^2_{g_m} = 18.61$$

Using these values to derive elements of the **P** and **G** matrices for t=1 we get:

$$\mathbf{P}_{(1)} = \begin{bmatrix} 43.61 & 6.11 \\ 6.11 & 13.61 \end{bmatrix} \qquad \mathbf{G}_{(1)} = \begin{bmatrix} 18.61 \\ 6.11 \end{bmatrix} \quad \text{and} \quad \mathbf{b}_{(1)} = \mathbf{P}_{(1)}^{-1}\mathbf{G}_{(1)} = \begin{bmatrix} .3883 \\ .2746 \end{bmatrix}$$

Accuracy is $$r_{(1)} = \sqrt{\frac{\mathbf{b}'_{(1)}\mathbf{G}_{(1)}}{\sigma^2_{g_{(1)}}}} = 0.69$$

Using the recursive equations, this accuracy can be used to predict response to selection from t=1 to t=2 and to derive the genetic variance and selection in t=2.

Note that, compared to generation 0, selection reduced the variance among sires and dams and, as a consequence, the relative importance of observations on full-sibs is lower for the index used for selection in t=1 and the relative importance of observations on the individual has increased.

The reduced importance of full-sib information can also be illustrated by comparing accuracy of the index to the accuracy of selecting on own phenotype alone, which is equal to $h_{(0)}$ and $h_{(1)}$ for t=0 and t=1, respectively. Based on this, the efficiency of the index excluding information from the full-sibs is 0.71/0.77 = 0.92 and 0.65/0.69 = 0.95 before and after one round of selection.

5.4 Incorporating Gametic Phase Disequilibrium in BLUP EBV

The previous section described methods to incorporate the effect of selection on genetic variance components in derivation of selection indexes based on recursive equations. In principle, these

methods can also be applied to the selection indexes described in section 4.5, method 1, to approximate BLUP EBV. Examples are in Wray and Hill (1989) and Villaneuva et al. (1993).

For BLUP EBV, however, an alternative method can be used to incorporate the Bulmer effect, which facilitates direct derivation of steady state parameters. This method is based on the second approach for approximating BLUP EBV described in section 4.5 and utilizes the important property of BLUP EBV that their prediction error variance (PEV) does not depend on selection, but only on the amount of information used, with information defined as the number and type of records that is available on the individual itself and its relatives. This was described by Henderson (1975), using the argument that PEV's are based on the inverse of the coefficient matrix, which depends on the design matrices, the matrix of additive genetic relationships, and genetic parameters in the base population:

$$\sigma^2_{\varepsilon} = \text{Var}(\boldsymbol{\varepsilon}) = \text{Var}(\hat{\boldsymbol{g}} - \boldsymbol{g}) = \mathbf{C}_{22} \qquad (5.29)$$

where $\boldsymbol{\varepsilon}$, $\hat{\boldsymbol{g}}$, and $\boldsymbol{g}$ are vectors of prediction errors, EBV, and BV, respectively, and $\mathbf{C}_{22}$ is the part of the inverse of the coefficient matrix of the mixed model equations that corresponds to animal breeding values. Elements of $\mathbf{C}_{22}$ do not depend on selection. Therefore, the PEV of a particular animal with a particular amount of information in an unselected population is the same as if that animal was in a selected population (but with selection accounted for through ancestor information). Thus, to get the PEV of an EBV, the mixed model equations can be set up ignoring the effect of selection on genetic variance and solved for. The same applies to approximations of BLUP EBV using the selection index methods described in section 4.5. Thus, the variance of prediction errors can be derived as:

$$\sigma^2_{\varepsilon_{(0)}} = (1 - r^2_{(0)})\,\sigma^2_{g_{(o)}} \qquad (5.30)$$

where the subscript 0 (t=0) refers to parameters derived for an unselected population, and $r_{(0)}$ is the accuracy of the BLUP EBV, derived using an index that ignores the effect of selection, following section 4.5.

Although selection doesn't affect the PEV, and, therefore, remains equal to $\sigma^2_{\varepsilon_{(0)}}$, PEV can also be derived based on the accuracy and genetic variance in the selected population as:

$$\sigma^2_{\varepsilon_{(t)}} = (1 - r^2_{(t)})\,\sigma^2_{g_{(t)}} \qquad (5.31)$$

Thus, using the property that PEV is unaffected by selection:

$$\sigma^2_{\varepsilon_{(t)}} = \sigma^2_{\varepsilon_{(0)}}$$
$$(1 - r^2_{(t)})\,\sigma^2_{g_{(t)}} = (1 - r^2_{(0)})\,\sigma^2_{g_{(o)}}$$

which, solving for $r^2_{(t)}$ results in:

$$r^2_{(t)} = 1 - (1 - r^2_{(0)})\,\sigma^2_{g_{(o)}} / \sigma^2_{g_{(t)}} \qquad (5.32)$$

This equation expresses the accuracy of EBV in a selected population in terms of the accuracy of EBV in an unselected population and the ratio of genetic variance in the unselected and selected population. Equation (5.32) holds for any generation and for any group of individuals.

Together with the recursive equation (5.11) for genetic variance:

$$\sigma^2_{g_{(t+1)}} = \tfrac{1}{4}(1 - k_s\, r^2_{s_{(t)}})\,\sigma^2_{g_{(t)}} + \tfrac{1}{4}(1 - k_d\, r^2_{d_{(t)}})\,\sigma^2_{g_{(t)}} + \tfrac{1}{2}\,\sigma^2_{g_{(o)}}$$

equation (5.32) provides a recursive system to derive genetic variance, accuracy of selection, and response to selection, as illustrated in Table 5.2 for selection on BLUP EBV that are described in section 4.5, method 2. Note that it is assumed that full pedigree information is available in generation zero.

Table 5.2. Recursive prediction of genetic variance, accuracy, and response with selection on BLUP EBV. Selected fractions are 0.2 and 0.5 for males and females, respectively, for a trait with heritability 0.25 and phenotypic variance 100. Selection is on BLUP EBV from a hierarchical mating structure with 20 mates per sire and 10 offspring per dam. Accuracy in generation zero is derived in section 4.5.

t	$\frac{(i_s+i_d)}{2}$	k_s	k_d	$\sigma_{g(0)}^2$	$\sigma_{g(t)}^2$ from t-1	$r_{(0)}$	$r_{(t)}=\sqrt{\frac{(1-(1-r_{(0)}^2)\sigma_{g(0)}^2}{\sigma_{g(t)}^2}}$	$\bar{g}_{(t+1)} = \bar{g}_{(t)} + {}^1/_2(i_s+i_d)r_{(t)}\sigma_{g(t)}$	$R_{(t)} = \bar{g}_{(t+1)} - \bar{g}_{(t)}$	$\sigma^{*2}_{gs(t)} = (1-r_{(t)}^2 k_s)\sigma_{g(t)}^2$	$\sigma^{*2}_{gd(t)} = (1-r_{(t)}^2 k_d)\sigma_{g(t)}^2$	$\sigma_{g(t+1)}^2 = {}^1/_2\sigma^{*2}_{gs(t)} + {}^1/_2\sigma^{*2}_{gd(t)} + {}^1/_2\sigma_{g(0)}^2$
0	1.1	0.78	0.64	25	25.00	0.704	0.704	3.871	3.871	15.326	17.074	20.600
1	1.1	0.78	0.64	25	20.60	0.704	0.623	6.979	3.108	14.363	15.490	19.963
2	1.1	0.78	0.64	25	19.96	0.704	0.607	9.961	2.982	14.224	15.261	19.871
3	1.1	0.78	0.64	25	19.87	0.704	0.604	12.924	2.963	14.204	15.228	19.858
4	1.1	0.78	0.64	25	19.86	0.704	0.604	15.884	2.960	14.201	15.223	19.856
5	1.1	0.78	0.64	25	19.86	0.704	0.604	18.843	2.960	14.200	15.223	19.856

Table 5.2 shows that, similar to mass selection, the impact of the Bulmer effect reaches a steady state after 5 generations of selection.

5.4.1 Asymptotic Genetic Variance and Response to Selection

Equations (5.32) and (5.11) can also be used to directly derive steady state parameters, following Dekkers (1992). Assuming for simplicity equal selection in males and females, using equation (5.32), accuracy at the limit is:

$$r^2_{(L)} = 1 - (1-r^2_{(0)})\ \sigma^2_{g_{(o)}} / \sigma^2_{g_{(L)}} \qquad (5.33)$$

Simplifying equation (5.11) for equal selection among males and females, genetic variance at the limit is:

$$\sigma^2_{g_{(L)}} = ½(1-k\ r^2_{(L)})\sigma^2_{g_{(L)}} + ½\,\sigma^2_{g_{(o)}} \qquad (5.34)$$

Rearranging results in:

$$\sigma^2_{g_{(L)}} = \sigma^2_{g_{(o)}} /(1-k\ r^2_{(L)}) \qquad (5.35)$$

and substituting equation (5.33) gives an equation that expresses genetic variance at the limit in terms of parameters for t=0:

$$\sigma^2_{g_{(L)}} = [1+k(1-r^2_{(0)})]\,\sigma^2_{g_{(o)}} /(1+k) \qquad (5.36)$$

Equations (5.36) and (5.33) can then be used to derive response at the limit as:

$$R_{(L)} = i\ r_{(L)}\ \sigma_{g_{(L)}}$$

Response at t=0 is:

$$R_{(0)} = i\ r_{(0)}\ \sigma_{g_{(o)}}$$

Therefore, response at the limit relative to response without accounting for the effect of selection on genetic variance under BLUP selection is equal to:

$$R_{(L)}/R_{(0)} = r_{(L)}\,\sigma_{g(L)} / r_{(0)}\,\sigma_{g(o)}$$

which, using equations (5.36) and (5.33) simplifies to:

$$R_{(L)}/R_{(0)} = \frac{1}{\sqrt{1+k}} \qquad (5.37)$$

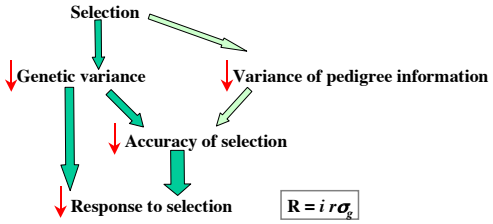
Therefore, the reduction in response under BLUP selection depends only on selection intensity, and not on initial accuracy or heritability, as is the case for mass selection.

Linkage Disequilibrium under BLUP Selection

Predicting Response to Selection

Effect of Linkage Disequilibrium Induced by Selection on Response to Selection

(infinitesimal model, no inbreeding)



BLUP Selection - Example

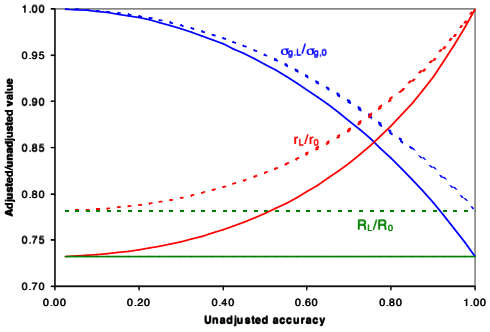
$p_s = p_d = 0.2$ ⇒ $i=1.4$ $k=0.79$

$\sigma_{g,0} = 1$

$r_0^2 = 0.15$ ⇒ PEV=(1-.15)*1	= 0.85	$r_0^2 = 0.80$ ⇒ PEV = 0.85
$\sigma^2_{g,L} = \frac{1+k(1-r_0^2)}{1+k}\sigma^2_{g,0}$	= 0.93	= 0.65
$\sigma^2_{g,L}/\sigma^2_{a,0}$	= 0.93	= 0.65
$r_L^2 = 1 - (1-r_0^2)\frac{\sigma^2_{g,0}}{\sigma^2_{g,L}}$	= 0.09	= 0.69
r_L^2/r_0^2	= 0.60	= 0.86
$R_L/R_0 = \frac{1}{\sqrt{1+k}}$	= 0.747	= 0.747

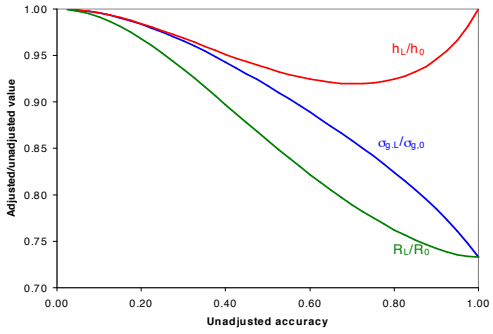
Effect of BLUP Selection on Asymptotic Parameters

$r_s=r_d$ Broken lines: $p_s=p_d=.5$ Solid lines: $p_s=p_d=.05$



Effect of Mass Selection on Asymptotic Parameters

$p_s= p_d=.05$ $r_s=r_d=h$

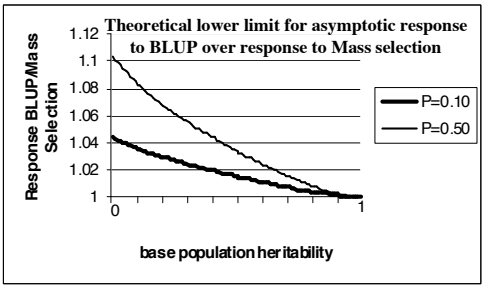


Reduction with BLUP selection > Reduction with Mass selection

Is BLUP selection better than Mass selection at the limit?

(Ignore effect of Inbreeding)

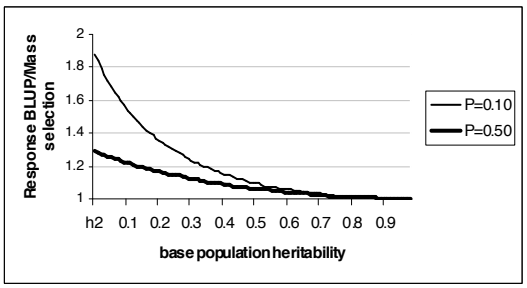
Lower limit to BLUP accuracy: 1 progeny per parent



Comparison of asymptotic Response to BLUP and Mass Selection

(Ignore effect of Inbreeding)

For BLUP accuracy: 2/p progeny per parent



When selection intensities and initial selection accuracy's are different for the two sexes, similar procedures can be used to derive the following results (Dekkers, 1992):

$$R_{(L)}/R_{(0)} = \frac{i_s\sqrt{2\frac{r^2_{s_{(0)}}}{r^2_{d_{(0)}}} - k_d(\frac{r^2_{s_{(0)}}}{r^2_{d_{(0)}}} - 1)} + i_d\sqrt{2 + k_s(1 - \frac{r^2_{s_{(0)}}}{r^2_{d_{(0)}}})}}{(i_s\frac{r_{s_{(0)}}}{r_{d_{(0)}}} + i_d)\sqrt{2 + k_s + k_d}} \qquad (5.38)$$

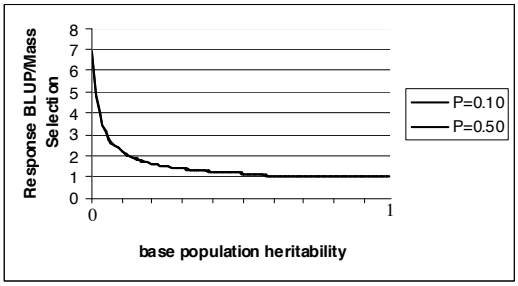
When initial selection accuracy's are equal for both sexes, this equation simplifies to:

$$R_{(L)}/R_{(0)} = \sqrt{\frac{2}{2 + k_s + k_d}} \qquad (5.39)$$

Comparison of asymptotic Response to BLUP and Mass Selection

(Ignore effect of Inbreeding)

For BLUP accuracy: many progeny per parent



Example Programs with Four Selection Paths

Breeding plan		I	II	III	IV
Pathway	p_{SS}	.05	.03	.03	.03
	p_{SD}	.20	.10	.10	.10
	p_{DS}	.05	.01	.01	.01
	p_{DD}	.90	.80	.80	.80
r_s		.85	.85	.85	.85
r_d		.65	.65	.50	.75
$\sigma^2_{Sg,L}/\sigma^2_{g,0}$		.736	.732	.773	.699
$\sigma^2_{Dg,L}/\sigma^2_{g,0}$		.774	.759	.783	.741
$r_{S,L}/r_{S,0}$		.929	.927	.942	.914
$r_{D,L}/r_{D,0}$		.775	.753	.408	.853
R_L/R_0		.759	.743	.685	.752

Relative reductions in response to BLUP due to LD range from 22 to 32% for most practical programs

5.5 Selection Across Multiple Age Groups

The recursive equations developed in the previous sections can be expanded to selection across multiple age groups. Following the notation and derivations of Chapter 2, the genetic mean in year $t+1$ can be predicted as: $\bar{g}_{(t+1)} = \frac{1}{2}\bar{g}^*_{s_{(t)}} + \frac{1}{2}\bar{g}^*_{d_{(t)}}$

where $\bar{g}_{s_{(t)}}$ is the mean genetic value of sires selected at time t, which can be derived as a weighted average of genetic means of selected sires from each age group i at time t:

$$\bar{g}^*_{s_{(t)}} = \frac{1}{P_s}\sum p_{si}\, w_{si}\, \bar{g}^*_{si_{(t)}}$$

with $$\bar{g}^*_{si_{(t)}} = \bar{g}_{si_{(t)}} + i_{si}\, r_{si_{(t)}}\, \sigma_{gi_{(t)}}$$

and $$\bar{g}^*_{d_{(t)}} = \frac{1}{P_d}\sum p_{di}\, w_{di}\, \bar{g}^*_{di_{(t)}}$$

with $$\bar{g}^*_{di_{(t)}} = \bar{g}_{di_{(t)}} + i_{di} r_{di_{(t)}}\, \sigma_{gi_{(t)}}$$

From equation (5.7), genetic variance among selected sires from age group i at time t is equal to:

$$\sigma^{*2}_{g_{si(t)}} = (1 - k_{si}\, r^2_{si_{(t)}})\, \sigma^{\ 2}_{g_{si(t)}}$$

where k_{si} is the variance reduction factor corresponding to the selection intensity among sires of age group i: $k_{si} = i_{si}\,(i_{si} - x_{si})$

Genetic variance among all selected sires at time t is the pooled genetic variance of selected sires within each age group, augmented by the genetic variance between age groups:

$$\sigma^{*2}_{g_{s(t)}} = \frac{1}{P_s} \sum p_{si}\, w_{si}\, \sigma^{*2}_{g_{si(t)}} + \frac{1}{P_s} \sum p_{si}\, w_{si}\, (\bar{g}^{*}_{si_{(t)}} - \bar{g}^{*}_{s_{(t)}})^2 \qquad (5.40)$$

Similarly for dams:

$$\sigma^{*2}_{g_{d(t)}} = \frac{1}{P_d} \sum p_{di}\, w_{di}\, \sigma^{*2}_{g_{di(t)}} + \frac{1}{P_d} \sum p_{di}\, w_{di}\, (\bar{g}^{*}_{di_{(t)}} - \bar{g}^{*}_{d_{(t)}})^2 \qquad (5.41)$$

And genetic variance at time t+1 can be computed using equation (5.8) as:

$$\sigma^{\ 2}_{g_{(t+1)}} = ¼\, \sigma^{*2}_{g_{s(t)}} + ¼\, \sigma^{*2}_{g_{d(t)}} + ½\, \sigma^{\ 2}_{g_0}$$

5.6 Effects of Sample size and Inbreeding

There are two additional factors that affect genetic variance in future generations under the infinitesimal model: sample size and inbreeding. Models to incorporate these effects will be presented in the following sections.

5.6.1 Effect of finite population size on genetic variance

Expected variances derived in the previous sections apply to infinite population sizes. When selecting n individuals out of a population, in addition to the effect of selection on genetic variance, variance is expected to be reduced further by a factor $(1-\frac{1}{n})$. Thus, extending equation (5.7):

$$\sigma^{*2}_{g} = (1-\frac{1}{n})(1 - k\, r^2_{\hat{g}g})\, \sigma^2_g \qquad (5.42)$$

This adjustment is needed because the variances predicted in the previous sections are expected population variances rather than expected sampling variances. Recalling from statistics, sample variance is estimated by dividing sums of squares by n, whereas population variance is estimated by dividing sums of squares by n-1. Thus, to convert an estimate of population variance to an estimate of sample variance, the population variance estimate must be multiplied by $(n\text{-}1)/n = (1-\frac{1}{n})$. It is clear that the impact of this adjustment will be minor for n>50.

5.6.2 Effect of Inbreeding on Genetic Variance

The coefficient of inbreeding of an individual is equal to the probability that two alleles drawn at random from a locus at that individual are identical by descent. Inbreeding, thereby, reduces the variance contributed by Mendelian sampling by a parent by a factor $(1\text{-}F_i)$, where F_i is the

coefficient of inbreeding of the parent. Averaging over all sires and dams that are used for breeding, Mendelian sampling variance contributed to the next generation then is equal to:

$$\sigma^2_{g_{m(t+1)}} = (1 - \tfrac{1}{2}(\overline{F}_{s(t)} + \overline{F}_{d(t)}))\tfrac{1}{2}\sigma^2_{g_{(o)}} \qquad (5.43)$$

where $\overline{F}_{s(t)}$ and $\overline{F}_{d(t)}$ are mean coefficients of inbreeding of sires and dams selected at time *t*.