Chapter 5

Genetic Covariances Between Animals

The prediction of breeding values using BLUP as shown in chapter 4 uses linear mixed effects models where the breeding value of each animal is included as a random effect. Linear mixed effect models in general and specifically Henderson's mixed model equations require us to be able to specify the variance-covariance matrix of all random effects. When using the animal model, the breeding value of each animal is included as a random effect in the linear mixed effects model. As a consequence of that we need to determine the covariance between the true breeding values of all animals. Figure 5.1 tries to display the structure of the required variance-covariance diagrammatically.

The variance-covariance matrix shown at the bottom of Figure 5.1 has the variances of the true breeding values on the diagonal and all the covariances between the true breeding values of all animals as offdiagonal elements. From the formulation of the linear mixed effect model in (4.5), we defined the variance-covariance matrix of the random effects u to be G (see equation (4.7)). When predicting breeding values with the animal model, the variance-covariance matrix of all components in the vector u is defined as

$$var(u) = G = A * \sigma_u^2 \tag{5.1}$$

where the matrix A is called **numerator relationship matrix**.

5.1 Similarity Between Individuals

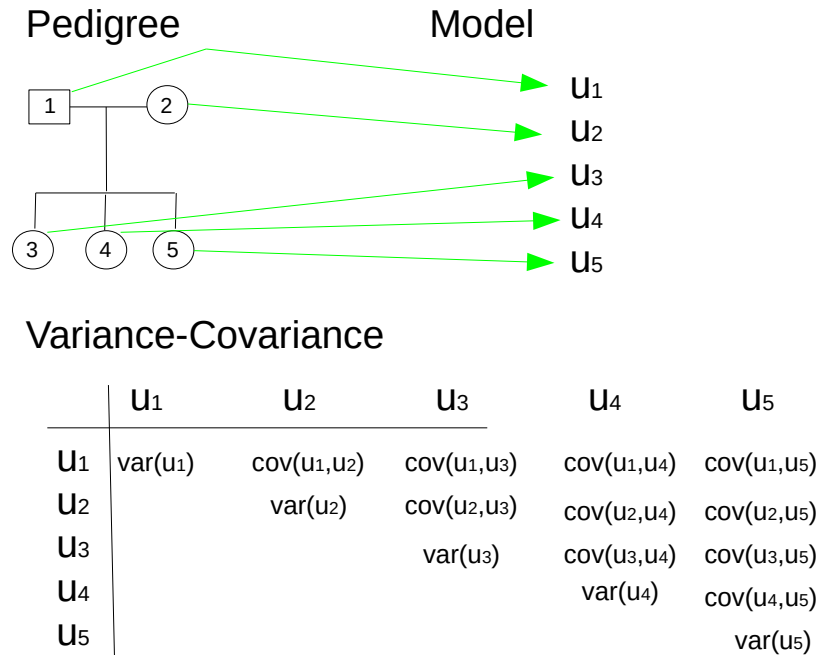At the genetic level there are two different kinds of similarity

Pedigree   Model

	u_1	u_2	u_3	u_4	u_5
u_1	$\text{var}(u_1)$	$\text{cov}(u_1,u_2)$	$\text{cov}(u_1,u_3)$	$\text{cov}(u_1,u_4)$	$\text{cov}(u_1,u_5)$
u_2		$\text{var}(u_2)$	$\text{cov}(u_2,u_3)$	$\text{cov}(u_2,u_4)$	$\text{cov}(u_2,u_5)$
u_3			$\text{var}(u_3)$	$\text{cov}(u_3,u_4)$	$\text{cov}(u_3,u_5)$
u_4				$\text{var}(u_4)$	$\text{cov}(u_4,u_5)$
u_5					$\text{var}(u_5)$

Variance-Covariance

Figure 5.1: Genetic Covariance Between Animals

1. Identity by descent (IBD)
2. Identity by state

Figure 5.2 illustrates the difference between the two type of identities. The type of graph shown in Figure 5.2 is called a **pedigree** which is commonly used to display the relationship between animals in a population. The rectangle symbols denote male animals and the round symbols stand for female animals. The horizontal connections between female and male animals denote a mating. All animals connected to a vertical line and follow below are progeny of the connected parents.

The notations inside of the symbols stand for the different genotypes of the animals on a given locus. The red arrows denote the path of two A_1-alleles which are copies of the same ancestral allele. These two copies are called **identical by descent** (IBD). The green arrows show the path of two alleles which are identical by state which do not originate from the same copy of any given ancestral alleles.
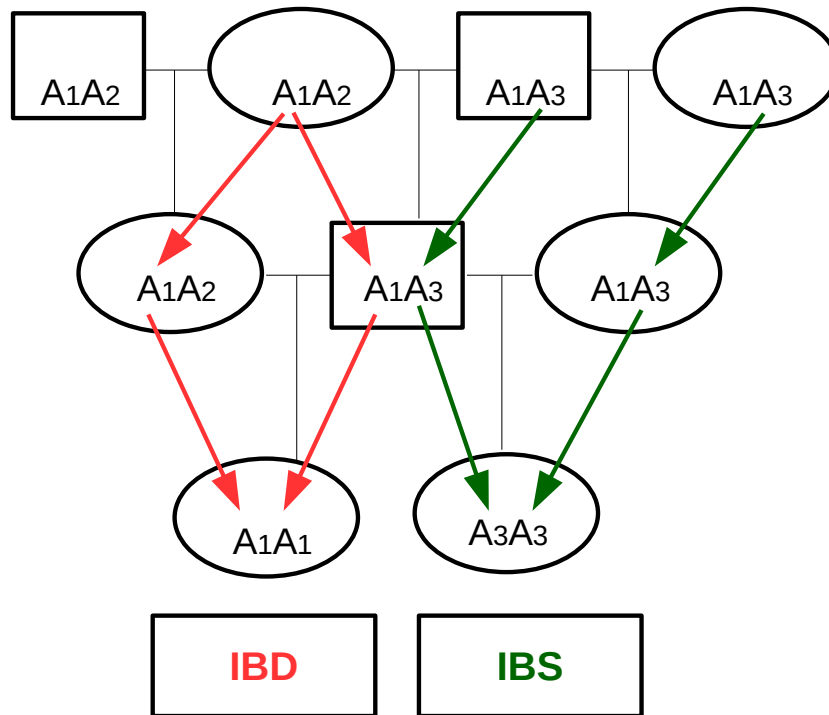
Figure 5.2: Identity by State Versus Identity by Descent

5.2 Numerator Relationship Matrix

The probability of identical genes by descent (IBD) occurring in two individuals is termed the co-ancestry or the coefficient of kinship [[Falconer and Mackay, 1996]]. The additive genetic relationship between two individuals is twice their co-ancestry. The matrix that expresses the additive genetic relationship among animals in a population is called the **numerator relationship matrix** A. The matrix A is symmetric and its diagonal elements $(A)_{ii}$ are equal to $1 + F_i$ where F_i is the **coefficient of inbreeding** of animal i. The coefficient of inbreeding F_i indicates whether an animal i is inbred or not. F_i is defined to be half the additive genetic relationship between the parents of i. Hence the diagonal element $(A)_{ii}$ of matrix A corresponds to twice the probability that two gametes taken at random from an animal i will carry IBD-alleles.

The off-diagonal elements $(A)_{ij}$ equals the numerator of the coefficient of relationship between animals i and j. Multiplying the matrix A by the additive genetic variance σ_u^2 leads to the covariance among breeding values. Thus if u_i is the breeding value of animal i then

$$var(u_i) = (A)_{ii}\sigma_u^2 = (1 + F_i)\sigma_u^2 \tag{5.2}$$

5.2.1   Algorithm To Compute A

The matrix A can be computed using either the

1. path coefficient method or
2. recursive method.

The second method is especially suitable for an implementation by a software program. In what follows the recursive method to compute the components of A is described now. Initially, animals in a pedigree are numbered from 1 to n and ordered such that parents precede their progeny. The following rules are then used to compute the components of A.

- If both parents s and d of animal i are known then
  - the diagonal element $(A)_{ii}$ corresponds to: $(A)_{ii} = 1 + F_i = 1 + \frac{1}{2}(A)_{sd}$ and
  - the off-diagonal element $(A)_{ji}$ is computed as: $(A)_{ji} = \frac{1}{2}((A)_{js} + (A)_{jd})$
  - because A is symmetric $(A)_{ji} = (A)_{ij}$
- If only one parent s is known and assumed unrelated to the mate
  - $(A)_{ii} = 1$
  - $(A)_{ij} = (A)_{ji} = \frac{1}{2}((A)_{js}$
- If both parents are unknown
  - $(A)_{ii} = 1$
  - $(A)_{ij} = (A)_{ji} = 0$

5.2.2   Numeric Example

We are given the following pedigree and we want to compute the matrix A using the recursive method described in 5.2.1.

Table 5.1: Example Pedigree To Compute Additive Genetic Relationship Matrix

Calf	Sire	Dam
3	1	2
4	1	NA

$$
\begin{array}{ccc}
5 & 4 & 3 \\
6 & 5 & 2
\end{array}
$$

The first step of the computations of A are the numbering and the ordering of all the animals. This is already done in the pedigree shown in Table 5.1. The components of A are computed row-by-row starting with $(A)_{11}$.

$$
(A)_{11} = 1 + F_1 = 1 + 0 = 1
$$
$$
(A)_{12} = 0 = (A)_{21}
$$
$$
(A)_{13} = \frac{1}{2}((A)_{11} + (A)_{12}) = 0.5 = (A)_{31}
$$
$$
(A)_{14} = \frac{1}{2}(A)_{11} = 0.5 = (A)_{14}
$$
$$
(A)_{15} = \frac{1}{2}(A)_{14} + (A)_{13}) = 0.5 = (A)_{51}
$$
$$
(A)_{16} = \frac{1}{2}(A)_{15} + (A)_{12}) = 0.25
$$

The same computations are also done for all the other components of the matrix A. The final result for the matrix looks as follows

$$
A = \begin{bmatrix}
1 & 0 & 0.5 & 0.5 & 0.5 & 0.25 \\
0 & 1 & 0.5 & 0 & 0.25 & 0.625 \\
0.5 & 0.5 & 1 & 0.25 & 0.625 & 0.5625 \\
0.5 & 0 & 0.25 & 1 & 0.625 & 0.3125 \\
0.5 & 0.25 & 0.625 & 0.625 & 1.125 & 0.6875 \\
0.25 & 0.625 & 0.5625 & 0.3125 & 0.6875 & 1.125
\end{bmatrix}
$$

As a result, we can see from the components of the above shown matrix A that animals 1 and 2 are not related to each other. Furthermore from the diagonal elements of A, it follows that animals 5 and 6 are inbred while animals 1 to 4 are not inbred.

5.3   The Inverse Numerator Relationship Matrix

The general form of Henderson's mixed model equations as shown in (4.10) depend on four matrices

1. Design matrix X for the fixed effects
2. Design matrix Z for the random effects
3. The inverse variance-covariance matrix R^{-1} for the residuals e and

4. The inverse variance-covariance matrix G^{-1} for the random breeding values a.

When using the animal model as described in section @ref(#animalmodel), we have seen in (5.1) that G corresponds to the product of the numerator relationship matrix A and the genetic additive variance σ_u^2. But the mixed model equations depend on the inverse G^{-1} of G which means

$$G^{-1} = (A * \sigma_u^2)^{-1} = A^{-1} * \frac{1}{\sigma_u^2} \tag{5.3}$$

From (5.3) we can see that in order to be able to set up the mixed model equations for an animal model, we need the inverse numerator relationship matrix A^{-1}. Because in practical routine predictions of breeding values, we have something in the order of 10 million animals that we predict breeding values for. Hence the matrix A has 10 million rows and 10 million columns. A matrix of that size cannot be inverted explicitly with commonly known methods such as described under https://en.wikipedia.org/wiki/Invertible_matrix. This would have been the end of the BLUP animal model, if not Henderson published in [[Henderson, 1976]] a set of simple rules to directly construct the matrix A^{-1} without setting up the numerator relationship matrix.

5.4   Structure of A^{-1}

When looking at a concrete example of an inverse of a numerator relationship matrix as shown below, we can discover some important facts. Let us assume the following pedigree.

Table 5.2: Pedigree Used To Compute Inverse Numerator Relationship Matrix

Calf	Sire	Dam
1	NA	NA
2	NA	NA
3	NA	NA
4	1	2
5	3	2

The numerator relationship matrix A for the pedigree shown in Table 5.2 is shown in (5.4).

$$A = \begin{bmatrix} 1 & 0 & 0 & 0.5 & 0 \\ 0 & 1 & 0 & 0.5 & 0.5 \\ 0 & 0 & 1 & 0 & 0.5 \\ 0.5 & 0.5 & 0 & 1 & 0.25 \\ 0 & 0.5 & 0.5 & 0.25 & 1 \end{bmatrix} \tag{5.4}$$

The matrix A^{-1} shown below corresponds to the inverse of the matrix A from (5.4).

$$A^{-1} = \begin{bmatrix} 1.5 & 0.5 & 0 & -1 & 0 \\ 0.5 & 2 & 0.5 & -1 & -1 \\ 0 & 0.5 & 1.5 & 0 & -1 \\ -1 & -1 & 0 & 2 & 0 \\ 0 & -1 & -1 & 0 & 2 \end{bmatrix}$$

From the above shown inverse A^{-1}, we recognize that A^{-1} has a simpler structure than A itself. The reason for this is that non-zero elements occur only at matrix elements of A^{-1} corresponding to parents and progeny or to mates. Furthermore off-diagonal elements corresponding to mates are positive and elements corresponding to parents and progeny are negative. These observations were used by Henderson in [Henderson, 1976] to come up with the rules described below.

5.5 Henderson's Rule To Set Up A^{-1}

We denote the row or column index corresponding to an animal of interest as i and the row or column index belonging to the animals father as s and the row or column index corresponding to animal i's mother as d. The rules differentiate the following three cases

1. both parents of animal i are known
2. only one parent of animal i is known
3. both parents of animal i are unknown

5.5.1 Both Parents Known

- add 2 to the diagonal-element (i, i)
- add -1 to off-diagonal elements (s, i), (i, s), (d, i) and (i, d)
- add $\frac{1}{2}$ to elements (s, s), (d, d), (s, d), (d, s)

5.5.2 Only One Parent Known

We assume that sire s is known

- add $\frac{4}{3}$ to diagonal-element (i, i)
- add $-\frac{2}{3}$ to off-diagonal elements (s, i), (i, s)
- add $\frac{1}{3}$ to element (s, s)

5.5.3 Both Parents Unknown

- add 1 to diagonal-element (i, i)

The application of Henderson's rules to construct A^{-1} directly will be a problem in one of the coming exercises. When applying the above described rules, it has to be noted well that this simple version of the rules are only valid for a pedigree without inbreeding. We will see in section @ref(#derivationofhendersonsrules) how to extend the simple rules such that they can be used for a general pedigree with inbreeding.

5.6 Derivation of Henderson's Rules

Henderson's rules can be related to the so-called `LDL`-decomposition of the numerator relationship matrix A. The result of this decomposition consists of the equivalence between matrix A and the product of three matrices L, D and L^T. The matrix L is a lower triangular matrix and the matrix D is a diagonal matrix. The reason for why we are doing this decomposition of A is that the matrices L and D can much easier be inverted than the matrix A. The `LDL`-decomposition is a general procedure in linear algebra and it can be applied to any symmetric and positive-definite matrix not only to numerator relationship matrices. But when the `LDL`-decomposition is applied to a numerator relationship matrix, the matrices L and D do also have a special genetic meaning. This meaning is revealed in the following derivation.

5.6.1 Decomposition of True Breeding Value and its Variance

The true breeding value (u_i) of animal i can be decomposed into the true breeding values of animal i's parents s and d and the mendelian sampling term m_i

$$u_i = \frac{1}{2}u_s + \frac{1}{2}u_d + m_i \tag{5.5}$$

Applying the decomposition shown in (5.10) for all animals in the pedigree and combining the decompositions into a matrix-vector notation, we get

$$u = P \cdot u + m \tag{5.6}$$

where

u	vector of true breeding values for all animals
P	matrix linking animals to their parents
m	vector of mendelian sampling terms

Equation (5.6) can be interpreted as regression of the true breeding values onto parental breeding values. In such a regression the random term m is like a residual term. The genetic meaning of m corresponds to the deviation of u_i from the full-sib average of the true breeding values within the nuclear family with parents s and d. The term m is called **mendelian sampling** term. The source of m is the fact that during meiosis, parental alleles are randomly assigned to each progeny. Bulmer [[Bulmer, 1971]] has shown that m_i are independent from true breeding values of parents s and d. Using this result, we take the variance on both sides of (5.10) leading to

$$\begin{aligned}
var(u_i) &= var(\frac{1}{2}u_s + \frac{1}{2}u_d + m_i) \\
&= \frac{1}{4}var(u_s) + \frac{1}{4}var(u_d) + \frac{1}{2}cov(u_s, u_d) + var(m_i)
\end{aligned} \tag{5.7}$$

From (5.1) together with the definition of the numerator relationship matrix A, we know that

$$\begin{aligned}
var(u_i) &= (1 + F_i)\sigma_u^2 \\
var(u_s) &= (1 + F_s)\sigma_u^2 \\
var(u_d) &= (1 + F_d)\sigma_u^2 \\
cov(u_s, u_d) &= (A)_{sd}\sigma_u^2 = 2F_i\sigma_u^2
\end{aligned} \tag{5.8}$$

5.6.2 Variance of Mendelian Sampling Terms

Inserting the relations from (5.8) into (5.7) and solving for $var(m_i)$ gives the following result

$$var(m_i) = var(u_i) - \frac{1}{4}var(u_s) - \frac{1}{4}var(u_d) - \frac{1}{2}cov(u_s, u_d)$$

$$= (1 + F_i)\sigma_u^2 - \frac{1}{4}(1 + F_s)\sigma_u^2 - \frac{1}{4}(1 + F_d)\sigma_u^2 - \frac{1}{2} * 2 * F_i * \sigma_u^2$$

$$= \left(\frac{1}{2} - \frac{1}{4}(F_s + F_d)\right)\sigma_u^2 \tag{5.9}$$

In case where only parent s of animal i is known the terms in (5.10) and (5.9) change to

$$u_i = \frac{1}{2}u_s + \frac{1}{2}u_d + m_i \tag{5.10}$$

$$u_i = \frac{1}{2}u_s + m_i$$

$$var(m_i) = \left(1 - \frac{1}{4}(1 + F_s)\right)\sigma_u^2$$

$$= \left(\frac{3}{4} - \frac{1}{4}F_s\right)\sigma_u^2 \tag{5.11}$$

When both parents are unknown, we get

$$u_i = m_i$$

$$var(m_i) = \sigma_u^2 \tag{5.12}$$

5.6.3   Decomposition of A

The true breeding values u_s of sire s and u_d of dam d can be decomposed in a similar way as shown for the true breeding value u_i in (5.10).

$$u_s = \frac{1}{2}u_{ss} + \frac{1}{2}u_{ds} + m_s$$

$$u_d = \frac{1}{2}u_{sd} + \frac{1}{2}u_{dd} + m_d \tag{5.13}$$

where

ss	sire of s
ds	dam of s
sd	sire of d
dd	dam of d

Using (5.13) together with (5.10) leads to the following expression for u_i

$$
\begin{aligned}
u_i &= \frac{1}{2}u_s + \frac{1}{2}u_d + m_i \\
&= \frac{1}{2}(\frac{1}{2}u_{ss} + \frac{1}{2}u_{ds} + m_s) + \frac{1}{2}(\frac{1}{2}u_{sd} + \frac{1}{2}u_{dd} + m_d) + m_i \\
&= \frac{1}{4}u_{ss} + \frac{1}{4}u_{ds} + \frac{1}{4}u_{sd} + \frac{1}{4}u_{dd} + \frac{1}{2}m_s + \frac{1}{2}m_d + m_i
\end{aligned}
$$

This type of decomposition can also be done for the grand-parents of animal i and further back until we get to true breeding values of animals with un-known parents. Animals of ancestor generations with unknown parents are called **founder population**. The process of decomposing true breeding values back through all generations of a pedigree is called **recursive decomposition** of animal i's true breeding value. Because we know from (5.12) that the decomposition of true breeding values u_k for an animal k with unknown parents is $u_k = m_k$, the result of the recursive decomposition of u_i is a sum of mendelian sampling terms linking the ancestors of i back to the founder population.

Ordering all animals in a pedigree according to their age and writing the result of the recursive decomposition of all true breeding values in matrix-vector notation leads to

$$
u = L \cdot m \tag{5.14}
$$

The matrix L is a lower triangular matrix. The row corresponding to animal i is the average of the rows in L corresponding to parents s and d of i. The matrix L contains the path of the gene flow from the base population to all animals in the population. From equation (5.14), we are computing the variance of all true breeding values which leads to

$$
\begin{aligned}
var(u) &= var(L \cdot m) \\
&= L \cdot var(m) \cdot L^T \\
&= L \cdot (D\sigma_u^2) \cdot L^T \\
&= L \cdot D \cdot L^T \sigma_u^2 \qquad\qquad = A\sigma_u^2
\end{aligned}
$$

From (??), the `LDL`-decomposition of A follows directly as $A = LDL^T$. The single components m_i are independent of each other. This also means that $cov(m_i, m_j) = 0$ for $i \neq j$. Hence the matrix D is a diagonal matrix. In section @ref(#variancemendeliansamplingterm) we have seen that $var(m_i)$ always contain σ_u^2 as a factor. Hence it is reasonable to express $var(m)$ as a product of a diagonal matrix D times the genetic additive variance

σ_u^2. The diagonal elements of matrix D are computed as shown in section @ref(#variancemendeliansamplingterm).

The mixed model equations require the inverse numerator relationship matrix A^{-1}. Thanks to the `LDL`-decomposition of A, we can compute A^{-1} as

$$A^{-1} = (L \cdot D \cdot L^T)^{-1} = (L^T)^{-1} \cdot D^{-1} \cdot L^{-1} \tag{5.15}$$

The inverse D^{-1} is easy to compute, because D is a diagonal matrix. As a consequence of that D^{-1} is also a diagonal matrix where the elements $(D^{-1})_{ii}$ correspond to the inverse of the elements of the original matrix D, i.e. $(D^{-1})_{ii} = 1/(D)_{ii}$. The matrix L^{-1} is also a lower triangular matrix and can easily computed based on the two relations for the vector m. Based on (5.6), we know

$$m = u - Pu = (I - P)u \tag{5.16}$$

and from (5.14), we get

$$m = L^{-1}u \tag{5.17}$$

Setting both expressions for m in (5.16) and (5.17) equal can be used to obtain L^{-1}

$$m = (I - P)u = L^{-1}u \tag{5.18}$$

Therefore,

$$L^{-1} = I - P \tag{5.19}$$

5.6.4   General Version of Henderson's Rule

Based on the decomposition of A^{-1} shown in (5.15), the general version of Henderson's rule where inbreeding of animals can be accounted for can be summarized as

- Start with a matrix A^{-1} where all elements are set to 0.
- Let d^i be the i-th diagonal element of D^{-1} for animal i, assuming i has parents s and d.
- Then add the following contributions to A^{-1}

  - d^i to the element (i, i)
  - $-d^i/2$ to the elements (s, i), (i, s), (d, i), (i, d)
  - $d^i/4$ to the elements (s, s), (s, d), (d, s), (d, d)

The contributions to rows and columns corresponding to parents s and d are only added, if they are known. Because the elements d^i are dependent on the inbreeding coefficients F_s and F_d of parents s and d, we have to find an efficient way to compute inbreeding coefficients for all animals in a population.

5.7 Computing Inbreeding Coefficients

Inbreeding coefficients can be computed using different methods. From all these methods, we are just showing the one method which was described in [[Quaas, 1976]]. This method is based on the properties of a second decomposition of the numerator relationship matrix A which is called the **cholesky decomposition**. The cholesky decomposition of a matrix A corresponds to

$$A = RR^T \tag{5.20}$$

where the matrix R is a lower triangular matrix. At this point we have to be clear about this. In practical routine evaluations, we will not explicitly compute this decomposition, because we do not want to construct A explicitly. We are just using the properties of (5.20) to find an efficient way to compute the diagonal elements $(A)_{ii}$ of A without constructing the complete matrix A. The diagonal elements $(A)_{ii}$ are important, because they contain the inbreeding coefficients (F_i), as $(A)_{ii} = 1 + F_i$. Based on (5.20), $(A)_{ii}$ can be computed from the components of R as

$$(A)_{ii} = \sum_{j=1}^{i}(R)_{ij}^2 \tag{5.21}$$

This can be shown with a small 3×3 matrix A

$$\begin{bmatrix} (A)11 & (A)12 & (A)13 \\ (A)21 & (A)22 & (A)23 \\ (A)31 & (A)32 & (A)33 \end{bmatrix} = \begin{bmatrix} (R)11 & 0 & 0 \\ (R)21 & (R)22 & 0 \\ (R)31 & (R)32 & (R)33 \end{bmatrix} * \begin{bmatrix} (R)11 & (R)21 & (R)31 \\ 0 & (R)22 & (R)32 \\ 0 & 0 & (R)33 \end{bmatrix}$$

For the above shown example, the diagonal elements $(A)_{ii}$ are computed as

$$(A)_{11} = (R)_{11}^2$$
$$(A)_{22} = (R)_{21}^2 + (R)_{22}^2$$
$$(A)_{33} = (R)_{31}^2 + (R)_{32}^2 + (R)_{33}^2$$

This shows that diagonal elements $(A)_{ii}$ can be computed using just all the components of row i in R up to the diagonal. Next, we have to show how to compute the components of R.

5.7.1   Recursive Computation of R

We are setting the two decompositions of A equal which leads to

$$A = R * R^T = L * D * L^T \tag{5.22}$$

Let us write the matrix R as a product of two matrices L and S where L corresponds to the matrix from the LDL-decomposition and insert that product into (5.22)

$$A = R * R^T = L * S * (L * S)^T = L * S * S^T * L^T = L * D * L^T \tag{5.23}$$

From (5.23), we conclude that $D = S \cdot S^T$ where S is also a diagonal matrix with elements $(S)_{ii} = \sqrt{(D)_{ii}}$. For our small example we get

$$\begin{bmatrix} (R)11 & 0 & 0 \\ (R)21 & (R)22 & 0 \\ (R)31 & (R)32 & (R)33 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ (L)21 & 1 & 0 \\ (L)31 & (L)32 & 1 \end{bmatrix} * \begin{bmatrix} (S)11 & 0 & 0 \\ 0 & (S)22 & 0 \\ 0 & 0 & (S)33 \end{bmatrix} \tag{5.24}$$

From (5.24), we can see that the diagonal elements $(R)_{ii}$ are equal to $(S)_{ii}$. Therefore

$$(R)_{ii} = (S)_{ii} = \sqrt{(D)_{ii}} \tag{5.25}$$

Earlier, we have seen that diagonal elements $(D)_{ii}$ of D correspond to

$$(D)_{ii} = \frac{1}{2} - \frac{1}{4}(F_s + F_d) = 1 - 0.25((A)_{ss} + (A)_{dd}) \tag{5.26}$$

and hence

$$(R)_{ii} = \sqrt{1 - 0.25((A)_{ss} + (A)_{dd})} \tag{5.27}$$

The components $(A)_{ss}$ and $(A)_{dd}$ correspond to diagonal elements of A of parents of s and d which were computed earlier.

The off-diagonal elements of R are computed as

$$(R)_{ij} = (L)_{ij} * (S)_{jj} \tag{5.28}$$

One property of the matrix L is that any element $(L)_{ij}$ is equal to the average of elements $(L)_{sj}$ and $(L)_{dj}$, if s and d are parents of animal i. Using this we get

$$
\begin{aligned}
(R)_{ij} &= (L)_{ij} * (S)_{jj} \\
&= \frac{1}{2}((L)_{sj} + (L)_{dj}) * (S)_{jj} \\
&= \frac{1}{2}((R)_{sj} + (R)_{dj})
\end{aligned}
\tag{5.29}
$$

With that we have shown how to compute all elements of R recursively. This requires an ordering of all animals according to their age.