

Züchtungslehre - Inverse Verwandtschaftsmatrix

Peter von Rohr

2016-10-28

Einleitung

Die genetische Verwandtschaft zwischen Tieren ist sehr wichtig bei der Zuchtwertschätzung. Wie wir im Kapitel zur Verwandtschaft schon gesehen hatten, zeigen verwandte Tiere eine erhöhte Wahrscheinlichkeit Kopien des gleichen Ahnen-Allels zu tragen. Im Kapitel BLUP-Zuchtwertschätzung werden wir sehen, dass genetische Kovarianzen zwischen Tieren als Informationsquellen verlangt werden. Die genetischen Kovarianzen zwischen Tieren sind durch die Verwandtschaftsbeziehungen zwischen den Tieren bestimmt. Technisch gesehen wird die additiv genetische Verwandtschaftsmatrix A und insbesondere ihre Inverse A^{-1} eine wichtige Rolle spielen bei der BLUP-Zuchtwertschätzung.

Verwandtschaftsmatrix

Im Kapitel “Verwandtschaft und Inzucht” wurde bereits besprochen, wie die Verwandtschaftsgrade zwischen Tieren und die Inzuchtkoeffizienten von Tieren berechnet werden. Wir haben auch einen rekursiven Algorithmus angeschaut, welcher es uns erlaubt die ganze Verwandtschaftsmatrix aufzustellen. Die Verwandtschaftsmatrix enthält sehr viel Informationen über die Struktur einer Population. Als Nebenprodukt erhalten wir auch die Inzuchtkoeffizienten aller Tiere in unserer Population. Es gibt auch spannende Möglichkeiten die Verwandtschaftsmatrix graphisch darzustellen. Eine Möglichkeit ist die sogenannten Heatmap. Als Beispiel können wir die Verwandtschaftsmatrix aus dem Kapitel “Verwandtschaft und Inzucht”, wie folgt in eine Heatmap verwandeln. Das Pedigree in Listenformat ist nachfolgend gezeigt.

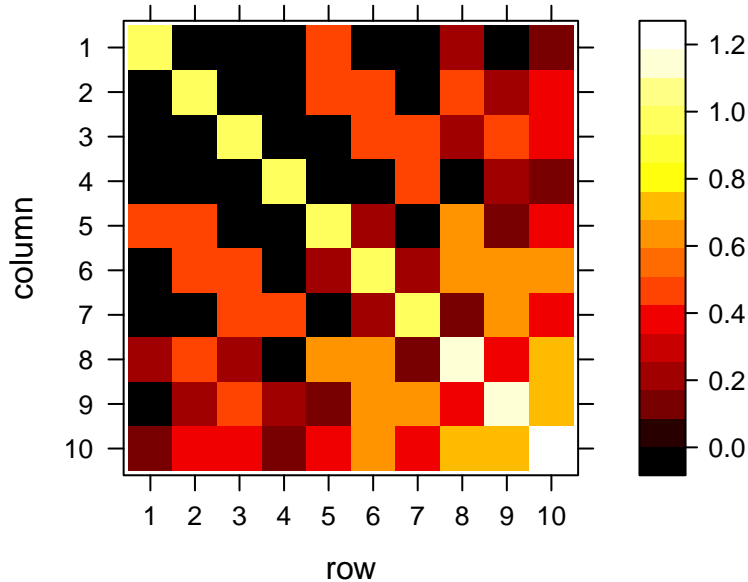
```
##   sire  dam
## 1 <NA> <NA>
## 2 <NA> <NA>
## 3 <NA> <NA>
## 4 <NA> <NA>
## 5     1    2
## 6     3    2
## 7     3    4
## 8     6    5
## 9     6    7
## 10    8    9
```

Anhand der Verwandtschaftsmatrix A

$$A = \begin{bmatrix} 1.0000 & 0.0000 & 0.0000 & 0.0000 & 0.5000 & 0.0000 & 0.0000 & 0.2500 & 0.0000 & 0.1250 \\ 0.0000 & 1.0000 & 0.0000 & 0.0000 & 0.5000 & 0.5000 & 0.0000 & 0.5000 & 0.2500 & 0.3750 \\ 0.0000 & 0.0000 & 1.0000 & 0.0000 & 0.0000 & 0.5000 & 0.5000 & 0.2500 & 0.5000 & 0.3750 \\ 0.0000 & 0.0000 & 0.0000 & 1.0000 & 0.0000 & 0.0000 & 0.5000 & 0.0000 & 0.2500 & 0.1250 \\ 0.5000 & 0.5000 & 0.0000 & 0.0000 & 1.0000 & 0.2500 & 0.0000 & 0.6250 & 0.1250 & 0.3750 \\ 0.0000 & 0.5000 & 0.5000 & 0.0000 & 0.2500 & 1.0000 & 0.2500 & 0.6250 & 0.6250 & 0.6250 \\ 0.0000 & 0.0000 & 0.5000 & 0.5000 & 0.0000 & 0.2500 & 1.0000 & 0.1250 & 0.6250 & 0.3750 \\ 0.2500 & 0.5000 & 0.2500 & 0.0000 & 0.6250 & 0.6250 & 0.1250 & 1.1250 & 0.3750 & 0.7500 \\ 0.0000 & 0.2500 & 0.5000 & 0.2500 & 0.1250 & 0.6250 & 0.6250 & 0.3750 & 1.1250 & 0.7500 \\ 0.1250 & 0.3750 & 0.3750 & 0.1250 & 0.3750 & 0.6250 & 0.3750 & 0.7500 & 0.7500 & 1.1875 \end{bmatrix}$$

lässt sich dann die folgende Heatmap erzeugen. Die nachfolgenden R-Statements zeigen gleich, wie eine Heatmap in R erzeugt wird, vorausgesetzt, dass das Pedigree in einer Matrix mit Namen `matApedEx1` gespeichert ist.

```
library(lattice)
new.palette=colorRampPalette(c("black","red","yellow","white"),space="rgb")
levelplot(matApedEx1[1:ncol(matApedEx1),ncol(matApedEx1):1],col.regions=new.palette(20))
```



Zerlegung der Verwandtschaftsmatrix

Die Verwandtschaftsmatrix A kann in ein Produkt von drei Faktoren zerlegt werden. Diese Zerlegung lautet

$$A = LDL^T \quad (1)$$

wobei L eine linke untere Dreiecksmatrix ist und D einer Diagonalmatrix entspricht. Aufgrund dieser Zerlegung lässt sich die Inverse A^{-1} der Verwandtschaftsmatrix A sehr einfach berechnen.

Herleitung der Zerlegung von A

Die Herleitung der Zerlegung in (1) basiert auf der Tatsache, dass die Varianz der Zuchtwerte dem Produkt aus Verwandtschaftsmatrix A und genetisch-additiver Varianz σ_a^2 entspricht. In Matrix-Vektor-Schreibweise heisst das

$$\text{var}(a) = A * \sigma_a^2 \quad (2)$$

wobei a der Vektor mit den Zuchtwerten aller Tiere in der Population darstellt und A der Verwandtschaftsmatrix entspricht.

Der Zuchtwert a_i eines Tieres i mit Mutter d und Vater s lässt sich zerlegen als

$$a_i = \frac{1}{2} a_s + \frac{1}{2} a_d + m_i \quad (3)$$

wobei a_s und a_d die Zuchtwerte der Eltern s und d von Tier i ist und m_i dem ‘‘Mendelian Sampling’’-Effekt entspricht. Mendelian Sampling-Effekte entstehen durch die zufällige Auswahl der Elternallele für die Nachkommen. Durch diesen zufälligen Auswahlprozess können bei Geschwistern eine Häufung von Allelen mit positiver (oder negativer) Wirkung auftreten. Somit haben Vollgeschwister nicht den gleichen Zuchtwert. Diese Variation der Zuchtwerte unter Vollgeschwister wird mit den m_i -Effekten modelliert.

In Matrix-Vektor-Schreibweise können wir die Zerlegung in (3) für die gesamte Population schreiben als

$$a = P * a + m \tag{4}$$

Ein Beispiel

Zur Veranschaulichung verwenden wir das folgende Beispiel. Gegeben sei das folgende Pedigree

```
##   sire  dam
## 1 <NA> <NA>
## 2 <NA> <NA>
## 3   1   2
## 4   1 <NA>
## 5   4   3
## 6   5   2
```

Die Zerlegung (3) für Tier 3 mit bekannten Eltern 1 und 2 lautet

$$a_3 = \frac{1}{2} a_1 + \frac{1}{2} a_2 + m_3$$

Analog dazu für Tier 4 mit bekanntem Vater 1 und unbekannter Mutter

$$a_4 = \frac{1}{2} a_1 + m_4$$

Für Tiere mit unbekanntem Eltern, wie 1 oder 2 ist der Zuchtwert einfach dem m_i Effekt. Somit ist

$$a_1 = m_1$$

und

$$a_2 = m_2$$

Fassen wir die Zerlegungen aller Tiere zusammen, so resultiert

$$\begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \\ a_6 \end{bmatrix} = \begin{bmatrix} 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.50 & 0.50 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.50 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.50 & 0.50 & 0.00 & 0.00 \\ 0.00 & 0.50 & 0.00 & 0.00 & 0.50 & 0.00 \end{bmatrix} * \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \\ a_6 \end{bmatrix} + \begin{bmatrix} m_1 \\ m_2 \\ m_3 \\ m_4 \\ m_5 \\ m_6 \end{bmatrix}$$

Rekursive Zerlegung aller Zuchtwerte

In der Zerlegung des Zuchtwertes a_i für Tier i in (3) kommen die Zuchtwerte a_s und a_d der Eltern s und d vor. Diese können analog zu a_i zerlegt werden in

$$a_s = \frac{1}{2} a_{ss} + \frac{1}{2} a_{sd} + m_s$$

wobei a_{ss} und a_{sd} die Zuchtwerte der Eltern von s sind. Analog kann der Zuchtwert für a_d zerlegt werden.

$$a_d = \frac{1}{2} a_{ds} + \frac{1}{2} a_{dd} + m_d$$

Setzen wir diese Zerlegungen in (3) ein, so erhalten wir eine neue Zerlegung des Zuchtwertes a_i in die Zuchtwerte der Grosseltern und in die entsprechenden m -Effekte.

$$\begin{aligned} a_i &= \frac{1}{2} \left(\frac{1}{2} a_{ss} + \frac{1}{2} a_{sd} + m_s \right) + \frac{1}{2} \left(\frac{1}{2} a_{ds} + \frac{1}{2} a_{dd} + m_d \right) + m_i \\ &= \frac{1}{4} a_{ss} + \frac{1}{4} a_{sd} + \frac{1}{4} a_{ds} + \frac{1}{4} a_{dd} + \frac{1}{2} m_s + \frac{1}{2} m_d + m_i \end{aligned} \quad (5)$$

Diese rekursive Zerlegung der Zuchtwerte lässt sich fortführen bis wir bei Ahnen sind von i , welche keine bekannten Eltern mehr haben. Somit haben wir die Zuchtwerte zu Linearkombinationen der m -Effekte zerlegt. In Matrix-Vektor-Schreibweise sieht diese Zerlegung wie folgt aus.

$$a = L * m \quad (6)$$

wobei L eine rechte untere Dreiecksmatrix ist mit lauter Einsen auf der Diagonalen. Die Offdiagonalelemente zeigen für jedes Tier den Pfad zu den verwandten Foundertieren der Population, wobei Tiere ohne Eltern als Foundertiere bezeichnet werden. Ist Tier i ein Nachkomme von s und d , so lassen sich die Offdiagonalelemente der i -ten Zeile als Mittelwert zwischen den Zeilen s und d berechnen.

Für unser Beispiel sieht die Zerlegung in (6) wie folgt aus.

$$\begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \\ a_6 \end{bmatrix} = \begin{bmatrix} 1.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 1.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.50 & 0.50 & 1.00 & 0.00 & 0.00 & 0.00 \\ 0.50 & 0.00 & 0.00 & 1.00 & 0.00 & 0.00 \\ 0.50 & 0.25 & 0.50 & 0.50 & 1.00 & 0.00 \\ 0.25 & 0.62 & 0.25 & 0.25 & 0.50 & 1.00 \end{bmatrix} * \begin{bmatrix} m_1 \\ m_2 \\ m_3 \\ m_4 \\ m_5 \\ m_6 \end{bmatrix}$$

Varianzen der Zuchtwerte

Kehren wir zurück zur Zerlegung des Zuchtwerts a_i aus (3) und berechnen daraus die Varianz $var(a_i)$ so folgt

$$var(a_i) = var\left(\frac{1}{2} a_s + \frac{1}{2} a_d + m_i\right) = \frac{1}{4} var(a_s) + \frac{1}{4} var(a_d) + \frac{1}{2} cov(a_s, a_d) + var(m_i) \quad (7)$$

wobei aufgrund der Elemente aus der Verwandtschaftsmatrix A :

$$\begin{aligned}
\text{var}(a_i) &= (1 + F_i)\sigma_a^2 \\
\text{var}(a_s) &= (1 + F_s)\sigma_a^2 \\
\text{var}(a_d) &= (1 + F_d)\sigma_a^2 \\
\text{cov}(a_s, a_d) &= a_{sd}\sigma_a^2 = 2F_i\sigma_a^2
\end{aligned} \tag{8}$$

In (7) wurde auch verwendet, dass keine Kovarianzen zwischen den Zuchtwerten und den m -Effekten bestehen. Setzen wir die Beziehungen aus (8) in (7) ein und lösen nach $\text{var}(m_i)$, dann folgt

$$\begin{aligned}
\text{var}(m_i) &= \text{var}(a_i) - \left[\frac{1}{4} \text{var}(a_s) + \frac{1}{4} \text{var}(a_d) + \frac{1}{2} \text{cov}(a_s, a_d) \right] \\
&= (1 + F_i)\sigma_a^2 - \frac{1}{4} (1 + F_s)\sigma_a^2 - \frac{1}{4} (1 + F_d)\sigma_a^2 - \frac{1}{2} 2F_i\sigma_a^2 \\
&= \left[\frac{1}{2} - \frac{1}{4} (F_s + F_d) \right] \sigma_a^2
\end{aligned} \tag{9}$$

Ist ein Elternteil von i unbekannt, dann lautet die Zerlegung des Zuchtwerts a_i

$$a_i = \frac{1}{2} a_d + m_i$$

und daraus abgeleitet ist

$$\text{var}(m_i) = \left[1 - \frac{1}{4} (1 + F_d) \right] \sigma_a^2 = \left[\frac{3}{4} - \frac{1}{4} F_d \right] \sigma_a^2 \tag{10}$$

Sind beide Eltern unbekannt, dann ist

$$a_i = m_i$$

und

$$\text{var}(m_i) = \sigma_a^2 \tag{11}$$

Die Berechnungen der Varianzen der m -Effekte war eine Vorbereitung für die Berechnung der Varianz der Zuchtwerte. In Matrix-Vektor-Schreibweise hatten wir die Zuchtwerte nach (6) in Linearkombinationen der m -Effekte zerlegt. Berechnen wir aufgrund von (6) die Kovarianz-Matrix des Vektors a der Zuchtwerte so folgt

$$\text{var}(a) = \text{var}(L * m) = L * \text{var}(m) * L^T \tag{12}$$

wobei $\text{var}(m)$ eine Diagonalmatrix mit den Elementen $\text{var}(m_i)$ ist. Die Elemente $\text{var}(m_i)$ werden aufgrund von (9), (10) oder (11) berechnet, je nachdem ob i zwei bekannte Eltern, einen bekannten Elternteil oder unbekannte Eltern hat. Die Kovarianzmatrix $\text{var}(m)$ ist diagonal, da die m -Effekte unabhängig sind von einander. Somit treten zwischen m_i und m_j für $i \neq j$ keine Kovarianzen auf. Da in allen Elementen $\text{var}(m_i)$ die genetisch additive Varianz σ_a^2 vorkommt, können wir diese als skalen Faktor ausklammern. Somit gilt

$$\text{var}(m) = D * \sigma_a^2$$

Setzen wir das in (12) so folgt

$$\text{var}(a) = L * \text{var}(m) * L^T = L * D * L^T \sigma_a^2 \quad (13)$$

Da wir von Gleichung (2) wissen, dass die Varianz $\text{var}(a)$ der Zuchtwerte gleich der Verwandtschaftsmatrix A mal die genetisch additive Varianz σ_a^2 ist, können wir das in (13) einsetzen. Dies führt zur sogenannten **LDL**-Zerlegung der Verwandtschaftsmatrix A .

$$\text{var}(a) = A * \sigma_a^2 = L * D * L^T \sigma_a^2 \quad (14)$$

Aus (14) folgt

$$A = L * D * L^T$$

was wir schon am Anfang dieses Kapitels in (1) postuliert hatten.

Inverse von A

Der Grund, weshalb wir die Verwandtschaftsmatrix A zerlegt hatten ist, dass die Berechnung der Inversen A^{-1} so einfacher ist, als über die explizite Inversion von A direkt. Aufgrund der Zerlegung von A in (1) erhalten wir für die Inverse A^{-1}

$$A^{-1} = (L * D * L^T)^{-1} = (L^T)^{-1} * D^{-1} * L^{-1} = (L^{-1})^T * D^{-1} * L^{-1} \quad (15)$$

Die Matrix D ist eine Diagonalmatrix und somit einfach zu invertieren. Die Inverse D^{-1} ist wieder eine Diagonalmatrix mit Elementen $1/\text{var}(m_i)$. Die Matrix L^{-1} können wir durch das Gleichsetzen von den Gleichungen (4) und (6).

$$\begin{aligned} a &= P * a + m \\ a &= L * m \end{aligned} \quad (16)$$

Lösen wir die beiden Gleichungen in (16) nach m auf, dann erhalten wir

$$\begin{aligned} m &= a - P * a = (I - P) * a \\ m &= L^{-1} * a \end{aligned} \quad (17)$$

wobei I die Einheitsmatrix ist und P entspricht der Matrix, welche in Zerlegung (4) die Nachkommen- und die Eltern-Zuchtwerte miteinander verknüpft. Da in (17) beide Gleichungen einen Ausdruck für den Vektor m enthalten, können wir die Gleichungen gleichsetzen und somit die Matrix L^{-1} bestimmen als

$$L^{-1} = I - P$$

Unser Beispiel

Für unser Beispiel ist die Matrix D gegeben als

$$D = \begin{bmatrix} 1.00000 & 0.00000 & 0.00000 & 0.00000 & 0.00000 & 0.00000 \\ 0.00000 & 1.00000 & 0.00000 & 0.00000 & 0.00000 & 0.00000 \\ 0.00000 & 0.00000 & 0.50000 & 0.00000 & 0.00000 & 0.00000 \\ 0.00000 & 0.00000 & 0.00000 & 0.75000 & 0.00000 & 0.00000 \\ 0.00000 & 0.00000 & 0.00000 & 0.00000 & 0.50000 & 0.00000 \\ 0.00000 & 0.00000 & 0.00000 & 0.00000 & 0.00000 & 0.46875 \end{bmatrix}$$

Die Inverse D^{-1} ist

$$D^{-1} = \begin{bmatrix} 1.00000 & 0.00000 & 0.00000 & 0.00000 & 0.00000 & 0.00000 \\ 0.00000 & 1.00000 & 0.00000 & 0.00000 & 0.00000 & 0.00000 \\ 0.00000 & 0.00000 & 2.00000 & 0.00000 & 0.00000 & 0.00000 \\ 0.00000 & 0.00000 & 0.00000 & 1.33333 & 0.00000 & 0.00000 \\ 0.00000 & 0.00000 & 0.00000 & 0.00000 & 2.00000 & 0.00000 \\ 0.00000 & 0.00000 & 0.00000 & 0.00000 & 0.00000 & 2.13333 \end{bmatrix}$$

Die Matrix P zur Berechnung von L^{-1} hatten wir aufgestellt als

$$P = \begin{bmatrix} 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.50 & 0.50 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.50 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.50 & 0.50 & 0.00 & 0.00 \\ 0.00 & 0.50 & 0.00 & 0.00 & 0.50 & 0.00 \end{bmatrix}$$

Somit ist L^{-1}

$$L^{-1} = \begin{bmatrix} 1.00000 & 0.00000 & 0.00000 & 0.00000 & 0.00000 & 0.00000 \\ 0.00000 & 1.00000 & 0.00000 & 0.00000 & 0.00000 & 0.00000 \\ -0.50000 & -0.50000 & 1.00000 & 0.00000 & 0.00000 & 0.00000 \\ -0.50000 & 0.00000 & 0.00000 & 1.00000 & 0.00000 & 0.00000 \\ 0.00000 & 0.00000 & -0.50000 & -0.50000 & 1.00000 & 0.00000 \\ 0.00000 & -0.50000 & 0.00000 & 0.00000 & -0.50000 & 1.00000 \end{bmatrix}$$

Somit haben die Faktoren von A^{-1} und können diese direkt aufstellen als

$$A^{-1} = \begin{bmatrix} 1.83333 & 0.50000 & -1.00000 & -0.66667 & 0.00000 & 0.00000 \\ 0.50000 & 2.03333 & -1.00000 & 0.00000 & 0.53333 & -1.06667 \\ -1.00000 & -1.00000 & 2.50000 & 0.50000 & -1.00000 & 0.00000 \\ -0.66667 & 0.00000 & 0.50000 & 1.83333 & -1.00000 & 0.00000 \\ 0.00000 & 0.53333 & -1.00000 & -1.00000 & 2.53333 & -1.06667 \\ 0.00000 & -1.06667 & 0.00000 & 0.00000 & -1.06667 & 2.13333 \end{bmatrix}$$