Introduction, Background, and Mathematical Foundation of Quantitative Genetics

Jack Dekkers\(^1\) and Jean-Luc Jannink\(^2\)

\(^1\) Dept. Animal Science, Iowa State University
\(^2\) Currently at Cornell University, formerly Dept. Agronomy, Iowa State University

Quantitative genetics is the study of continuous traits (such as height or weight) and its underlying mechanisms. It is based on extending the principles of Mendelian and populations genetics to quantitative traits.

Mendelian inheritance:
1. Law of segregation: A trait is influenced by a pair of alleles but each individual only passes a single, random allele on to its progeny.
2. Law of independent assortment: Alleles of different factors combine independently in the gamete.

Population Genetics is the study of the allele frequency distribution and change under the influence of the four evolutionary forces: natural selection, genetic drift, mutation, and migration.

Falconer and Mackay:
“Quantitative genetics theory consists of the deduction of the consequences of Mendelian inheritance when extended to the properties of populations and to the simultaneous segregation of genes at many loci.”

For the purposes of this class: Quantitative genetics = A set of concepts based on the theory of inheritance that help us understand and dissect the genetic basis of quantitative traits and predict what the consequences of different breeding choices will be and therefore allow us to make decisions that lead to the most desirable outcomes.

Quantitative traits
Quantitative genetics covers all traits that are determined by many genes.

- Continuous traits are quantitative traits with a continuous phenotypic range. They are usually polygenic, and may also have a significant environmental influence.

- Traits whose phenotypes are expressed in whole numbers, such as number of offspring, or number of bristles on a fruit fly. These traits can be either treated as approximately continuous traits or as threshold traits.

- Some qualitative traits can be treated as if they have an underlying quantitative basis, expressed as a threshold trait (or multiple thresholds). E.g. diseases that are controlled by multiple traits but for which phenotype is observed as healthy/diseased.

See also Lynch and Walsh Chapter 1 and “Philosophical and Historical Overview.pdf”
Mendelian Genetics

**Discrete traits**
- Theory of heredity at individual locus level
- Inheritance of genes (alleles) at a locus from parent to progeny
  - Law of segregation
  - Law of indep. assortment

Darwin’s Evolutionary Genetics – quantitative traits
- Focus on variation as spice of evolution
- Resemblance among relatives
  - Some heritable variation
- Differential reproductive success
  *But: no clear model of inheritance*
  Progeny resemble parents yet need to differ from parents to maintain variation

Galton’s Biometrical Genetics – quantitative traits
- Progeny resemble their parents
- Regression towards mediocrity
  - Children of tall (short) parents tend to be shorter (taller) than parents
- *But: how is variation maintained?*
  - Johannson: $P = G + E$
  - $G =$ single inherited block
  - Evolution through mutation

Population Genetics

**Individual loci**
- Theory of allele/genotype frequencies in populations
- Theory of changes in frequencies due to evolutionary forces: Natural selection
  - Genetic drift
  - Mutation

Multifactorial model for quantitative traits

George Udny Yule (1902)
- multiple genetic factors + environment
  - continuous variation
  - regression towards mediocrity

Experimental evidence from corn breeders
  - outbreak of variation in F2

Quantitative Genetics Theory
- Theory underlying the inheritance of quantitative traits
- Falconer and McKay: “the deduction of the consequences of Mendelian inheritance when extended to the properties of populations and to the simultaneous segregation of genes at many loci.”
- Theory of population changes in quantitative trait as a result of selection, genetic drift (inbreeding), mutation, migration (crossing)
- A set of concepts based on the theory of inheritance that help us understand and dissect the genetic basis of quantitative traits and predict what the consequences of different breeding choices will be and therefore allow us to make decisions that lead to the most desirable outcomes
“...genetics is meant to explain two apparently antithetical observations – that organisms resemble their parents and differ from their parents. That is, genetics deals with both the problem of heredity and the problem of variation.” Lewontin, 1974.

Francis Galton (1822-1911): regression toward mediocrity – progeny of parents with extreme phenotypes tend to be closer to average.

The modern synthesis of Quantitative Genetics was founded by R.A. Fisher, Sewall Wright, and J.B.S. Haldane, based on evolutionary concepts and population genetics, and aimed to predict the response to selection given data on the phenotype and relationships of individuals.

Analysis of Quantitative trait loci, or QTL, is a more recent addition to the study of quantitative genetics. A QTL is a region in the genome that affects the trait or traits of interest.

Some Basic Quantitative Genetic Concepts and Models

Quantitative genetics dwells primarily on developing theory or mathematical models that represent our understanding of phenomena of interest, and uses that theory to make predictions about how those phenomena will behave under specific circumstances. The model that exists to explain observations of quantitative traits contains the following components:

- Loci that carry alleles that affect phenotype – so-called quantitative trait loci or QTL
- Many such quantitative trait loci
- Alleles at QTL that act in pairs (2 alleles per locus) but that are passed on to progeny individually
- Which of the parent’s alleles are passed on to progeny occurs at random (i.e. a random one of the pair of alleles that a parent has at a locus is passed on to a given progeny), which introduces variability among progeny
- Loci that affect phenotype sometimes show independent assortment (unlinked loci); sometimes not (linked loci)
- Environmental factors influence the trait

In order to develop the quantitative genetic theory and models and to deduce its consequences or predictions it might make, quantitative geneticists have translated these concepts and their behavior into mathematical and statistical terms/models. The most basic model of quantitative genetics is that the phenotypic value ($P$) of an individual is the combined effect of the individual’s genotypic value ($G$) and the environmental deviation ($E$):

$$P = \mu + G + E$$

where $\mu$ is the trait mean

$G$ is the combined effect of all the genes that affect the trait.

$E$ is the combined effect of all environmental effects that affect the phenotype of the individual.

The simplest model to describe inheritance of a quantitative trait (under a lot of assumptions that will be covered later), is that the genotypic value of the offspring can be expressed in terms of the genotypic values of its sire (s) and dam (d), based on the fact that half of the genes that the offspring have come from each parent:

$$G_o = \frac{1}{2} G_s + \frac{1}{2} G_d + RA_s + RA_d$$

Here the terms $RA_s$ and $RA_d$ are random assortment or Mendelian sampling terms, which reflect that parents pass on a random half of their alleles (i.e. a random one of two alleles at each locus).

Developing these quantitative genetic models and deducing their consequences, e.g. the consequences of natural or artificial selection on the trait and the population, then involves manipulating the mathematical terms, that is doing algebra and even a little calculus sometimes (!). Quantitative geneticists were really pioneers in this type of mathematical treatment of biological phenomena and as a result the early growth of quantitative genetics was almost synonymous with the early growth of
statistics. Indeed, R.A. Fisher is hailed as a founder of quantitative genetics but also of analysis of variance and randomization procedures in statistics. The early geneticists Galton and Pearson originated the concepts of regression and correlation. Anyway, the upshot for us here is that we will be deeply involved with the mathematical manipulation and statistical evaluation of our representations of the basic quantitative genetic model. We will review some of the rules of probability and statistics, such as variance, covariance, correlation and regression, and will give a hint at how they may relate to the quantitative genetic model.

Mathematical Foundations for Quantitative Genetics

See also Lynch and Walsh Chapters 2 and 3

Random Variables

In principle, we are interested in the random and non-random processes that determine the value of variables. If the variable of interest is which allele a heterozygous (Mm) father passed on to his daughter for a given marker locus, the rule of random segregation indicates that this is a random process. If the variable of interest is the height of the son of a tall woman, some portion of the variable will be non-random (we expect a relatively tall son) and some portion will be random (we don’t know exactly what the height will be). Either way, we can identify a random variable with a symbol (say $X^p$ to designate the paternally inherited marker allele, or $Y$ to designate height). Common notation is to use capitals for the name of a variable (e.g. $X$ or $Y$) and regular font to represent the value (or class) of that variable. E.g. $X=x$ indicates the event that variable $X$ has value $x$.

Sample Space

The sample space is the set of possible values that a random variable can take. So, for example $X^p \in [M, m]$ (i.e., the progeny inherits either allele $M$ or allele $m$ from its heterozygous $Mm$ father), and $1 < Y < 2.5$ if height is measured in meters. Note that these two example random variables are very different. Random variable $X^p$ can take on just two states (one of the two alleles that the parent has), it is a categorical variable, while $Y$ can take on all values between 1 and 2.5, it is a continuous variable. Nevertheless, many of the mathematical manipulations we will discuss below can be applied equally to either type variable.

Probability (~ frequency)

We designate the probability of an event A as Pr(A). For example, if the event A is “the daughter received marker allele $M$ from her heterozygous $Mm$ father” then $Pr(A) = Pr(X^p = M)$. In this case $Pr(A) = \frac{1}{2}$. The probability function Pr(·) has certain rules assigned to it, just like, for example multiplication has rules assigned to it. For example if event A is “any possible event in the sample space of events” then $Pr(A) = 1$. Thus, the probability that $X^p = M$ or $X^p = m$ for a progeny of a heterozygous $Mm$ father is equal to $\frac{1}{2} + \frac{1}{2} = 1$. Intuitively, though, it is most useful to think of the Pr(A) as the chance that event A will happen. If you look at many events (N events, with N very big) and you count $N_A$, the number of times event A happens, then we can interpret Pr(A) as a frequency, i.e. $Pr(A) = N_A/N$. As examples related to the random variables we gave above, if the father is a heterozygote, then Mendel’s law of segregation say $Pr(X^p = M) = Pr(X^p = m) = \frac{1}{2}$. For the height $Y$ of the son of a tall woman, we can guess that $Pr(1.5 < Y \leq 1.6) < Pr(1.8 < Y \leq 1.9)$, that is, the son is less likely to be in a short ten centimeter bracket than a relatively tall ten centimeter bracket.
**Probability Density** (~ frequency distribution for continuous variables)

The second example leads to the question what is $\Pr(Y = 1.8)$? And the answer, oddly, is zero. That is, given that $Y$ can take on an infinite number of values in the range $[1, 2.5]$, there is a probability of zero that it will take on any specific value. Intuitively, though, we want to be able to express the idea that the chance that the height will be some tall value is greater than the chance it will be some short value. To do this we define the probability density $f(y) = \Pr(y < Y \leq y + \epsilon)/\epsilon$ as $\epsilon$ comes increasingly close to zero. This probability density will be useful to discuss random variables that vary continuously (such as the value of a quantitative trait). Using the probability density function (or pdf) and integration, we can calculate the probability that $Y$ is contained in a certain bracket as

$$\Pr(1.5 < Y \leq 1.6) = \int_{1.5}^{1.6} f(y) dy.$$ 

The most prominent pdf that we will use is that of the normal distribution, i.e. the bell-shaped curve, which is illustrated in Figure 1.

**Expected Value** (~ mean or average)

The expected value of a random variable is a measure of its location in the sample space, and can be thought of as a mean or an average. It takes slightly different forms depending on whether the variable is categorical or continuous. Consider a categorical variable $X$ with sample space $x_1, x_2, \ldots, x_k$. The expected value of $X$ is essentially calculated as a weighted average of the values that $X$ can take on, with weights equal to the probability with which $X$ takes on each value: $E(X) = \sum_{i=1}^{k} x_i \Pr(X = x_i)$.

**Example 1:** The number of florets per spikelet in oat (= variable $X$) is affected by a recessive allele that inhibits development of tertiary kernels (this example is slightly fictitious but serves its purpose). Note that the expected value of a categorical trait may not belong to any of the categories of the trait: the expected value for the number of florets per spikelet is $E(X) = 2.75$ though any given spikelet obviously has a whole number of florets.

**Table 1 Example for computing expectations for a categorical variable**

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Probability (frequency) = $\Pr(X=x_i)$</th>
<th>Number of florets per spikelet $X=x_i$</th>
<th>$x_i \cdot \Pr(X=x_i)$</th>
<th>$x_i^2 \cdot \Pr(X=x_i)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>t / t</td>
<td>0.25</td>
<td>3</td>
<td>0.75</td>
<td>2.25</td>
</tr>
<tr>
<td>T / t</td>
<td>0.50</td>
<td>3</td>
<td>1.50</td>
<td>4.50</td>
</tr>
<tr>
<td>T / T</td>
<td>0.25</td>
<td>2</td>
<td>0.50</td>
<td>1.00</td>
</tr>
<tr>
<td><strong>Sum</strong></td>
<td><strong>1.00</strong></td>
<td><strong>-</strong></td>
<td><strong>E(X) = 2.75</strong></td>
<td><strong>E(X^2) = 7.75</strong></td>
</tr>
</tbody>
</table>
Example 2. Now consider the continuous variable height discussed above. The sample space for $Y$ given was $1 < Y < 2.5$, and the pdf is $f(y) = Pr(y < Y \leq y + e)/e$ as $e$ comes increasingly close to zero. Its expected value is $E(Y) = \int_1^{2.5} y f(y) dy$. Here, instead of multiplying the value of a category by the probability of that category as we did above, we multiply the value by its probability density and integrate over the sample space of the continuous variable. Note that integration is the continuous variable equivalent of summation for categorical variables and the pdf is the equivalent of the probability of each value occurring.

Example 3. Consider again a categorical variable $X$ with sample space $x_1, x_2, \ldots, x_k$. Now assume that there is a function $g(X)$, and we want the expected value of $g(X)$. This expectation is again computed as a weighted average, but now the average of $g(X)$, rather than $X$ itself. The formula for the expectation of $g(X)$ is: $E[\sum_{i=1}^{k} g(x_i) \Pr(X = x_i)]$.

Here, $E$ means that the expectation is taken over all possible values of variable $X$. E.g., referring back to Example 1, the expectation of $g(X) = X^2$ is equal to 7.75, as calculated in the last column in Table 1.

Properties of Expectations

Assuming $X$ and $Y$ are random variables and $a$ is a constant (e.g. $a=5$):

- $E(a) = a$ The expectation of a constant is that constant.
- $E(aX) = aE(X)$ The expectation of the product of a random variable by a constant is the product of the constant and the expectation of the random variable.
- $E(X + Y) = E(X) + E(Y)$ The expectation of a sum of two variables is the sum of their expectations. Note that $E(XY) = E(X)E(Y)$ ONLY IF $X$ and $Y$ are independent – see later.

Joint Probability (~ joint frequency)

The joint probability is the probability for given values of two or more random variables to occur together. The joint probability that random variable $X = x$ and random variable $Y = y$ is denoted $Pr(X = x, Y = y)$.

As an example, assume two genetic loci A and B. The genotypes of a set of individuals are obtained for both loci, resulting in two random variables ($G_A$ and $G_B$). One obtains a table of the joint probability of carrying specific genotypes at each of the two loci:

<table>
<thead>
<tr>
<th>Genotype for locus A ($G_A$)</th>
<th>Genotype for locus B ($G_B$)</th>
<th>Marginal Prob. for $G_A$</th>
</tr>
</thead>
<tbody>
<tr>
<td>bb</td>
<td>Bb</td>
<td>BB</td>
</tr>
<tr>
<td>aa</td>
<td>0.10</td>
<td>0.04</td>
</tr>
<tr>
<td>Aa</td>
<td>0.14</td>
<td>0.18</td>
</tr>
<tr>
<td>AA</td>
<td>0.06</td>
<td>0.10</td>
</tr>
<tr>
<td>Marg.Prob.$G_B$</td>
<td>0.30</td>
<td>0.32</td>
</tr>
</tbody>
</table>

The entries in the body of this table are the joint probabilities. So, for example the joint probability that an individual has genotypes Aa and BB is: $Pr(G_A = Aa, G_B = BB) = 0.16$.  


Marginal probability

Marginal probability is used in Table 2 to show the probabilities of, for example, \( GB = bb \), as the sum down a column of joint probabilities. That is,

\[
Pr(GB = bb) = Pr(GB = bb, GA = aa) + Pr(GB = bb, GA = Aa) + Pr(GB = bb, GA = AA)
\]

\[
= 0.1 + 0.14 + 0.6 = 0.30
\]

What works in the columns for \( GB \) also works in the rows to get marginal probabilities for \( GA \).

In general if \( \{E_1, E_2, \ldots, E_n\} \) is a mutually exclusive and exhaustive set of events (i.e. a set of non-overlapping events that includes the complete parameter space for the variables involved), then marginal probabilities for event \( I \) can be calculated as the sum of joint probabilities of event \( I \) and events \( E_i: \Pr(I) = \sum_{i=1}^{n} \Pr(E_i, I) \)

In Table 2, for example, events \( GA = aa, GA = Aa, \) and \( GA = AA \) are mutually exclusive and exhaustive events and marginal probabilities for \( GB \) can be obtained by summing the joint probabilities in a column of Table 2.

Conditional probability

Intuitively, the conditional probability is the probability of a certain event to occur when you already know that another event is true. Alternately, it is the probability of obtaining a given value for one variable (say, \( X = x \)), conditional on the fact that the value of another variable (say \( Y = y \)) has already been observed. This conditional probability is denoted \( Pr(X = x \mid Y = y) \). First, in order to obtain a given value for \( X \) (say \( X = x \)) while \( Y \) has another value (say \( Y = y \)), both conditions have to hold. So we need the joint probability \( Pr(X = x, Y = y) \). Second, because we know that \( Y = y \), the parameter space for \( X \) is restricted to the subset of events where \( Y = y \). All this to help you intuit the definition of conditional probability:

\[
Pr(X = x \mid Y = y) = \frac{Pr(X = x, Y = y)}{Pr(Y = y)}
\]

In words, the probability of \( X = x \) given \( Y = y \), is the joint probability of \( X = x \) and \( Y = y \) divided by the marginal probability of \( Y = y \).

Referring back to Table 2, the probability of Aa cows having genotype BB is the probability of \( GB = BB \) conditional on \( GA = Aa \), which is:

\[
Pr(GB = BB \mid GA = Aa) = \frac{Pr(GB = BB, GA = Aa)}{Pr(GA = Aa)} = \frac{0.16}{0.48} = 0.333.
\]

One way to interpret this conditional probability is as follows: assuming that we have a total of 100 individuals, then on average 48 (=0.48*100) will be Aa and of those, on average 16 (=0.16*100) will be BB. Thus, the proportion of Aa cows that are BB = 16/48 = 0.333.

Bayes’ Theorem

Sometimes, the conditional probability of \( X \) given \( Y \) is more difficult to derive than the conditional probability of \( Y \) given \( X \). We can then use conditional probabilities to convert one into the other, as follows: \( Pr(X = x \mid Y = y) = \frac{Pr(X = x, Y = y)}{Pr(Y = y)} \). Then, using \( Pr(Y = y \mid X = x) = \frac{Pr(X = x, Y = y)}{Pr(X = x)} \), we can write this as:

\[
Pr(X = x \mid Y = y) = \frac{Pr(Y = y \mid X = x) Pr(X = x)}{Pr(Y = y)}
\]

This is known as Bayes’ Theorem.
For example, suppose somebody tosses a coin three times and gets three heads. What is the probability that this is a double-headed coin, instead of a fair coin?

Let $X$ represent a variable that denotes the state of the coin, i.e. $X = \text{‘double’}$ of $X = \text{‘fair’}$

Let $Y$ represent the data, in our case $Y = 3$ heads in three tosses.

Thus, we are looking for the following conditional probability: $\Pr(X = \text{double} | Y = 3)$

Using Bayes’ theorem, we can also write this as:

$$\Pr(X = \text{double} | Y = 3) = \frac{\Pr(Y = 3 | X = \text{double}) \Pr(X = \text{double})}{\Pr(Y = 3)}$$

Considering each of the three probabilities:

- $\Pr(Y=3|X=\text{double}) = 1$ because every toss will give heads for a double-headed coin
- $\Pr(X = \text{double})$ is known as the ‘prior’ probability of a random coin being double-headed, rather than fair. So what proportion of all coins is double-headed. Let’s say that that is 0.01.
- $\Pr(Y=3)$ is the probability of getting 3 heads out of 3 tosses for a randomly chosen coin, which can be a double-headed coin with prob=0.01 and a fair coin with prob=0.99

Thus $\Pr(Y=3) = \Pr(Y=3|X=\text{double})\Pr(X=\text{double}) + \Pr(Y=3|X=\text{fair})\Pr(X=\text{fair})$

Filling these probabilities into the Bayes’ theorem equation gives:

$$\Pr(X = \text{double} | Y = 3) = \frac{1 \times 0.01}{0.134} = 0.075$$

Statistical independence

Random variable $X$ is statistically independent of $Y$ if the probabilities of obtaining different categories of $X$ are the same irrespective of the value of $Y$.

That is, $\Pr(X = x_i | Y = y_j) = \Pr(X = x_i | Y = y_k) = \Pr(X = x_i)$ for all $i$, $j$, and $k$.

In other words, the conditional probabilities are equal to the marginal probabilities. It follows from the definition of conditional probability that if $X$ is statistically independent of $Y$, the joint probability is equal to the product of their marginal probabilities:

$$\Pr(X = x_i, Y = y_j) = \Pr(X = x_i)\Pr(Y = y_j).$$

For the example in Table 2, $G_A$ and $G_B$ are NOT independent because, e.g.:

$$\Pr(G_B = BB | G_A = Aa) = 0.333 \text{ is NOT equal to } \Pr(G_B = BB) = 0.38.$$  

Also, $\Pr(G_B = BB, G_A = Aa) = 0.16 \text{ is NOT equal to the product of the marginal probabilities:}$

$$\Pr(G_B = BB)\Pr(G_A = Aa) = 038 \times 0.48 = 0.1824$$

Conditional expectation ($\sim$ conditional mean or average)

The expectation (=mean) for variable $X$ conditional on variable $Y$ being equal to $y$ is:

$$E(X|Y = y) = \sum_{i=1}^{k} x_i \Pr(X = x_i | Y = y)$$

and, for continuous variables, $E(X|Y = y) = \int x f(x|Y = y)dx$

So conditional expectation is also computed as a weighted average, but now with weights being equal to the conditional probabilities.
For example, in the oat example of Table 1, consider the expectation for the number of florets per spikelet, conditional on the fact that the line carries at least one T allele. From Table 1, first computing the conditional probabilities:

\[
\Pr(G = T/t | G \text{ contains } T) = \frac{\Pr(G = T/t, G \text{ contains } T)}{\Pr(G \text{ contains } T)} = \frac{0.5}{0.5 + 0.25} = \frac{2}{3}
\]

\[
\Pr(G = T/T | G \text{ contains } T) = \frac{0.25}{0.5 + 0.25} = \frac{1}{3}
\]

Then the conditional expectation is:

\[
E(X | G \text{ contains } T) = 3 \times \frac{2}{3} + 2 \times \frac{1}{3} = \frac{8}{3} = 2.67
\]

Note that this expectation is slightly lower than the overall \(E(X) = 2.75\). So, if we know that the line carries one T allele, we expect the number of florets per spikelet to be slightly lower than average.

**Variance**

The variance of a random variable is a measure of the spread of a variable over the sample space. Intuitively, we want to know how far we can expect the value of a given variable on average to be from its expected value. That is, we want to know something about the average deviation of the random variable from its expected location. The way to obtain a variance is to find the average of the squared deviation from the mean:

\[
\text{var}(Y) = E\{(Y - \mu_Y)^2\} \quad \text{where } \mu_Y = E(Y)
\]

\[
= E\{Y^2 - 2Y \mu_Y + \mu_Y^2\} = E(Y^2) - 2E(Y \mu_Y) + \mu_Y^2 = E(Y^2) - 2 \mu_Y \mu_Y + \mu_Y^2
\]

Thus:

\[
\text{var}(Y) = E\{Y - \mu_Y\}^2 = E(Y^2) - \mu_Y^2
\]

Looking back at Table 1, the number of florets per spikelet given different genotypes,

\[
\text{var}(X) = 7.75 - (2.75)^2 = 0.1875
\]

Note from your statistics class that when we have a sample of \(N\) observations for a random variable \(X\) (instead of frequencies of the variable attaining certain values), the variance of the sample can be computed as:

\[
\text{var}(X) = \frac{\sum_{i=1}^{N} (x_i - \bar{x})^2}{N} \quad \text{or as} \quad \text{var}(X) = \frac{\sum_{i=1}^{N} x_i^2}{N} - \frac{\bar{x}^2}{N}
\]

where \(\bar{x}\) is the average of \(X\).

Realizing that taking the average is sample equivalent to taking the expectation of a variable, note that these equations are similar to the equations for variances based on expectations, as given above.

**Covariance**

The covariance between variables \(X\) and \(Y\) quantifies the (linear) relationship or dependence between \(X\) and \(Y\) based on the extent to which they “co-vary”.

\[
\text{Cov}(X, Y) = E\{(X - \mu_X)(Y - \mu_Y)\}
\]

\[
= E(XY) - \mu_X \mu_Y \quad \text{where } E(XY) = \sum_x \sum_y x_y \Pr(X = x_i, Y = y_j)
\]

Example: The covariance between the genotypic value and the phenotypic value will play a big role in quantitative genetic inferences. Refer back to Table 1, the number of florets per spikelet, conditional on the oat genotype. In Table 1, the genotypic value for the number of florets per spikelet \(G\) is considered the same as the phenotypic value for the number of florets per spikelet \(P\). In that case, the covariance between the genotypic and phenotypic values is equal to the variance of the phenotypic values (0.1875, see above). But consider a slightly more complicated situation in which the environment also contributes to determining the phenotype so that:
Table 3 Example for computing covariances

<table>
<thead>
<tr>
<th>Genotype, $T$</th>
<th>Probability</th>
<th>Genotypic value $G$</th>
<th>Phenotypic value $P$</th>
<th>$Pr(T) \times GP$</th>
<th>$E$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t / t$</td>
<td>0.20</td>
<td>2.8</td>
<td>3</td>
<td>1.68</td>
<td>0.2</td>
</tr>
<tr>
<td>$t / t$</td>
<td>0.05</td>
<td>2.8</td>
<td>2</td>
<td>0.28</td>
<td>-0.8</td>
</tr>
<tr>
<td>$T / t$</td>
<td>0.30</td>
<td>2.6</td>
<td>3</td>
<td>2.34</td>
<td>0.4</td>
</tr>
<tr>
<td>$T / t$</td>
<td>0.20</td>
<td>2.6</td>
<td>2</td>
<td>1.04</td>
<td>-0.6</td>
</tr>
<tr>
<td>$T / T$</td>
<td>0.05</td>
<td>2.2</td>
<td>3</td>
<td>0.33</td>
<td>0.8</td>
</tr>
<tr>
<td>$T / T$</td>
<td>0.20</td>
<td>2.2</td>
<td>2</td>
<td>0.88</td>
<td>-0.2</td>
</tr>
<tr>
<td><strong>Expectation:</strong></td>
<td></td>
<td>2.55</td>
<td>2.55</td>
<td>6.55</td>
<td>0</td>
</tr>
</tbody>
</table>

With this environmental effect, the covariance between genetic and phenotypic values is:

$$\text{Cov}(G, P) = E(GP) - E(G)E(P) = 6.55 - (2.55)^2 = 0.0475.$$  

Check that for this specific example, $\text{Cov}(G,P) = \text{Var}(G) = 0.0475$

The variance of phenotype is greater: $\text{Var}(P) = 0.2475$

The model that relates phenotype to genotype is: $P = G + E$ where $E$ represents the effect of environment. So, for the first row in Table 3 the $E = 3 - 2.8 = +0.2$. For the second row: $E = 2 - 2.8 = -0.8$.

Environmental effects are in the last column of Table 3. Note that $E(E) = 0$. You can also check that:

- $\text{Cov}(G,E) = 0$ (i.e. environmental effects are independent of genetic effects)
- $\text{Cov}(P,E) = 0.2$
- $\text{Var}(E) = 0.2$

Properties of Variance and Covariance

Assuming again that $a$ is a constant:

- $\text{Var}(a) = 0$  The variance of a constant is zero
- $\text{Var}(aX) = a^2\text{Var}(X)$  The variance of the product of a variable by a constant is the product of the constant squared and the variable’s variance
- $\text{Cov}(X,Y) = \text{Cov}(Y,X)$
- $\text{Cov}(X,aY) = a\text{Cov}(X,Y)$
- $\text{Cov}(X,Y+Z) = \text{Cov}(X,Y) + \text{Cov}(X,Z)$
- $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)$  The variance of a sum is the sum of variance plus twice the covariance

(for the Table 3 example: $\text{Var}(P) = \text{Var}(G+E) = \text{Var}(G) + \text{Var}(E) + 2\text{Cov}(G,E) =$

\[= 0.0475 + 0.2 + 2 \times 0 \]  = 0.2475

Generalizing the equation for $\text{Var}(X+Y)$ to the sum of many variables:

- $\text{Var}(\Sigma X_i) = \Sigma \text{Var}(X_i) + 2\Sigma_{i<j}\text{Cov}(X_i, X_j)$  If $X$’s are independent $\Rightarrow \text{Var}(\Sigma X_i) = \Sigma \text{Var}(X_i)$

Also: $\text{Var}(X-Y) = \text{Var}[X+(-Y)] = \text{Var}(X) + \text{Var}(-1\times Y) + 2\text{Cov}[X,(-1\times Y)] =$

\[= \text{Var}(X) + (-1)^2\text{Var}(Y) + 2\times(-1)\times\text{Cov}(X,Y) = \]

\[= \text{Var}(X) + \text{Var}(Y) - 2\text{Cov}(X,Y) \]
**Covariance**

\[
\text{Cov}(X, X) = \text{E}(XX) - \text{E}(X)\text{E}(X) = \text{E}(X^2) - [\text{E}(X)]^2 = \text{Var}(X)
\]

- the covariance of a variable with itself is its variance

**If X and Y are independent:**

\[
\text{E}(XY) = \Sigma \Sigma x_i y_j \text{Pr}(X = x_i, Y = y_j) = \Sigma \Sigma x_i y_j \text{Pr}(X = x_i)\text{Pr}(Y = y_j) = [\Sigma x_i \text{Pr}(X = x_i)] [\Sigma y_j \text{Pr}(Y = y_j)] = \text{E}(X)\text{E}(Y)
\]

So that \(\text{Cov}(X, Y) = \text{E}(XY) - \text{E}(X)\text{E}(Y) = 0\)

**Correlation**

The correlation measures the (linear) relationship between two variables on a standardized scale, by dividing their covariance by the product of their standard deviations:

\[
\text{r}_{XY} = \text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}
\]

Note that: \(-1 \leq \text{r}_{XY} \leq 1\)

For the example of Table 3: \(r_{GP} = \frac{\text{Cov}(G, P)}{\sqrt{\text{Var}(G)\text{Var}(P)}} = \frac{0.0475}{\sqrt{0.0475*0.2475}} = 0.438\)

Based on rearrangement of the correlation equation, we get the following expression for the covariance, which we also frequently use:

\[
\text{Cov}(X, Y) = r_{XY} \sqrt{\text{Var}(X)\text{Var}(Y)}
\]

**Regression**

A repeated theme in quantitative genetics is the estimation of quantities associated with individuals or parameters associated with populations when those quantities or parameters are themselves not directly observable. The most obvious example is the desire to estimate an individual’s genotypic value for a trait when the only information we have available derives from the individual’s phenotype. Regression is used for this kind of estimation.

**Definition:** The regression of \(Y\) on \(X\) is the expected value of \(Y\) conditional on having a certain value for variable \(X\):

\[
\hat{y} = \text{E}(Y|X)
\]

This is also called the **best (linear) predictor** of \(Y\) given \(X\).

Regression can be used to define a model:

\[
y = \hat{y} + e\quad \text{where } e \text{ is called the residual, which is the deviation of the observed value for } Y \text{ from its expected value conditional on } X.
\]
For quantitative variables, the predicted value for $Y$ can be derived using linear regression:

$$\hat{y} = E(Y|X) = \mu_Y + b_{yx}(x-\mu_X)$$

with

- $\mu_Y = E(Y)$
- $\mu_X = E(X)$
- $b_{yx} = \text{coefficient of regression of } Y \text{ on } X = \text{expected change in } Y \text{ per 1 unit increase in } X$

Given data, $b_{yx}$ can be derived by fitting the following linear regression model:

$$y = \mu_Y + b_{yx}(x-\mu_X) + e$$

Using least squares (see Lynch & Walsh p39), $b_{yx}$ can be derived to be equal to:

$$b_{yx} = \frac{\text{Cov}(Y,X)}{\text{Var}(X)}$$

Note that $b_{yx}$ can also be expressed in terms of the correlation coefficient:

$$b_{yx} = \frac{\text{Cov}(Y,X)}{\text{Var}(X)} = r_{xy} \sqrt{\text{Var}(Y)\text{Var}(X)} / \text{Var}(X) = r_{xy} \frac{\text{Var}(Y)}{\text{Var}(X)}$$

So the important equations to remember for the regression coefficient are:

$$b_{yx} = \frac{\text{Cov}(Y,X)}{\text{Var}(X)} = r_{xy} \frac{\text{Var}(Y)}{\text{Var}(X)}$$

Note that these only hold for simple regression with a single independent variable ($X$).

For the example of Table 3, suppose we want to predict the genotypic value of an individual based on its observed phenotypic value. We would use the following regression model:

$$G = \bar{G} + b_{GP}(P - \bar{P}) + e$$

with $\bar{G} = E(G) = E(P) = 2.55$

The regression coefficient can be computed as:

$$b_{GP} = \frac{\text{Cov}(G,P)}{\text{Var}(P)} = \frac{0.0475}{0.2475} = 0.192$$

or

$$b_{GP} = r_{GP} \sqrt{\frac{\text{Var}(G)}{\text{Var}(P)}} = 0.438 \sqrt{\frac{0.0475}{0.2475}} = 0.192$$

So the prediction model is: $\hat{G} = \bar{G} + b_{GP}(P - \bar{P}) = 2.55 + 0.192(P - 2.55)$.

Results are in Table 4. The last column in this table shows the prediction error: $\hat{e} = G - \hat{G}$

<table>
<thead>
<tr>
<th>Genotype, $T$</th>
<th>Probability</th>
<th>$G$</th>
<th>$P$</th>
<th>$E$</th>
<th>$\hat{G}$</th>
<th>$\hat{e}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>t / t</td>
<td>0.20</td>
<td>2.8</td>
<td>3</td>
<td>0.2</td>
<td>2.636</td>
<td>0.164</td>
</tr>
<tr>
<td>t / t</td>
<td>0.05</td>
<td>2.8</td>
<td>2</td>
<td>-0.8</td>
<td>2.444</td>
<td>0.356</td>
</tr>
<tr>
<td>T / t</td>
<td>0.30</td>
<td>2.6</td>
<td>3</td>
<td>0.4</td>
<td>2.636</td>
<td>-0.036</td>
</tr>
<tr>
<td>T / t</td>
<td>0.20</td>
<td>2.6</td>
<td>2</td>
<td>-0.6</td>
<td>2.444</td>
<td>0.156</td>
</tr>
<tr>
<td>T / T</td>
<td>0.05</td>
<td>2.2</td>
<td>3</td>
<td>0.8</td>
<td>2.636</td>
<td>-0.436</td>
</tr>
<tr>
<td>T / T</td>
<td>0.20</td>
<td>2.2</td>
<td>2</td>
<td>-0.2</td>
<td>2.444</td>
<td>-0.244</td>
</tr>
</tbody>
</table>

Expectation: 2.55 2.55 0 2.55 0.0004
Properties of Regression

1. The average of predicted values is equal to the average of Y’s: $E(\hat{Y}) = E(Y) = \mu_Y$

   $$E(\hat{Y}) = E[\mu_Y + b_{yx}(x-\mu_X)] = E(\mu_Y) + E[b_{yx}(x-\mu_X)] = \mu_Y + b_{yx}[E(x)-\mu_X] = \mu_Y$$

   This also implies that the regression line always passes through the mean of both $X$ and $Y$; substituting $\mu_X$ for $x$ into the prediction equation gives $\hat{y} = \mu_Y$

2. The average value of the residual is zero: $E(e) = 0$.

   $$E(e) = E(Y - \hat{Y})$$
   from regression model
   $$= E(Y) - E(\hat{Y})$$
   property of expectation
   $$= 0$$
   from property 1 above

3. The expectation of the residual is zero for all values of $X$: $E(e|X) = 0$

   $$E(e|X) = E(Y - \hat{Y}|X)$$
   from regression model
   $$= E(Y|X) - E(\hat{Y}|X)$$
   property of expectation
   $$= \hat{Y} - \hat{Y} = 0$$
   by definition of regression

   This implies that predictions of $Y$ are on average equal to the true $Y$ across the range of possible values for $X$.

4. Accuracy of prediction = Corr($\hat{Y}, Y$) = $r_{\hat{Y}Y}$

   The accuracy of the prediction equation is equal to the correlation of $\hat{Y}$ with its true value $y$. We can derive accuracy as:

   $$\text{Accuracy} = r_{\hat{Y}Y} = \frac{\text{Cov}(\hat{Y}, Y)}{\sqrt{\text{Var}(\hat{Y})\text{Var}(Y)}} = \frac{\text{Cov}(\mu_Y + b_{yx}(x-\mu_X), y)}{\sqrt{\text{Var}(\mu_Y + b_{yx}(x-\mu_X))\text{Var}(y)}}$$

   Since $\mu_Y$ and $\mu_X$ are constants, this simplifies to:

   $$\text{Accuracy} = r_{\hat{Y}Y} = \frac{b_{yx}\text{Cov}(x, y)}{\sqrt{b_{yx}^2\text{Var}(x)\text{Var}(y)}} = \frac{\text{Cov}(x, y)}{\sqrt{\text{Var}(x)\text{Var}(y)}} = r_{XY}$$

   So the accuracy of a prediction equation based on simple (= 1-variable) regression is equal to the correlation between the dependent and independent variables.

5. Decomposition of variance in $Y$ into that explained by the prediction and unexplained variance

   Using the above equation, we can also show that the variance of $Y$ is the sum of the variance explained by the regression on $X$ and residual variance (note that Cov($X,e)=0$):

   $$\text{Var}(y) = \text{Var}(\mu_Y + b_{yx}(x-\mu_X) + e) = b_{yx}^2\text{Var}(x) + \text{Var}(e) = [\text{Cov}(y,x)]^2/\text{Var}(x) + \text{Var}(e)$$

   Note that because Cov($y,x$) = $r_{XY} \sqrt{\text{Var}(x)\text{Var}(y)}$ the first term can also be written as:

   $$[\text{Cov}(y,x)]^2/\text{Var}(x) = r_{XY}^2 \text{Var}(x) \text{Var}(y) / \text{Var}(x) = r_{XY}^2 \text{Var}(y) = r_{\hat{Y}Y}^2 \text{Var}(y)$$

   This is the variance in $Y$ that is explained by the $X$ through the prediction model.

   By subtraction we get $\text{Var}(e) = [1 - r_{XY}^2] \text{Var}(y)$. This is the unexplained/residual variance.

   Thus, variance of $Y$ can be decomposed as: $\text{Var}(y) = r_{XY}^2\text{Var}(y) + [1 - r_{XY}^2] \text{Var}(y)$

   Note that the variance of predicted values is equal to the explained variance:
\[
\text{Var}(\hat{y}) = \text{Var}[\mu_Y + b_{YX}(x-\mu_X)] = b_{YX}^2 \text{Var}(x) = \left\{ \frac{\text{Cov}(y,x)}{\text{Var}(x)} \right\}^2 \text{Var}(x) = \left[ \frac{\text{Cov}(y,x)}{\text{Var}(x)} \right]^2 \text{Var}(x) = \frac{\text{Cov}(y,x)^2}{\text{Var}(x) \text{Var}(y)} \text{Var}(y) = r_{YX}^2 \text{Var}(y)
\]

So the variance of predicted values is equal to the variance explained by the model, which depends on the correlation between \(Y\) and \(X\).

The above equations apply when prediction is based on one variable \((x)\), in which case \(r_{yY} = r_{XY}^2\).

In general, prediction can be based on multiple \(x\)’s = multiple regression. In that case the partitioning of variance is: \(\text{Var}(\hat{y}) = \text{Var}(\hat{y}) + \text{Var}(e) = r_{Yy}^2 \text{Var}(y) + [1-r_{Yy}^2] \text{Var}(y)\)

6. Residuals are uncorrelated with the predictor variable, \(X\):\[
\text{Cov}(X,e) = 0
\]
\[
\text{Cov}(x,e) = \text{Cov}[x, y - \hat{y}] = \text{Cov}[x, y -(\mu_Y + b_{YX}(x-\mu_X))] = \\
= \text{Cov}(x,y) - \text{Cov}(x,\mu_Y) - b_{YX} \text{Cov}(x,x) - b_{YX} \text{Cov}(x,\mu_X) \\
= \text{Cov}(x,y) - 0 - b_{YX} \text{Var}(x) - 0 \\
= \text{Cov}(x,y) - \frac{\text{Cov}(y,x)}{\text{Var}(x)} \text{Var}(x) = 0
\]

7. Residuals are uncorrelated with the predictions:\[
\text{Cov}(\hat{y},e) = 0
\]
\[
\text{Cov}(\hat{y},e) = \text{Cov}[\hat{y}, y - \hat{y}] = \text{Cov}(\hat{y}, y) - \text{Var}(\hat{y}) = \\
= \text{Cov}(x,y) - \text{Cov}(x,\mu_Y) - b_{YX} \text{Cov}(x,x) - b_{YX} \text{Cov}(x,\mu_X) \\
= \text{Cov}(x,y) - 0 - b_{YX} \text{Var}(x) - 0 \\
= \text{Cov}(x,y) - \frac{\text{Cov}(y,x)}{\text{Var}(x)} \text{Var}(x) = 0
\]

Properties 6 and 7 imply that all information on \(Y\) that is contained in \(X\) is captured in the predicted values, as the residual is uncorrelated to both \(X\) and the predicted values.

Some Applications to Quantitative Genetic Theory

The standard quantitative genetics model equation for the observed phenotype of an individual \(i\) for a quantitative trait \((P_i)\) is that it is the sum of the effect of genetics (the genotypic value \(G_i\)) and the effect of environment \((E_i): P_i = \mu_P + G_i + E_i\)

In practice, we only observe phenotype and cannot directly observe \(G_i\) or \(E_i\). However, if we could observe both \(P_i\) and \(G_i\) for a group of individuals, we could plot genotypic against phenotypic values, as in the figure below.

Using such a hypothetical plot, or model, and statistics such as correlation, covariance, variance, and regression, we can specify important population parameters such as heritability \((h^2)\) and make a number of inferences or predictions, such as predicting an individual’s genotypic value or ‘breeding value’ from its observed phenotype:
1) Covariance and correlation between phenotypic and genotypic values:
Based on \( P_i = \mu_P + G_i + E_i \)
\[
\text{Cov}(P, G) = \text{Cov}(\mu + G + E, G) = \\
= \text{Cov}(G, G) + \text{Cov}(G, E) = \text{Var}(G)
\]
The last step assumes that \( \text{Cov}(G, E) = 0 \), i.e. that the environment that an individual receives is independent of its genotypic value. The result of this covariance, \( \text{Var}(G) \), which is often denoted \( \sigma_G^2 \), is the \textit{genetic variance} in the population, i.e. the variance of genotypic values of individuals in a population. This in contrast to the phenotypic variance, \( \text{Var}(P) \), often denoted \( \sigma_P^2 \), which is the variance of phenotypic values of individuals in a population.

Then, the correlation between phenotypic and genotypic values can be derived as:
\[
r_{P,G} = \frac{\text{Cov}(G, P)}{\sqrt{\text{Var}(G)\text{Var}(P)}} = \frac{\sigma_G^2}{\sqrt{\sigma_G^2\sigma_P^2}} = \frac{\sigma_G}{\sigma_P}
\]
Thus, the correlation between genotypic and phenotypic values of individuals in a population is equal to the ratio of the genetic and phenotypic standard deviations for the trait.

The square of this correlation, therefore, is equal to the ratio of the genetic and phenotypic variances, or to the proportion of phenotypic variance that is genetic. This proportion is also defined as the \textit{heritability} of the trait (= \( h^2 \)). Thus:
\[
(r_{P,G})^2 = \frac{\sigma_G^2}{\sigma_P^2} = h^2
\]

2) Regression of genotypic on phenotypic values:
Using the above model and referring to the figure, we can also set up a regression equation between the genotypic and phenotypic values to predict \( G \):
\[
G_i = \mu_G + b_{G,P}(P_i - \mu_P) + e_i \quad \text{where } b_{G,P} \text{ is the coefficient of regression of } G \text{ on } P.
\]
This regression coefficient can be derived as:
\[
b_{G,P} = \frac{\text{Cov}(G, P)}{\text{Var}(P)} = \frac{\sigma_G^2}{\sigma_P^2} = h^2
\]
Thus, the slope of the regression of genotypic on phenotypic values is equal to heritability

3) Prediction of genotypic values:
The above regression model can be used to predict an individual’s genotypic value based on it’s observed phenotype, using the following prediction equation:
\[
\hat{G}_i = \mu_G + h^2(P_i - \mu_P)
\]
In practice, we often set \( \mu_G \) to zero, because we’re primarily interested in ranking individuals in a population. Thus:
\[
\hat{G}_i = h^2(P_i - \mu_P)
\]
As an \textit{example} (see figure), assume a dairy cow produces 6500 kg milk, which is its phenotypic value \( (P) \). The mean production of the herd she is in is 6000 kg \( (= \mu_P) \).

Milk production is a trait with an (assumed known) heritability of 0.3, a phenotypic standard deviation of 1200 kg \( (\sigma_P=1200) \). Using \( h^2 = \frac{\sigma_G^2}{\sigma_P^2} \) and, thus, \( \sigma_G^2 = h^2\sigma_P^2 \), the genetic standard deviation for milk yield is equal to \( \sigma_G = h\sigma_P = \sqrt{0.3 \times 1200} = 657.3 \text{ kg} \)

Then, this cow’s genotypic value can be predicted to be:
\[
\hat{G}_i = h^2(P_i - \mu_P) = 0.3 \times (6500-6000) = +150 \text{ kg}
\]
So this cow’s genotypic value is expected to be 150 kg greater than the average in this herd.
We can also attach an accuracy to this prediction, based on the previously derived result that the correlation between predicted and true values based on linear regression is equal to the correlation to the dependent (Y) and independent (X) variables:  \( r_{G,G} = r_{G,P} = h \)

Thus, when predicting an individual’s genotypic value based on its phenotypic value, the accuracy of this prediction will be equal to the square-root of heritability of the trait.

When we predict genotypic values for all individuals in a population in this manner, and take the variance of these predicted values, we expect this variance to be equal to (based on property 5):

\[
\text{Var}(\hat{y}) = r_{XY}^2 \text{Var}(y)
\]

which in this case simplifies to:  \( \text{Var}(\hat{G}) = h^2 \sigma_G^2 = h^4 \sigma_p^2 \)

And, using property 5 above, the variance of prediction errors \( (e_i = G - \hat{G}) \) is equal to:

\[
\text{Var}(e) = (1 - r_{XY}^2) \text{Var}(y)
\]

which in this case simplifies to:  \( \text{Var}(e) = (1-h^2) \sigma_G^2 \)

For the example, the variance of predicted values is:  \( \text{Var}(\hat{G}) = h^2 \sigma_G^2 = 0.7 \times 657.32 = 129600 \)

and the variance of prediction errors is:  \( \text{Var}(e) = (1-h^2) \sigma_G^2 = 0.3 \times 657.32 = 302400 \)

Note that these two variances sum to the genetic variance:  \( 129600 + 302400 = 432040 = 657.32 \)

Based on \( \text{Var}(e) = 302400 \text{kg}^2 \) we can also add a confidence interval to our prediction (see later).

4) **Regression of offspring phenotype on parent phenotype**

One of the problems with predicting genotypic values, as described above, is that it requires you to know the heritability of the trait. Luckily, we can also get estimates of heritability for a trait from phenotypic data. We do this by observing how similar the phenotype of offspring is to that of their parents; if these are very similar, we expect the trait to be more heritable.

When we have phenotypes observed on offspring and their sires, we can estimate heritability by regressing the phenotype of the offspring on that of their parents, as illustrated below:

The regression model is:

\[
P_o = \mu_o + b_{PoPs}(P_s-\mu_s) + e
\]

The regression coefficient can be derived as:

\[
b_{PoPs} = \frac{\text{Cov}(P_o, P_s)}{\text{Var}(P_s)} = \frac{\text{Cov}(G_o + E_o, G_s + E_s)}{\sigma_p^2} = \frac{\text{Cov}(G_o, G_s) + \text{Cov}(G_o, E_s) + \text{Cov}(E_o, G_s) + \text{Cov}(E_o, E_s)}{\sigma_p^2} = \frac{\text{Cov}(G_o, G_s)}{\sigma_p^2}
\]

The last step assumes that the environment that the offspring progeny received is independent of the phenotype of the sire (a sometimes strong assumption), making the last 3 covariance terms 0.

To derive the covariance between the genotypic value of offspring and that of their sire, we can express the genotypic value of the offspring in terms of the genotypic values of its sire (s) and dam (d), based on the fact that half of the genes that the offspring have come from each parent:
\[
G_o = \frac{1}{2} G_s + \frac{1}{2} G_d + RA_s + RA_d
\]

Here the terms \( RA_s \) and \( RA_d \) are random assortment or Mendelian sampling terms, which reflect that parents pass on a random half of their alleles (i.e. a random one of two alleles at each locus). Using this genetic model (which has quite a number of assumptions, which well be covered later), we can continue our derivation as:

\[
\text{Cov}(G_o, G_s) = \text{Cov}(\frac{1}{2}G_s + \frac{1}{2}G_d + RA_s + RA_d, G_s) = \\
= \text{Cov}(\frac{1}{2}G_s, G_s) + \text{Cov}(\frac{1}{2}G_d, G_s) + \text{Cov}(RA_s, G_s) + \text{Cov}(RA_d, G_s)
\]

Assuming random mating and the fact that Mendelian sampling terms are independent (see later), the last three covariance terms are zero, resulting in:

\[
b_{PoPs} = \frac{\text{Cov}(G_o, G_s)}{\sigma_p^2} = \frac{\text{Cov}(\frac{1}{2}G_s, G_s)}{\sigma_p^2} = \frac{\frac{1}{2}\sigma_g^2}{\sigma_p^2} = \frac{1}{2}h^2
\]

Thus, heritability of a trait can be estimated based on phenotypes of relatives, by measuring the degree of resemblance between relatives, using statistics such as linear regression. More on this later.

**Some Distributions useful in Population and Quantitative Genetics**

**Bernoulli distribution.**

Named after the mathematician Daniel Bernoulli, 1700-1782. A Bernoulli random variable is characterized by one parameter, that is typically designated \( p \) and is sometimes called the “probability of success”. The random variable can have one of two values: 1 with probability \( p \) and 0 with probability \( 1 - p \).

If \( Y \) is a Bernoulli random variable with probability \( p \), its expectation is:

\[
E(Y) = \sum_{i=1}^{2} y_i \Pr(Y = y_i) = 0(1 - p) + 1(p) = p
\]

Its variance is

\[
\text{var}(Y) = E(Y^2) - E(Y)^2 = [0^2(1 - p) + 1^2(p)] - p^2 = p - p^2 = p(1 - p)
\]

The Bernoulli distribution is used in population and quantitative genetics in relation to the presence or inheritance of alleles at a locus. For example, for a locus with two possible alleles, \( A \) and \( a \), and with the frequency of allele \( A \) in the population denoted by \( p \), then the process of drawing one allele at this locus from a population can be specified by a Bernoulli distribution by specifying a variable \( Y \) that is equal to 1 if allele \( A \) is drawn and equal to 0 if allele \( a \) is drawn.

**Binomial Distribution**

The Binomial distribution is based on the Bernoulli distribution. A binomial random variable is the sum of \( k \) independent Bernoulli random variables all with parameter \( p \). The binomial is therefore characterized by two parameters, \( k \) and \( p \) and can have integer values from 0 to \( k \). If \( X \) is binomially distributed with \( k \) trials and \( p \) probability of success: \( X \sim \text{Binomial}(k, p) \), then:

From the properties of expectation of a sum, the **expected value** of \( X \) is \( kp \): \( E(X) = kp \).

From the properties of variance of a sum of independent variables, the **variance** of \( X \) is \( kp(1 - p) \): \( \text{var}(X) = kp(1 - p) \)
The probability density function \( \Pr(X = x) \) is

\[
\Pr(X = x) = \binom{k}{x} p^x (1 - p)^{k-x} \text{ where } \binom{k}{x} = \frac{k!}{x!(k-x)!} \text{ and } a! = 1*2*3*...*a
\]

When considering population or quantitative genetics, the Binomial Distribution could correspond to the process of randomly drawing \( k \) alleles at a locus from a population.

**Normal or Gaussian distribution.**  
This is perhaps the most important distribution in quantitative genetics, as phenotypes for most quantitative traits approximately follow a normal distribution, or can be transformed to follow a normal distribution. This is a property of the fact that phenotype is the sum of many genetic factors and of many environmental factors. Following the Central Limit Theorem of statistics, this is expected to result in a Normal distribution, even if the distribution of variables that are included in the sum is not Normal. See also Falconer and MacKay Chapter 6.

The probability distribution function for a variable \( y \) that has a Normal distribution with mean \( \mu \) and standard deviation \( \sigma \), denoted by \( y \sim N(\mu, \sigma^2) \) is:

\[
\Pr(y) = z = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y - \mu)^2}{2\sigma^2}}
\]

It is often useful to work with the *Standard* Normal distribution, which has mean zero and standard deviation 1: \( N(0,1) \) Any Normally distributed variable can be ‘standardized’ to a variable that follows \( N(0,1) \) by subtracting the mean and dividing by the standard deviation:

If \( y \sim N(\mu, \sigma^2) \) then \( y' = (y - \mu)/\sigma \) follows \( N(0,1) \)

**Truncated Normal distribution.**  
In plant and animal breeding, we often are interested in using individuals with the highest phenotype for breeding. If phenotype \( (y) \) is Normally distributed \( (y \sim N(\mu, \sigma^2)) \) then it is of interest to know something about the distribution of phenotypes of the selected individuals. This is the Truncated Normal distribution, as illustrated in the figure above:
Selecting a proportion \( p \) of individuals from a population based on phenotype \( (y) \) is equivalent to truncating the Normal distribution at a truncation point \( T \), such that a fraction \( p \) falls above the truncation point.

The mean phenotype for the selected individuals is denoted by \( \mu_S \) (see Figure).

The difference between the mean of the selected individuals over that of all individuals is called the selection differential:
\[
S = \mu_S - \mu
\]

**Maximum Likelihood Estimation**

Maximum Likelihood (ML) is a procedure for estimating parameters from an observed set of data. It was introduced by Fisher and is widely used in population and quantitative genetics.

The basic idea of ML estimation is to find the value of the parameter(s) that is ‘most likely’ to have produced the data that is observed, i.e. that maximizes the likelihood of getting the data that you got.

As a simple example to illustrate the concept of ML estimation, consider the following observed genotype frequencies.

**Table 1. Falconer and Mackay, p. 1, blood group categories in Iceland:**

<table>
<thead>
<tr>
<th>Blood Group</th>
<th>Counts</th>
<th>Probabilities</th>
</tr>
</thead>
<tbody>
<tr>
<td>MM</td>
<td>233</td>
<td>( P = \frac{233}{747} = 0.312 )</td>
</tr>
<tr>
<td>MN</td>
<td>385</td>
<td>( H = \frac{385}{747} = 0.515 )</td>
</tr>
<tr>
<td>NN</td>
<td>129</td>
<td>( Q = \frac{129}{747} = 0.173 )</td>
</tr>
<tr>
<td>Total</td>
<td>747</td>
<td>( \frac{747}{747} = 1.000 )</td>
</tr>
</tbody>
</table>

\( P, H, \) and \( Q \) are the estimated genotype frequencies – obtained by counting.

To estimate allele/gene frequencies, we could obtain these simply by counting: 2 * 747 alleles were sampled; the number of M alleles is \( (2P + H) * 747 \). Thus, the allele/gene frequency of allele M is
\[
p = \frac{(2P + H) * 747}{2 * 747} = P + \frac{1}{2} H
\]

So \( p = 0.312 + 0.515 / 2 = 0.57 \) and \( q = 0.173 + 0.515 / 2 = 0.43 \).

This estimates of allele frequency obtained by counting is actually an ML estimate: for the example of Table 1, if 57% of all alleles in the sample is M (vs. N, as is observed in the sample), then the ML estimate of \( p \), the frequency of M in the population that the sample came from, is 0.57, because that is the value of \( p \) that is most likely to have produced a sample with 57% of alleles being M.

A more formal derivation of this estimate uses the Binomial distribution to specify the Likelihood of the data as a function of the parameter (= Likelihood function): if out of \( n \) alleles sampled \( n_M \) are M, then the likelihood to get these counts given the population frequency of M is equal to the probability that the value of a Binomial variable with parameters \( n \) and \( p \) is equal to \( n_M \):
\[
Likelihood(\text{data} | p) = Pr(\text{data} | p) = \binom{n}{n_M} p^{n_M} (1 - p)^{n-n_M}
\]

For the data in Table 1 \( n = 2*747 = 1494 \) and \( n_M = 2*233+385 = 851 \)

So:
\[
Likelihood(\text{data} | p) = \binom{1494}{851} p^{851} (1 - p)^{643}
\]
Now the ML estimate of $p$ is the value of $p$ that maximizes the above function. To find this value we can take the first derivative of the Likelihood and set it equal to zero. However, it is often easier to first take the natural log of the Likelihood and to maximize it for $p$:

$$L (\text{data} \mid p) = \ln \left( \binom{n}{n_M} p^{n_M} (1 - p)^{n - n_M} \right)$$

Then, using some algebra, this can be ‘simplified’ to:

$$L (\text{data} \mid p) = \ln \left( \binom{n}{n_M} \right) + \ln(p^{n_M}) + \ln[(1 - p)^{n - n_M}] =$$

$$= \ln \left( \binom{n}{n_M} \right) + n_M \ln(p) + (n - n_M) \ln(1 - p)$$

The first derivative of the LogLikelihood with respect to $p$ is:

$$n_M \frac{1}{p} + (n - n_M) \frac{-1}{1 - \hat{p}} = 0 \Rightarrow n_M \frac{1}{\hat{p}} = (n - n_M) \frac{1}{1 - \hat{p}} \Rightarrow 1 - \hat{p} = \frac{n - n_M}{n_M} \frac{n}{n_M}$$

$$\Rightarrow \frac{1}{\hat{p}} - 1 = \frac{n}{n_M} - 1$$

$$\Rightarrow \hat{p} = \frac{n_M}{n}, \text{ i.e. count estimate.}$$

This is obviously a simple example, where we don’t need ML estimation to obtain a good estimate (we can just count).

Another (obvious) example is the following: Suppose $n$ values, $y_1, y_2, \ldots, y_n$, are sampled independently from an underlying Normal distribution with unknown mean $\mu$ and variance 1. What is the MLE for $\mu$ given the data?

Let’s denote the data by a vector $y = (y_1, y_2, \ldots, y_n)$. Using the probability density function of the Normal distribution with mean $\mu$ and standard deviation 1, the likelihood for a given data point $y_i$ given the mean, $\mu$, is: $\text{Likelihood}(y_i \mid \mu) = \text{Pr}(y_i \mid \mu) = \frac{1}{\sqrt{2 \pi}} e^{-\frac{(y_i - \mu)^2}{2}}$ Because each observation is independent, the likelihood function for all observation $y$ is the product of $n$ normal density functions:

$$\text{Likelihood}(y \mid \mu) = \text{Pr}(y \mid \mu) = \prod_{i=1}^{n} \frac{1}{\sqrt{2 \pi}} e^{-\frac{(y_i - \mu)^2}{2}} = \left(2 \pi \right)^{-n/2} \sum_{i=1}^{n} e^{-\frac{(y_i - \mu)^2}{2}}$$

Again, taking the natural log of the likelihood:

$$L(y \mid \mu) = -\left(\frac{n}{2}\right) \ln(2\pi) - \frac{1}{2} \sum_{i=1}^{n} (y_i - \mu)^2$$
Maximizing by taking the first derivative gives:

$$\frac{\partial L(\mu | y)}{\partial \mu} = \sum_{i=1}^{n} (y_i - \mu) = n(\bar{y} - \mu)$$

where \( \bar{y} \) is the average of the observations

Setting this equal to zero gives: \( n(\bar{y} - \mu) = 0 \) \( \rightarrow \) the MLE of \( \mu \) is: \( \hat{\mu} = \bar{y} \)

Again, this is obvious but it does illustrate the principle behind the use of ML to estimate parameters in more complex situation. For example, if we want to estimate a parameter such as heritability from data \( (y) \) we have observed in a pedigreed population, we can formally state the problem by that of finding the MLE of heritability, given the observed data; i.e. what is the most likely value of heritability that would have given rise to the data that we observed. To do this, we need to formulate the Likelihood function, or the log of the likelihood, and maximize it.

$$Likelihood( data | h^2 ) = Pr(data | h^2 )$$

This is the basis of ML procedures for estimation of genetic parameters.
A Review of Elementary Matrix Algebra

Notes developed by John Gibson for Economics of Animal Breeding Strategies notes
(Dekkers, Gibson, van Arendonk)

Dr. B.W. Kennedy originally prepared this review for use alongside his course in Linear Models in Animal Breeding. His permission to use these notes is gratefully acknowledged. Not all the operations outlined here are necessary for this course, but most would be necessary for some applications in animal breeding.

A much more complete treatment of matrix algebra can be found in "Matrix Algebra Useful for Statistics" by S.R. Searle. See also Chapter 8 of Lynch and Walsh.

A.1 Definitions

A matrix is an ordered array of numbers. For example, an experimenter might have observations on a total of 35 animals assigned to three treatments over two trials as follows:

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Trial</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6 4</td>
</tr>
<tr>
<td>2</td>
<td>3 9</td>
</tr>
<tr>
<td>3</td>
<td>8 5</td>
</tr>
</tbody>
</table>

The array of numbers of observations can be written as a matrix as

\[ M = \begin{bmatrix} 6 & 4 \\ 3 & 9 \\ 8 & 5 \end{bmatrix} \]

with rows representing treatments (1,2,3) and columns representing trials (1,2).

The numbers of observations then represent the elements of matrix \( M \). The order of a matrix is the number of rows and columns it consists of. \( M \) has order 3 x 2.

A vector is a matrix consisting of a single row or column. For example, observations on 3 animals of 3, 4 and 1, respectively, can be represented as column or row vectors as follows:

A column vector: \( \mathbf{x} = \begin{bmatrix} 3 \\ 4 \\ 1 \end{bmatrix} \)

A row vector: \( \mathbf{x}' = [3 \quad 4 \quad 1] \)

A scalar is a single number such as 1, 6 or -9.
A.2 Matrix Operations

A.2.1 Addition

If matrices are of the same order, they are conformable for addition. The sum of two conformable matrices, is the matrix of sums element by element of the two matrices. For example, suppose \( A \) represents observations on the first replicate of a 2 x 2 factorial experiment, \( B \) represents observations on a second replicate and we want the sum of each treatment over replicates. This is given by matrix \( S = A + B \).

\[
A = \begin{bmatrix} 2 & 5 \\ 1 & 9 \end{bmatrix}, \quad B = \begin{bmatrix} -4 & 6 \\ 5 & 2 \end{bmatrix}, \quad S = A + B = \begin{bmatrix} 2 - 4 & 5 + 6 \\ 1 + 5 & 9 - 2 \end{bmatrix} = \begin{bmatrix} -2 & 11 \\ 6 & 7 \end{bmatrix}.
\]

A.2.2 Subtraction

The difference between two conformable matrices is the matrix of differences element by element of the two matrices. For example, suppose now we want the difference between replicate 1 and replicate 2 for each treatment combination, i.e. \( D = A - B \),

\[
D = A + B = \begin{bmatrix} 2 + 4 & 5 - 6 \\ 1 - 5 & 9 + 2 \end{bmatrix} = \begin{bmatrix} 6 & -1 \\ -4 & 11 \end{bmatrix}.
\]

A.2.3 Multiplication

Scalar Multiplication

A matrix multiplied by a scalar is the matrix with every element multiplied by the scalar. For example, suppose \( A \) represents a collection of measurements taken on one scale which we would like to convert to an alternative scale, and the conversion factor is 3.

For a scalar \( \lambda = 3 \). \( \lambda A = 3 \begin{bmatrix} 2 & 5 \\ 1 & 9 \end{bmatrix} = \begin{bmatrix} 6 & 15 \\ 3 & 27 \end{bmatrix} \).

Vector Multiplication

The product of a row vector with a column vector is a scalar obtained from the sum of the products of corresponding elements of the vectors. For example, suppose \( \mathbf{v} \) represents the number of observations taken on each of 3 animals and that \( \mathbf{y} \) represents the mean of these observations on each of the 3 animals and we want the totals for each animal.
Matrix Multiplication

Vector multiplication can be extended to the multiplication of a vector with a matrix, which is simply a collection of vectors. The product of a vector and a matrix is a vector and is obtained as follows:

\[ v' = [3 \ 4 \ 1] \]
\[ y = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \]

\[ t = v'y = [3 \ 4 \ 1] \begin{bmatrix} 1 \\ 5 \\ 2 \end{bmatrix} = 3(1) + 4(5) + 1(2) = 25. \]

That is, each column (or row) of the matrix is treated as a vector multiplication.

This can be extended further to the multiplication of matrices. The product of two conformable matrices is illustrated by the following example:

\[ A \times B = \begin{bmatrix} 2 & 5 \\ 1 & 9 \end{bmatrix} \begin{bmatrix} 4 & -6 \\ -5 & 2 \end{bmatrix} \]
\[ = \begin{bmatrix} 2(4) + 5(-5) & 2(-6) + 5(2) \\ 1(4) + 9(-5) & 1(-6) + 9(2) \end{bmatrix} \]
\[ = \begin{bmatrix} -17 & -2 \\ -41 & 12 \end{bmatrix}. \]

For matrix multiplication to be conformable, the number of columns of the first matrix must equal the number of rows of the second matrix.
A.2.4 Transpose

The transpose of a matrix is obtained by replacing rows with corresponding columns and vice-versa, e.g.

\[ M' = \begin{bmatrix} 6 & 4 \\ 3 & 9 \\ 8 & 5 \end{bmatrix}' = \begin{bmatrix} 6 & 3 & 8 \\ 4 & 9 & 5 \end{bmatrix}. \]

The transpose of the product of two matrices is the product of the transposes of the matrices taken in reverse order, e.g.

\[(AB)' = B'A'\]

A.2.5 Determinants

The determinant of a matrix is a scalar and exists only for square matrices. Knowledge of the determinant of a matrix is useful for obtaining the inverse of the matrix, which in matrix algebra is analogous to the reciprocal of scalar algebra. If \(A\) is a square matrix, its determinant can be symbolized as \(|A|\). Procedures for evaluating the determinant of various order matrices follow.

The determinant of a scalar (1 x 1 matrix) is the scalar itself, e.g. for \(A = 6\), \(|A| = 6\). The determinant of a 2 x 2 matrix is the difference between the product of the diagonal elements and the product of the off-diagonal elements, e.g. for

\[ A = \begin{bmatrix} 5 & 2 \\ 6 & 3 \end{bmatrix} \]

\[ |A| = 5(3) - 6(2) = 3. \]

The determinant of a 3 x 3 matrix can be obtained from the expansion of three 2 x 2 matrices obtained from it. Each of the second order determinants is preceded by a coefficient of +1 or -1, e.g. for

\[ A = \begin{bmatrix} 5 & 2 & 4 \\ 6 & 3 & 1 \\ 8 & 7 & 9 \end{bmatrix} \]

Based on elements of the first row,

\[ |A| = 5(+1) \begin{vmatrix} 3 & 1 \\ 7 & 9 \end{vmatrix} + 2(-1) \begin{vmatrix} 6 & 1 \\ 8 & 9 \end{vmatrix} + 4(+1) \begin{vmatrix} 6 & 3 \\ 8 & 7 \end{vmatrix} \]

\[ = 5(27 - 7) - 2(54 - 8) + 4(42 - 24) \]

\[ = 5(20) - 2(46) + 4(18) \]

\[ = 100 - 92 + 72 = 80 \]
The determinant was derived by taking in turn each element of the first row, crossing out the row and column corresponding to the element, obtaining the determinant of the resulting 2 x 2 matrix, multiplying this determinant by +1 or -1 and the element concerned, and summing the resulting products for each of the three first row elements. The (+1) or (-1) coefficients for the \( ij \)th element were obtained according to \((-1)^{i+j}\). For example, the coefficient for the 12 element is \((-1)^{1+2} = (-1)^3\) = -1. The coefficient for the 13 element is \((1)^{1+3} = (-1)^4\) = 1. The determinants of each of the 2 x 2 sub-matrices are called minors. For example, the minor of first row element 2 is

\[
\begin{bmatrix}
6 & 1 \\
8 & 9
\end{bmatrix} = 46
\]

When multiplied by its coefficient of \((-1)\), the product is called the co-factor of element 12. For example, the co-factor of elements 11, 12 and 13 are 20, -46 and 18.

Expansion by the elements of the second row yields the same determinant, e.g.

\[
|A| = 6(-1) \begin{bmatrix}
2 & 4 \\
7 & 9
\end{bmatrix} + 3(+1) \begin{bmatrix}
5 & 4 \\
8 & 9
\end{bmatrix} + 1(-1) \begin{bmatrix}
5 & 2 \\
8 & 7
\end{bmatrix}
\]

\[
= -6 (18 - 28) - 3 (45 - 32) + 1 (35 - 16)
\]

\[
= 60 + 39 - 19 = 80
\]

Similarly, expansion by elements of the third row again yields the same determinant, etc.

\[
|A| = 8(+1) \begin{bmatrix}
2 & 4 \\
3 & 1
\end{bmatrix} + 7(-1) \begin{bmatrix}
5 & 4 \\
6 & 1
\end{bmatrix} + 9(+1) \begin{bmatrix}
5 & 2 \\
6 & 3
\end{bmatrix}
\]

\[
= 8 (2 - 12) - 7 (5 - 24) + 9 (15 - 12)
\]

\[
= -80 + 133 + 27 = 80
\]

In general, multiplying the elements of any row by their co-factors yields the determinant. Also, multiplying the elements of a row by the co-factors of the elements of another row yields zero, e.g. the elements of the first row by the co-factors of the second row gives

\[
5(-1) \begin{bmatrix}
2 & 4 \\
7 & 9
\end{bmatrix} + 2(+1) \begin{bmatrix}
5 & 4 \\
8 & 9
\end{bmatrix} + 4(-1) \begin{bmatrix}
5 & 2 \\
8 & 7
\end{bmatrix}
\]

\[
= -5 (18 - 28) + 2 (45 - 32) + 4 (35 - 16)
\]

\[
= 50 + 26 - 76 = 0
\]

Expansion for larger order matrices follows according to

\[
|A| = \sum_{j=1}^{n} a_{ij} (-1)^{i+j} |M_{ij}|
\]

for any i where n is the order of the matrix, i = 1, \ldots, n and j = 1,\ldots, n, \( a_{ij} \) is the \( ij \)th element, and \( |M_{ij}| \) is the minor of the \( ij \)th element.
A2.6 Inverse

As suggested earlier, the inverse of a matrix is analogous to the reciprocal in scalar algebra and performs an equivalent operation to division. The inverse of matrix \(A\) is symbolized as \(A^{-1}\). The multiplication of a matrix by its inverse gives an identity matrix \((I)\), which is composed of all diagonal elements of one and all off-diagonal elements of zero, i.e. \(A \times A^{-1} = I\). For the inverse of a matrix to exist, it must be square and have a non-zero determinant.

The inverse of a matrix can be obtained from the co-factors of the elements and the determinant.

The following example illustrates the derivation of the inverse.

\[
A = \begin{bmatrix}
5 & 2 & 4 \\
6 & 3 & 1 \\
8 & 7 & 9 \\
\end{bmatrix}
\]

i) Calculate the co-factors of each element of the matrix, e.g. the co-factors of the elements of the first row are \((+1) \begin{bmatrix} 3 & 1 \\ 7 & 9 \end{bmatrix}, (-1) \begin{bmatrix} 6 & 1 \\ 8 & 9 \end{bmatrix}, \text{ and } (+1) \begin{bmatrix} 6 & 3 \\ 8 & 7 \end{bmatrix} = 20, -46 \text{ and } 18.\]

Similarly, the co-factors of the elements of the second row are \(= 10, 13 \text{ and } -19\)

and the co-factors of the elements of the third row are \(= -10, 19 \text{ and } 3.\)

ii) Replace the elements of the matrix by their co-factors, e.g.

\[
A = \begin{bmatrix}
5 & 2 & 4 \\
6 & 3 & 1 \\
8 & 7 & 9 \\
\end{bmatrix}
\text{ yields } C = \begin{bmatrix}
20 & -46 & 18 \\
10 & 13 & -19 \\
-10 & 19 & 3 \\
\end{bmatrix}
\]

iii) Transpose the matrix of co-factors, e.g.

\[
C' = \begin{bmatrix}
20 & -46 & 18 \\
10 & 13 & -19 \\
-10 & 19 & 3 \\
\end{bmatrix}\text{ } = \begin{bmatrix}
20 & 10 & -10 \\
-46 & 13 & 19 \\
18 & -19 & 3 \\
\end{bmatrix}
\]

iv) Multiply the transpose matrix of co-factors by the reciprocal of the determinant to yield the inverse, e.g.

\[
|A| = 80, \frac{1}{|A|} = 1/80
\]

\[
A^{-1} = \frac{1}{80} \begin{bmatrix}
20 & 10 & -10 \\
-46 & 13 & 19 \\
18 & -19 & 3 \\
\end{bmatrix}
\]
v) As a check, the inverse multiplied by the original matrix should yield an identity matrix, i.e. \( A^{-1}A = I \), e.g.

\[
\begin{bmatrix}
  20 & 10 & -10 \\
  -46 & 13 & 19 \\
  18 & -19 & 3
\end{bmatrix}
\begin{bmatrix}
  5 & 2 & 4 \\
  6 & 3 & 1 \\
  8 & 7 & 9
\end{bmatrix}
= \begin{bmatrix}
  1 & 0 & 0 \\
  0 & 1 & 0 \\
  0 & 0 & 1
\end{bmatrix}.
\]

The inverse of a 2 x 2 matrix is:

\[
\begin{bmatrix}
  a & b \\
  c & d
\end{bmatrix}^{-1} = \frac{1}{ad - bc}
\begin{bmatrix}
  d & -b \\
  -c & a
\end{bmatrix}
\]

A.2.7 Linear Independence and Rank

As indicated, if the determinant of a matrix is zero, a unique inverse of the matrix does not exist. The determinant of a matrix is zero if any of its rows or columns are linear combinations of other rows or columns. In other words, a determinant is zero if the rows or columns do not form a set of linearly independent vectors. For example, in the following matrix:

\[
\begin{bmatrix}
  5 & 2 & 3 \\
  2 & 2 & 0 \\
  3 & 0 & 3
\end{bmatrix}
\]

rows 2 and 3 sum to row 1 and the determinant of the matrix is zero.

The rank of a matrix is the number of linearly independent rows or columns. For example, the rank of the above matrix is 2. If the rank of matrix \( \mathbf{A} \) is less than its order \( n \), then the determinant is zero and the inverse of \( \mathbf{A} \) does not exist, i.e. if \( r(\mathbf{A}) < n \) then \( \mathbf{A}^{-1} \) does not exist.

A.2.8 Generalized Inverse

Although a unique inverse does not exist for a matrix of less than full rank, generalized inverses do exist. If \( \mathbf{A}^* \) is a generalized inverse of \( \mathbf{A} \), it satisfies \( \mathbf{AA}^* = \mathbf{A} \). Generalized or g-inverses are not unique and there are many \( \mathbf{A}^* \) which satisfy \( \mathbf{AA}^* = \mathbf{A} \). There are also many ways to obtain a g-inverse, but one of the simplest ways is to follow these steps:

a) Obtain a full rank subset of \( \mathbf{A} \) and call it \( \mathbf{M} \).
b) Invert \( \mathbf{M} \) to yield \( \mathbf{M}^{-1} \).
c) Replace each element in \( \mathbf{A} \) with the corresponding element of \( \mathbf{M}^{-1} \).
d) Replace all other elements of \( \mathbf{A} \) with zeros.
e) The result is \( \mathbf{A}^- \), a generalized inverse of \( \mathbf{A} \).
Example

\[ A = \begin{bmatrix} 6 & 3 & 2 & 1 \\ 3 & 3 & 0 & 0 \\ 2 & 0 & 2 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix} \]

a) \( M \), a full rank subset, is

\[ M = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix} \]

b) \( M^{-1} = \begin{bmatrix} 1/3 & 0 & 0 \\ 0 & 1/2 & 0 \\ 0 & 0 & 1 \end{bmatrix} \).

c) Replacing elements of \( A \) with corresponding elements of \( M^{-1} \) and all other elements with 0’s gives

d) \( A^- = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1/3 & 0 & 0 \\ 0 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \).

A.2.9 Special Matrices

In many applications of statistics we deal with matrices that are the product of a matrix and its transpose, e.g.

\[ A = X'X \]

Such matrices are always symmetric, that is every off-diagonal element above the diagonal equals its counterpart below the diagonal. For such matrices

\[ X (X'X)^- X'X = X \]

and \( X(X'X)X' \) is invariant to \((X'X)^-\), that is, although there are many possible g-inverses of \( X'X \), any g-inverse pre-multiplied by \( X \) and post-multiplied by \( X'X \) yields the same matrix \( X \).

A.2.10 Trace

The trace of a matrix is the sum of the diagonal elements. For the matrix \( A \) of order \( n \) with elements \((a_{ij})\), the trace is defined as

\[ \text{tr} (A) = \sum_{i=1}^{n} a_{ii} \]
As an example, the trace of
\[
\begin{bmatrix}
3 & 1 & 4 \\
1 & 6 & 2 \\
4 & 2 & 5
\end{bmatrix}
\]
is \[3 + 6 + 5 = 14\]

For products of matrices, \(\text{tr}(AB) = \text{tr}(BA)\) if the products are conformable. This can be extended to the product of three or more matrices, e.g.

\[
\text{Tr}(ABC) = \text{tr}(BCA) = \text{tr}(CAB)
\]

**A.3 Quadratic Forms**

All sums of squares can be expressed as quadratic forms that is a \(y' Ay\). If \(y \sim (\mu, V)\), then \(E(y'Ay) = \mu'A\mu\)

**Exercises**

1. For
   \[
   A = \begin{bmatrix}
   6 & 3 \\
   0 & 5 \\
   -5 & 1
   \end{bmatrix}
   \quad B = \begin{bmatrix}
   3 & 8 \\
   2 & -4 \\
   5 & -1
   \end{bmatrix}
   \]
   Find the sum of \(A + B\).
   Find the difference of \(A - B\).

2. For \(A\) and \(B\) above and \(v' = [1 \quad 3 \quad -1]\), find \(v'A\) and \(v'B\).

3. For
   \[
   B' = \begin{bmatrix}
   3 & 2 & 5 \\
   8 & -4 & -1
   \end{bmatrix}
   \]
   and \(A\) as above. Find \(B'A\). Find \(AB'\).

4. For \(A\) and \(B\) above, find \(AB\).

5. Obtain determinants of the following matrices
   \[
   \begin{bmatrix}
   3 & 8 \\
   2 & -4
   \end{bmatrix}
   \]
   \[
   \begin{bmatrix}
   6 & 3 \\
   1 & 5 \\
   -5 & 2
   \end{bmatrix}
   \]
6. Show that the solution to \( Ax = y \) is \( x = A^{-1}y \).

7. Derive the inverses of the following matrices:

\[
\begin{bmatrix}
4 & 2 \\
6 & 1 \\
\end{bmatrix}
\]

\[
\begin{bmatrix}
2 & 1 & 3 \\
-5 & 1 & 0 \\
1 & 4 & -2 \\
\end{bmatrix}
\]

\[
\begin{bmatrix}
3 & 0 & 0 & 0 \\
0 & 4 & 0 & 0 \\
0 & 0 & 2 & 0 \\
0 & 0 & 0 & 5 \\
\end{bmatrix}
\]

8. For \( A = \begin{bmatrix} 1 & 1 & 3 \\ 2 & 3 & 0 \\ 4 & 5 & 6 \end{bmatrix} \) and \( B = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \), show that \( \text{tr}(AB) = \text{tr}(BA) \).