

Züchtungslehre - Herleitungen

Peter von Rohr

December 9, 2015

Erklärung

Auf den Folien des ersten Teils der Vorlesung Züchtungslehre wurden aus Platzgründen die Herleitungen von einigen Formeln weggelassen. Diese Herleitungen werden in diesem Dokument zusammengefasst. Das hier gezeigte Material soll als zusätzliche Erklärung und als Hilfestellung dienen und nicht den Prüfungsstoff erweitern.

Das hier gezeigte Material basiert auf folgenden Quellen

- Skript der Vorlesung Quantitative Genetik der ETHZ [1]
- Diss Andreas Hofer [2]
- Buch Linear Models for the Prediction of Animal Breeding Values von R.A. Mrode [3]

1 Selektionsindex

In der Vorlesung wurde der Selektionsindex I als Schätzer für den Gesamtzuchtwert H vorgestellt. Der Gesamtzuchtwert H entspricht der mathematischen Formulierung unseres Zuchtziels. Das Zuchtziel ist die Charakterisierung des optimalen Zuchttieres. In den meisten Fällen enthält das Zuchtziel mehrere Merkmale.

Der Gesamtzuchtwert ist als folgendes gewichtetes Mittel definiert.

$$H = \mathbf{v}^T * \mathbf{g} \quad (1)$$

wobei: \mathbf{v} der Vektor der wirtschaftlichen Gewichte der Merkmale im Gesamtzuchtwert ist und \mathbf{g} der Vektor der wahren Zuchtwerte der Merkmale im Gesamtzuchtwert.

1.1 Selektionsindex als Schätzung des Gesamtzuchtwerks

Werden die wirtschaftlichen Gewichte nach wirtschaftlichen Kriterien bestimmt und werden die Elterntiere nach dem Gesamtzuchtwert rangiert und selektiert, so resultiert ein optimaler erwarteter wirtschaftlicher Erfolg in der Nachkomengeneration.

Leider ist der Gesamtzuchtwert nicht beobachtbar oder messbar. Deshalb müssen wir uns mit einer Schätzung behelfen. Als eine mögliche Schätzung für den Gesamtzuchtwert H können wir den Selektionsindex I verwenden.

1.2 Definition des Selektionsindex

Der Selektionsindex wird, analog zum Gesamtzuchtwert als gewichtetes Mittel postuliert. Formell sieht die Definition des Selektionsindex wie folgt aus.

$$I = \mathbf{b}^T * \mathbf{p} \quad (2)$$

wobei: \mathbf{b} der Vektor der Indexgewichte ist und \mathbf{p} der Vektor der verfügbaren Informationsquellen.

1.3 Informationsquellen im Selektionsindex

Im einfachsten Fall sind im Gesamtzuchtwert und im Selektionsindex die gleichen Merkmale enthalten. In diesem Fall werden als Informationsquellen die aus dem Tiermodell geschätzten Zuchtwerte oder genomische Zuchtwerte verwendet und dann entsprechen die Indexgewichte den wirtschaftlichen Gewichten.

Hat es in einem Gesamtzuchtwert Merkmale, welche nur schwer messbar sind, oder welche erst spät im Leben eines Tieres beobachtbar sind, dann werden häufig Hilfsmerkmale im Selektionsindex berücksichtigt. Diese Hilfsmerkmale müssen eine hohe Korrelation zum eigentlichen Merkmal aufweisen, einfach messbar sein und möglichst früh im Leben eines Tieres beobachtbar sein.

Zusammenfassend kann gesagt werden, dass im Selektionsindex nach Möglichkeit die gleichen Merkmale wie im Gesamtzuchtwert verwendet werden. Falls dies zu teuer oder schwer umsetzbar ist, dann werden im Selektionsindex Hilfsmerkmale verwendet. Der Vektor \mathbf{p} ist somit gegeben. Die verbleibende Unbekannte ist der Vektor \mathbf{b} mit den Indexgewichten.

1.4 Herleitung der Indexgewichte

Der Vektor \mathbf{b} mit Indexgewichten wird so bestimmt, dass folgende Kriterien erfüllt sind.

1. Die Varianz der Abweichung zwischen H und I soll minimal sein, oder
2. Die Korrelation zwischen H und I soll maximal sein

1.4.1 Minimierung der Varianz der Abweichungen

Die Varianz der Abweichung zwischen Gesamtzuchtwert H und Selektionsindex I wird wie folgt minimiert. Dabei gilt es zu beachten, dass der Vektor der Indexgewichte \mathbf{b} als unbekannte Grösse gesucht wird.

$$\begin{aligned}
Var(H - I) &= Var(H) + Var(I) - 2 * Cov(H, I) \\
&= Var(\mathbf{v}^T * \mathbf{g}) + Var(\mathbf{b}^T * \mathbf{p}) - 2 * Cov(\mathbf{v}^T * \mathbf{g}, \mathbf{b}^T * \mathbf{p}) \\
&= \mathbf{v}^T * Var(\mathbf{g}) * \mathbf{v} + \mathbf{b}^T * Var(\mathbf{p}) * \mathbf{b} - 2 * \mathbf{v}^T * Cov(\mathbf{g}, \mathbf{p}) * \mathbf{b} \\
&= \mathbf{v}^T * \mathbf{C} * \mathbf{v} + \mathbf{b}^T * \mathbf{P} * \mathbf{b} - 2 * \mathbf{v}^T * \mathbf{G} * \mathbf{b} \tag{3}
\end{aligned}$$

wobei \mathbf{C} die Kovarianzmatrix zwischen den Merkmalen im Gesamtzuchtwert ist, \mathbf{P} die Kovarianzmatrix zwischen den Merkmalen im Selektionsindex ist, und \mathbf{G} die Kovarianzmatrix zwischen den Merkmalen im Gesamtzuchtwert und den Merkmalen im Selektionsindex ist.

In der Herleitung der Varianz der Abweichung des Selektionsindex vom Gesamtzuchtwert in Gleichung (3) wird im ersten Schritt die Definition der Varianz einer Differenz angewendet. Danach werden die Definitionen von H und I eingesetzt. Aus der Entwicklung der Formeln in Gleichung (3) können folgende Beziehungen zur Varianz des Gesamtzuchtwerthes, zur Varianz des Selektionsindex und zur Kovarianz zwischen Gesamtzuchtwert und Selektionsindex abgeleitet werden.

$$\begin{aligned}
Var(H) &= \mathbf{v}^T * \mathbf{C} * \mathbf{v} \\
Var(I) &= \mathbf{b}^T * \mathbf{P} * \mathbf{b} \\
Cov(H, I) &= \mathbf{v}^T * \mathbf{G} * \mathbf{b} \tag{4}
\end{aligned}$$

Zum Minimieren von $Var(H - I)$ muss dieser Ausdruck in Gleichung (3) nach \mathbf{b} abgeleitet werden (Siehe Anhang A für eine allgemeine Erklärung von Ableitungen nach Vektoren). Dabei gilt es zu beachten, dass die quadratische Form $\mathbf{b}^T * \mathbf{P} * \mathbf{b}$ analog zur skalaren Funktion x^2 abgeleitet wird, d.h. der quadratische Exponent wird in der Ableitung zum Faktor 2 und Grad der Funktion wird um eins reduziert, d.h. aus der quadratischen Form wird in der Ableitung die lineare Form $\mathbf{P} * \mathbf{b}$. Der Term $\mathbf{v}^T * \mathbf{C} * \mathbf{v}$ ist unabhängig von \mathbf{b} , d.h. es kommt darin kein \mathbf{b} vor und deshalb fällt dieser Term bei der Ableitung heraus. Der letzte Term in Gleichung 3 zeigt eine lineare Abhängigkeit in \mathbf{b} , was bedeutet, dass in der Ableitung noch der Konstante Faktor $-2 * \mathbf{G} * \mathbf{v}$ übrig bleibt.

$$\frac{\partial Var(H - I)}{\partial \mathbf{b}} = 2 * \mathbf{P} * \mathbf{b} - 2 * \mathbf{G} * \mathbf{v} \tag{5}$$

Die in Gleichung (5) gezeigte Ableitung wird $= 0$ gesetzt und daraus kann dann der unbekannte Vektor \mathbf{b} berechnet werden.

$$\begin{aligned}
\frac{\partial Var(H - I)}{\partial \mathbf{b}} &= 2 * \mathbf{P} * \mathbf{b} - 2 * \mathbf{G} * \mathbf{v} = 0 \\
2 * \mathbf{P} * \mathbf{b} &= 2 * \mathbf{G} * \mathbf{v} \\
\mathbf{P} * \mathbf{b} &= \mathbf{G} * \mathbf{v} \tag{6}
\end{aligned}$$

Die letzte Zeile in Gleichung (6) entspricht den in der Vorlesung vorgestellten Indexgleichungen. Werden diese Gleichungen nach \mathbf{b} aufgelöst, dann erhalten wir für den unbekannten Vektor \mathbf{b} der Indexgewichte folgende Beziehung.

$$\mathbf{b} = \mathbf{P}^{-1} * \mathbf{G} * \mathbf{v} \quad (7)$$

1.4.2 Maximierung der Korrelation zwischen Selektionsindex und Gesamtzuchtwert

Die Korrelation zwischen Selektionsindex und Gesamtzuchtwert wird als $r_{H,I}$ bezeichnet und hat folgende Form.

$$r_{H,I} = \frac{Cov(H, I)}{\sqrt{Var(I) * Var(H)}} \quad (8)$$

Das Ziel hier ist wiederum die Bestimmung des unbekannten Vektors der Indexgewichte \mathbf{b} so, dass die Korrelation $r_{H,I}$ maximal wird. Zu diesem Zweck muss wieder die Ableitung von $r_{H,I}$ nach \mathbf{b} berechnet und dann gleich 0 gesetzt werden. Derjenige Vektor für \mathbf{b} , welcher dabei herauskommt, ist dann unser gesuchter Vektor der Indexgewichte. Als Vereinfachung werden wir anstelle von der Ableitung von $r_{H,I}$ nach \mathbf{b} die Ableitung vom natürlichen Logarithmus von $r_{H,I}$ nach \mathbf{b} berechnen. Dies ist wie gesagt einfacher und wird zum gleichen Ergebnis für \mathbf{b} führen.

$$\begin{aligned} \log(r_{H,I}) &= \log\left(\frac{Cov(H, I)}{\sqrt{Var(I) * Var(H)}}\right) \\ &= \log(Cov(H, I)) - \frac{1}{2} \log(Var(I)) - \frac{1}{2} \log(Var(H)) \\ &= \log(\mathbf{v}^T * \mathbf{G} * \mathbf{b}) - \frac{1}{2} \log(\mathbf{b}^T * \mathbf{P} * \mathbf{b}) \\ &\quad - \frac{1}{2} \log(\mathbf{v}^T * \mathbf{C} * \mathbf{v}) \end{aligned} \quad (9)$$

Die Ableitung von $\log(r_{H,I})$ nach \mathbf{b} wird wie folgt berechnet.

$$\begin{aligned} \frac{\partial \log(r_{H,I})}{\partial \mathbf{b}} &= \frac{\mathbf{G} * \mathbf{v}}{\mathbf{v}^T * \mathbf{G} * \mathbf{b}} - \frac{1}{2} \frac{2 * \mathbf{P} * \mathbf{b}}{\mathbf{b}^T * \mathbf{P} * \mathbf{b}} - \frac{1}{2} \frac{0}{\mathbf{v}^T * \mathbf{C} * \mathbf{v}} \\ &= \frac{\mathbf{G} * \mathbf{v}}{\mathbf{v}^T * \mathbf{G} * \mathbf{b}} - \frac{\mathbf{P} * \mathbf{b}}{\mathbf{b}^T * \mathbf{P} * \mathbf{b}} \\ &= \frac{\mathbf{b}^T * \mathbf{P} * \mathbf{b} * \mathbf{G} * \mathbf{v} - \mathbf{v}^T * \mathbf{G} * \mathbf{b} * \mathbf{P} * \mathbf{b}}{\mathbf{v}^T * \mathbf{G} * \mathbf{b} * \mathbf{b}^T * \mathbf{P} * \mathbf{b}} \end{aligned} \quad (10)$$

Das Maximum von $\log(r_{H,I})$ finden wir durch Nullsetzen von

$$\frac{\partial \log(r_{H,I})}{\partial \mathbf{b}} = 0$$

Daraus folgt, dass

$$\begin{aligned}\mathbf{b}^T * \mathbf{P} * \mathbf{b} * \mathbf{G} * \mathbf{v} &= \mathbf{v}^T * \mathbf{G} * \mathbf{b} * \mathbf{P} * \mathbf{b} \\ \mathbf{G} * \mathbf{v} &= \frac{\mathbf{v}^T * \mathbf{G} * \mathbf{b}}{\mathbf{b}^T * \mathbf{P} * \mathbf{b}} * \mathbf{P} * \mathbf{b}\end{aligned}\quad (11)$$

Berücksichtigen wir hier zusätzlich die Anforderung, dass der Selektionsindex I ein erwartungstreuer Schätzer vom Gesamtzuchtwert H sein soll, dann folgt, dass der Regressionskoeffizient $b_{H,I}$ vom Gesamtzuchtwert auf den Selektionsindex = 1 sein muss. Aus dieser Anforderung folgt dann, dass

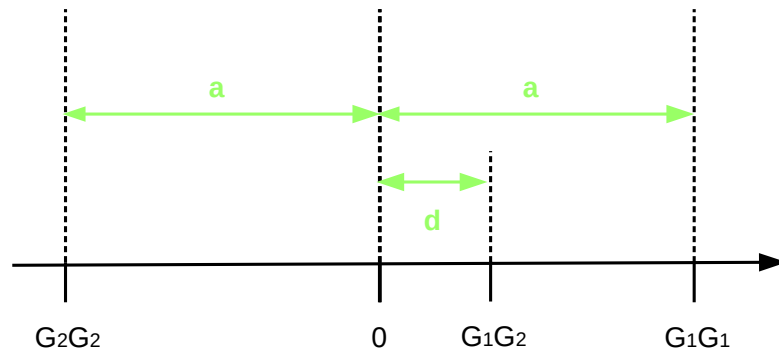
$$\begin{aligned}b_{H,I} &= \frac{Cov(H,I)}{Var(I)} \\ &= \frac{\mathbf{v}^T * \mathbf{G} * \mathbf{b}}{\mathbf{b}^T * \mathbf{P} * \mathbf{b}} \\ &= 1\end{aligned}\quad (12)$$

Setzen wir die Bedingung von Gleichung (12) in Gleichung (11) ein dann folgen die uns bekannten Indexgleichungen.

$$\mathbf{G} * \mathbf{v} = \mathbf{P} * \mathbf{b}$$

2 Genetische Varianz für das Ein-Lokus Modell

In der Vorlesung wurde das Konzept der genotypischen Werte für einen Locus mit zwei Allelen in einer Population im Hardy-Weinberg-Gleichgewicht vorgestellt. Dieses Konzept lässt sich mit folgendem Diagramm veranschaulichen



Unter der Annahme, dass die das Allel G_1 mit einer Häufigkeit von p in der Population auftritt, können die Frequenzen der Genotypen und ihre zugehörigen genotypischen Werte in folgender Tabelle zusammengefasst werden.

Genotyp	genotypischer Wert	Frequenz
G_1G_1	$V_{11} = a$	$f(G_1G_1) = p^2$
G_1G_2	$V_{12} = d$	$f(G_1G_2) = 2pq$
G_2G_2	$V_{22} = -a$	$f(G_2G_2) = q^2$

2.1 Populationsmittel

Als erste Kenngröße für die genotypischen Werte in einer bestimmten Population hatten wir das *Populationsmittel* berechnet. Das Populationsmittel ist als Erwartungswert der Zufallsvariablen der genotypischen Werte V definiert. Allgemein ist ein Erwartungswert für eine diskrete Zufallsvariable definiert als

$$E[X] = \sum_{x_i \in \mathcal{X}} x_i * f(x_i) \quad (13)$$

wobei \mathcal{X} : Menge aller möglichen x -Werte
 $f(x_i)$: Wahrscheinlichkeit dass x den Wert x_i annimmt

Das Populationsmittel als Erwartungswert der genotypischen Werte ist dann definiert als

$$\begin{aligned}
\mu = E[V] &= f(G_1G_1) * V_{11} + f(G_1G_2) * V_{12} + f(G_2G_2) * V_{22} \\
&= p^2 * a + 2pq * d - q^2 * a \\
&= (p^2 - q^2)a + 2pqd \\
&= (p + q)(p - q)a + 2pqd \\
&= (p - q)a + 2pqd
\end{aligned} \tag{14}$$

2.2 Zuchtwerte

Der Zuchtwert eines bestimmten Elterngenotyps entspricht der doppelten Abweichung des mittleren genotypischen Wertes der Nachkommen vom Populationsmittel μ .

Die Frequenzen der Nachkommen in Abhängigkeit des Elterngenotyps kann folgender Tabelle entnommen werden.

Mutter	Vater	
	$f(G_1) = p$	$f(G_2) = q$
G_1G_1		
$f(G_1) = 1$	$f(G_1G_1) = p$	$f(G_1G_2) = q$
G_1G_2		
$f(G_1) = 0.5$	$f(G_1G_1) = 0.5p$	$f(G_1G_2) = 0.5q$
$f(G_2) = 0.5$	$f(G_1G_2) = 0.5p$	$f(G_2G_2) = 0.5q$
G_2G_2		
$f(G_2) = 1$	$f(G_1G_2) = p$	$f(G_2G_2) = q$

2.2.1 Elterngenotyp G_1G_1

Der mittlere genotypische Wert der Nachkommen vom Elterngenotyp beträgt

$$\mu_{11} = pa + qd$$

Der Zuchtwert des Elterngenotyp kann nun berechnet werden.

$$\begin{aligned}
ZW_{11} &= 2 * (\mu_{11} - \mu) \\
&= 2(pa + qd - [(p - q)a + 2pqd]) \\
&= 2(pa + qd - (p - q)a - 2pqd) \\
&= 2(qd + qa - 2pqd) \\
&= 2(qa + qd(1 - 2p)) \\
&= 2q(a + d(1 - 2p)) \\
&= 2q(a + (q - p)d) \\
&= 2q\alpha
\end{aligned} \tag{15}$$

2.2.2 Elterngenotyp G_1G_2

Der mittlere genotypische Wert der Nachkommen vom Elterngenotyp G_1G_2 beträgt

$$\mu_{12} = 0.5pa + 0.5pd + 0.5qd - 0.5qa = 0.5(p - q)a + 0.5d$$

Der Zuchtwert (ZW_{12}) für den Elterngenotyp (G_1G_2)

$$\begin{aligned} ZW_{12} &= 2 * (\mu_{12} - \mu) \\ &= 2 (0.5(p - q)a + 0.5d - [(p - q)a + 2pqd]) \\ &= 2 (-0.5(p - q)a + (0.5 - 2pq)d) \\ &= (q - p)a + (1 - 4pq)d \\ &= (q - p)a + (p - q)^2d \\ &= (q - p)(a + (q - p)d) \\ &= (q - p)\alpha \end{aligned} \tag{16}$$

2.2.3 Elterngenotyp G_2G_2

Der mittlere genotypische Wert der Nachkommen vom Elterngenotyp G_2G_2 beträgt

$$\mu_{22} = pd - qa$$

Der Zuchtwert (ZW_{22}) für den Elterngenotyp (G_2G_2)

$$\begin{aligned} ZW_{22} &= 2 * (\mu_{22} - \mu) \\ &= 2 (pd - qa - [(p - q)a + 2pqd]) \\ &= 2 (-pa + (p - 2pq)d) \\ &= -2p(a + (2q - 1)d) \\ &= -2p(a + (q - p)d) \\ &= -2p\alpha \end{aligned} \tag{17}$$

In den Formeln der drei Zuchtwerte ist der Term $a + (q - p)d$ enthalten. Dieser Term wird mit α bezeichnet und entspricht dem sogenannten Allelsubstitutionseffekt. Bildet man die Differenzen zwischen den Zuchtwerten der Genotypen, welche sich jeweils durch ein G_1 Allel unterscheiden, dann sind diese Differenzen immer gleich α , oder formell heisst das, dass

$$ZW_{11} - ZW_{12} = ZW_{12} - ZW_{22} = \alpha$$

2.3 Dominanzabweichungen

Die Differenzen zwischen den genotypischen Werten und den Zuchtwerten entsprechen dem Populationsmittel plus der sogenannten Dominanzabweichungen D_{ij} für

den Genotypen G_iG_j . Für die drei Genotypen G_1G_1 , G_1G_2 und G_2G_2 sieht das wie folgt aus.

2.3.1 Genotyp G_1G_1

$$\begin{aligned}
V_{11} - ZW_{11} &= a - 2q\alpha \\
&= a - 2q(a + (q - p)d) \\
&= (1 - 2q)a - 2qd(q - p) \\
&= [(p - q)a + 2pqd] - 2q^2d \\
&= \mu + D_{11}
\end{aligned} \tag{18}$$

2.3.2 Genotyp G_1G_2

$$\begin{aligned}
V_{12} - ZW_{12} &= d - [(q - p)\alpha] \\
&= d - [(q - p)(a + (q - p)d)] \\
&= [(p - q)a + 2pqd] + 2pqd \\
&= \mu + D_{12}
\end{aligned} \tag{19}$$

2.3.3 Genotyp G_2G_2

$$\begin{aligned}
V_{22} - ZW_{22} &= -a + 2p\alpha \\
&= -a + 2p[a + (q - p)d] \\
&= (2p - 1)a + 2pqd - 2p^2d \\
&= [(p - q)a + 2pqd] - 2p^2d \\
&= \mu + D_{22}
\end{aligned} \tag{20}$$

2.3.4 Konsequenz und Zusammenfassung

Die Differenzen zwischen den genotypischen Werten und Zuchtwerten konnten für alle drei Genotypen als das Populationsmittel plus eine spezifische Dominanzabweichung dargestellt werden. Somit lässt sich der Genotypische Wert zerlegen in das Populationsmittel, den Zuchtwert und die Dominanzabweichung.

$$V_{ij} = \mu + ZW_{ij} + D_{ij} \tag{21}$$

Folgende Tabelle gibt eine Zusammenfassung über alle Werte für V_{ij} , ZW_{ij} und D_{ij} .

Genotyp	genotypischer Wert	Zuchtwert	Dominanzabweichung
G_iG_j	V_{ij}	ZW_{ij}	D_{ij}
G_1G_1	a	$2q\alpha$	$-2q^2d$
G_1G_2	d	$(q - p)\alpha$	$2pqd$
G_2G_2	$-a$	$-2p\alpha$	$-2p^2d$

2.4 Genetische Varianz

Die Varianz einer diskreten Zufallsvariablen X ist definiert als

$$Var[X] = \sum_{x_i \in \mathcal{X}} (x_i - \mu_X)^2 * f(x_i) \quad (22)$$

wobei \mathcal{X} : Menge aller möglichen x -Werte
 $f(x_i)$ Wahrscheinlichkeit dass x den Wert x_i annimmt
 μ_X Erwartungswert $E[X]$ von X

Für unsere genotypischen Werte V_{ij} heisst das

$$\begin{aligned} \sigma_G^2 = Var[V] &= (V_{11} - \mu)^2 * f(G_1G_1) \\ &\quad + (V_{12} - \mu)^2 * f(G_1G_2) \\ &\quad + (V_{22} - \mu)^2 * f(G_2G_2) \end{aligned} \quad (23)$$

wobei $\mu = (p - q)a + 2pqd$ das Populationsmittel

Für die Frequenzen setzen wir die Werte aufgrund des Hardy-Weinberg-Gleichgewichts ein. Aus der Zerlegung der genotypischen Werte in Gleichung (21) verwenden wir, dass $V_{ij} - \mu = ZW_{ij} + D_{ij}$. Setzen wir das in Gleichung (23) ein dann folgt daraus

$$\begin{aligned} \sigma_G^2 = Var[V] &= (ZW_{11} + D_{11})^2 * p^2 \\ &\quad + (ZW_{12} + D_{12})^2 * 2pq \\ &\quad + (ZW_{22} + D_{22})^2 * q^2 \end{aligned} \quad (24)$$

Für die Zuchtwerte (ZW_{ij}) und die Dominanzabweichungen (D_{ij}) setzen wir nun die Werte aus der Übersichtstabelle ein.

$$\begin{aligned} \sigma_G^2 = Var[V] &= (2q\alpha - 2q^2d)^2 * p^2 \\ &\quad + ((q - p)\alpha + 2pqd)^2 * 2pq \\ &\quad + (-2p\alpha - 2p^2d)^2 * q^2 \\ &= (4q^2\alpha^2 - 8q^3d\alpha + 4q^4d^2) * p^2 \\ &\quad + (q^2\alpha^2 - 2pq\alpha^2 + p^2\alpha^2 - 4(q - p)pqd\alpha + 4p^2q^2d^2) * 2pq \\ &\quad + (4p^2\alpha^2 + 8p^3d\alpha + 4p^4\alpha^2) * q^2 \\ &= 4p^2q^2\alpha^2 - 8p^2q^3d\alpha + 4p^2q^4d^2 \\ &\quad + 2pq^3\alpha^2 - 4p^2q^2\alpha^2 + 2p^3q\alpha^2 \\ &\quad - 8p^3q^2d\alpha + 8p^2q^3d\alpha + 8p^3q^3d^2 \\ &\quad + 4p^2q^2\alpha^2 + 8p^3q^2d\alpha + 4p^4q^2d^2 \end{aligned}$$

$$\begin{aligned}
&= 4p^2q^2\alpha^2 + 4p^2q^4d^2 \\
&\quad + 2pq^3\alpha^2 + 2p^3q\alpha^2 \\
&\quad + 8p^3q^3d^2 \\
&\quad + 4p^4q^2d^2 \\
&= 2pq\alpha^2(p^2 + 2pq + q^2) \\
&\quad + (2pqd)^2(p^2 + 2pq + q^2) \\
&= 2pq\alpha^2 + (2pqd)^2 \tag{25}
\end{aligned}$$

Das Resultat der Berechnung der genetischen Varianz σ_G^2 für einen Locus mit zwei Allelen zeigt, dass diese Varianz aus zwei Teilen besteht. Der erste Teil ist der additive Varianzanteil ($2pq\alpha^2$) und der zweite Teil heisst Dominanzvarianzanteil $((2pqd)^2)$.

Diese Herleitung zeigt, dass die in der Vorlesung vorgestellte genetische Varianz anhand der Definition der Varianz als zweites zentrales Moment einer diskreten Zufallsvariablen für die genotypischen Werte berechnet werden kann.

3 Varianz der Mendelian Sampling Effekte

3.1 Notation

In diesem Abschnitt verwenden wir eine leicht andere Notation, als in der Vorlesung. Der Grund dafür ist, dass mögliche Verwechslungen und Kollisionen zwischen gleichen Bezeichnungen von unterschiedlichen Grössen verhindert werden sollen.

3.1.1 Zuchtwerte

Zuchtwerte werden mit der Variablen “ u ” bezeichnet. Ein Vektor von Zuchtwerten wird mit dem kleinen \mathbf{u} in fetter Schrift bezeichnet. Der Zuchtwert u_i eines einzelnen Tieres i wird mit einem kleinen u in Normalschrift mit einem Subskript i gekennzeichnet.

Als Konsequenz der Bezeichnung der Zuchtwerte mit “ u ”, steht die Variable σ_u^2 für die additiv genetische Varianz.

Die hier verwendeten Bezeichnungen für die Zuchtwerte haben nichts zu tun mit der in der Vorlesung und in den Übungen verwendeten Matrix \mathbf{U} aus der Cholesky-Zerlegung der Verwandtschaftsmatrix.

3.1.2 Verwandtschaftsmatrix

Die Verwandtschaftsmatrix wird mit einem grossen “ \mathbf{A} ” in fetter Schrift bezeichnet. Die Komponenten aus dieser Matrix werden mit a_{ij} bezeichnet wobei der Index i für die i -te Zeile und der Index j für die j -te Kolonne steht.

3.2 Einführung

Im Gegensatz zu Abschnitt 2, in welchem wir Zuchtwerte an einem bestimmten Genort betrachtet haben, wechseln wir in diesem Abschnitt zum sogenannten “Infinitesimal”-Modell, in welchem angenommen wird, dass genetische Effekte aufgrund von unendlich vielen Genorten zustande kommen. Unter der Annahme eines solchen Modells und unter der Annahme, dass die betrachteten Effekte einer Normalverteilung folgen, kann der Zuchtwert u_i eines Tieres i mit bekannten Eltern s und d zerlegt werden in

$$u_i = \frac{1}{2} u_s + \frac{1}{2} u_d + m_i \quad (26)$$

wobei m_i für den sogenannten “Mendelian sampling” - Effekt steht.

Dieser Effekt entspricht der Abweichung des Zuchtwertes u_i vom erwarteten Vollgeschwisterdurchschnitt. Er kommt durch die zufällige Auswahl von Allelen während der Meiose bei den Eltern zustande. Die Zuchtwerte von Vollgeschwistern unterscheiden sich also nur in den unterschiedlichen Mendelian sampling Effekten, nicht aber in den Beiträgen der Elternzuchtwerte.

Fasst man die Zuchtwerte u_i aller Tiere zu einem Vektor \mathbf{u} zusammen, dann kann die Zerlegung aus Gleichung (26) in Matrixschreibweise übertragen werden.

$$\mathbf{u} = \mathbf{P} * \mathbf{u} + \mathbf{m} \quad (27)$$

wobei \mathbf{u} Vektor aller Zuchtwerte

\mathbf{m} Vektor aller Mendelian sampling Effekte

Die Matrix \mathbf{P} verbindet jedes Tier i zu seinen Eltern s und d . Das heisst auf der i -ten Zeile hat es zwei Elemente von $1/2$ in den Spalten s und d .

3.3 Beispiel

Nehmen wir folgendes Pedigree als Beispiel und stellen dafür die in Gleichung (27) gezeigte Zerlegung auf.

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

Für dieses kleine Beispielpedigree sieht die Zerlegung der Zuchtwerte wie folgt aus

$$\begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \\ u_5 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 1/2 & 0 \end{bmatrix} * \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \\ u_5 \end{bmatrix} + \begin{bmatrix} m_1 \\ m_2 \\ m_3 \\ m_4 \\ m_5 \end{bmatrix} \quad (28)$$

Die Matrixschreibweise der Zerlegung in Gleichung (28) zeigt, dass die Zuchtwerte für Tiere ohne Eltern den Mendelian sampling Effekten entsprechen. In unserem Beispielpedigree haben die Tiere 1, 2 und 4 keine bekannten Eltern. Für deren Zuchtwerte gilt also, dass $u_1 = m_1$, $u_2 = m_2$ und $u_4 = m_4$.

3.4 Rekursive Zerlegung

Kehren wir nun zurueck zur Zerlegung für ein bestimmtes Tier i . Wie aus Gleichung (26) ersichtlich ist, wird u_i in je die Hälfte der Elternzuchtwerte u_s und u_d und die Mendelian sampling Komponente m_i zerlegt. Die Zuchtwerte u_s und u_d können analog zur Zerlegung von u_i auch in die Komponenten der Elternzuchtwerte und der Mendelian sampling Effekte aufgeteilt werden. Konkret bedeutet das, dass u_s aufgeteilt wird in

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

und u_d wird zerlegt in

$$u_d = \frac{1}{2} u_{ds} + \frac{1}{2} u_{dd} + m_d$$

wobei u_{ss} und u_{sd} die Zuchtwerte der Eltern sd und ss von Vater s und u_{ds} und u_{dd} die Zuchtwerte der Eltern ds und dd der Mutter d sind.

Diese Art der rekursiven Zerlegung der Zuchtwerte wird so lange gemacht bis wir im Pedigree bei Tieren ohne Eltern angelangt sind. Deren Zuchtwerte entsprechen den entsprechenden Mendelian sampling Effekten. Somit kann jeder Zuchtwert u_i als eine Summe von Mendelian sampling Effekten dargestellt werden.

In unserem Beispieldedigree kann der Zuchtwert von Tier 5, wie folgt zerlegt werden.

$$\begin{aligned} u_5 &= \frac{1}{2} u_3 + \frac{1}{2} u_4 + m_5 \\ &= \frac{1}{2} \left(\frac{1}{2} u_1 + \frac{1}{2} u_2 + m_3 \right) + \frac{1}{2} m_4 + m_5 \\ &= \frac{1}{4} m_1 + \frac{1}{4} m_2 + \frac{1}{2} m_3 + \frac{1}{2} m_4 + m_5 \end{aligned} \quad (29)$$

Fassen wir wieder die Zuchtwerte aller Tiere in einer Population in einem Vektor zusammen, und ordnen die Tiere so, dass Elterntiere vor ihren Nachkommen erscheinen, dann lassen sich die in Gleichung (29) gefundene Beziehung zwischen Zuchtwerten und Mendelian sampling Effekten in Matrixschreibweise wie folgt ausdrücken.

$$\mathbf{u} = \mathbf{L} * \mathbf{m} \quad (30)$$

Matrix \mathbf{L} in Gleichung (30) ist eine untere Dreiecksmatrix, d.h. alle Elemente rechts der Diagonalen sind gleich null. Die Diagonalelemente der Matrix \mathbf{L} sind alle gleich eins. Für unser Beispiel sieht die Matrix \mathbf{L} wie folgt aus.

$$\begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \\ u_5 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 1/2 & 1/2 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 1/4 & 1/4 & 1/2 & 1/2 & 1 \end{bmatrix} * \begin{bmatrix} m_1 \\ m_2 \\ m_3 \\ m_4 \\ m_5 \end{bmatrix} \quad (31)$$

Bei näherer Betrachtung der Matrix \mathbf{L} fallen zwei Fakten auf.

1. Wie schon weiter oben erwähnt sind alle Diagonalelemente der Matrix \mathbf{L} gleich 1. Da aufgrund der Zerlegung in Gleichung (26), jeder Zuchtwert eines Tieres i die Mendelian Sampling Komponente m_i selber enthält, folgt daraus, dass das entsprechende Diagonalelement l_{ii} der Matrix \mathbf{L} gleich 1 sein muss.

2. Die Offdiagonalelemente l_{ij} wobei $i > j$ lassen sich berechnen als

$$l_{ij} = \frac{1}{2} * (l_{sj} + l_{dj})$$

wobei s und d die Eltern von i sind. Das Element l_{ij} in Matrix \mathbf{L} bezeichnet den Anteil vom Mendelian Sampling Effekt m_j , welcher im Zuchtwert u_i enthalten ist. Zum Beispiel entspricht das Element l_{53} dem Anteil vom Mendelian Sampling Effekt m_3 , welcher im Zuchtwert u_5 enthalten ist. Aufgrund der Zerlegung des Zuchtwertes u_5 in

$$u_5 = \frac{1}{2} u_3 + \frac{1}{2} u_4 + m_5$$

folgt, dass die Anteile von m_3 je zur Hälfte über den Vaterzuchtwert u_3 oder über den Mutterzuchtwert u_4 zustande kommen. Der Anteil von m_3 im Zuchtwert u_3 entspricht dem Element l_{33} in der Matrix \mathbf{L} und der Anteil von m_3 im Zuchtwert u_4 entspricht analog dem Element l_{43} . Somit lässt sich das Element l_{53} in unserem Beispiel berechnen als

$$l_{53} = \frac{1}{2} * l_{33} + \frac{1}{2} * l_{43} = \frac{1}{2} * 1 + \frac{1}{2} * 0 = \frac{1}{2}$$

3.5 Varianzen

Berechnen wir die Varianz des Vektors der Zuchtwerte \mathbf{u} aufgrund der Gleichung (30), so folgt daraus

$$\begin{aligned} Var(\mathbf{u}) &= Var(\mathbf{L} * \mathbf{m}) \\ &= \mathbf{L} * Var(\mathbf{m}) * \mathbf{L}^T \end{aligned} \quad (32)$$

Eine weiterführende Erklärung zur Bedeutung der Varianz von Vektoren ist im Anhang B aufgeführt. Die Varianz der Zuchtwerte \mathbf{u} entspricht der additiv genetischen Kovarianz-Matrix und diese ist gleich

$$Var(\mathbf{u}) = \mathbf{A} * \sigma_u^2 \quad (33)$$

wobei σ_u^2 der additiv genetischen Varianz entspricht.

Kombinieren wir die Gleichungen (33) und (32) so folgt daraus, dass

$$\begin{aligned} Var(\mathbf{u}) &= \mathbf{A} * \sigma_u^2 \\ &= \mathbf{L} * Var(\mathbf{m}) * \mathbf{L}^T \\ &= \mathbf{L} * \mathbf{D} * \mathbf{L}^T * \sigma_u^2 \end{aligned} \quad (34)$$

Aus Gleichung (34) folgt also, dass $Var(\mathbf{m}) = \mathbf{D} * \sigma_u^2$ ist. Da die Mendelian sampling Effekte unabhängig sind voneinander, d.h. $Cov(m_i, m_j) = 0$ für alle $i \neq j$, ist Matrix \mathbf{D} eine Diagonalmatrix. Die Diagonalelemente d_{ii} der Matrix \mathbf{D}

sind proportional zu den Varianzen der jeweiligen Mendelian sampling Effekte, wobei gilt, dass $d_{ii} = \text{Var}(m_i)/\sigma_u^2$

Die Varianz $\text{Var}(m_i)$ des Mendelian sampling Effektes m_i lässt sich aus der Zerlegung des Zuchtwertes u_i des Tieres i (siehe Gleichung (26)) berechnen.

$$u_i = \frac{1}{2} u_s + \frac{1}{2} u_d + m_i \quad (35)$$

Aufgrund der Zerlegung des Zuchtwertes u_i in Gleichung (35) können wir folgende Gleichung für die Varianzen aufstellen.

$$\begin{aligned} \text{Var}(u_i) &= \text{Var}\left(\frac{1}{2} u_s + \frac{1}{2} u_d + m_i\right) \\ &= \text{Var}\left(\frac{1}{2} u_s\right) + \text{Var}\left(\frac{1}{2} u_d\right) + \text{Var}(m_i) \\ &\quad + 2 * \text{Cov}\left(\frac{1}{2} u_s, \frac{1}{2} u_d\right) \\ &= \frac{1}{4} \text{Var}(u_s) + \frac{1}{4} \text{Var}(u_d) + \text{Var}(m_i) \\ &\quad + \frac{1}{2} \text{Cov}(u_s, u_d) \end{aligned} \quad (36)$$

In Gleichung (36) haben wir verwendet, dass die Mendelian Sampling Effekte unabhängig von den Zuchtwerten sind, d.h. dass $\text{Cov}(u_j, m_i) = 0$ für alle möglichen Werte für i und j .

Lösen wir Gleichung (36) nach $\text{Var}(m_i)$ auf und ersetzen wir die Varianzen der Zuchtwerte durch die jeweiligen Elemente aus der Verwandtschaftsmatrix mal die genetisch additive Varianz, erhalten wir folgendes Resultat für die Varianz $\text{Var}(m_i)$ der Mendelian sampling Effekte.

$$\begin{aligned} \text{Var}(m_i) &= \text{Var}(u_i) - \frac{1}{4} \text{Var}(u_s) - \frac{1}{4} \text{Var}(u_d) - \frac{1}{2} \text{Cov}(u_s, u_d) \\ &= (1 + F_i) \sigma_u^2 - \frac{1}{4} a_{ss} \sigma_u^2 - \frac{1}{4} a_{dd} \sigma_u^2 - \frac{1}{2} a_{sd} \sigma_u^2 \\ &= \left[1 + F_i - \frac{1}{4} a_{ss} - \frac{1}{4} a_{dd} - \frac{1}{2} a_{sd}\right] \sigma_u^2 \\ &= \left[1 + \frac{1}{2} a_{sd} - \frac{1}{4} a_{ss} - \frac{1}{4} a_{dd} - \frac{1}{2} a_{sd}\right] \sigma_u^2 \\ &= \left[1 - \frac{1}{4} a_{ss} - \frac{1}{4} a_{dd}\right] \sigma_u^2 \\ &= \left[1 - \frac{1}{4} (1 + F_s) - \frac{1}{4} (1 + F_d)\right] \sigma_u^2 \\ &= [0.5 - 0.25(F_s + F_d)] \sigma_u^2 \end{aligned} \quad (37)$$

Mit diesem Resultat für die Varianz $\text{Var}(m_i)$ lassen sich nun die Diagonalelemente d_{ii} der Matrix \mathbf{D} berechnen als

$$d_{ii} = \text{Var}(m_i)/\sigma_u^2 = 0.5 - 0.25(F_s + F_d) \quad (38)$$

3.6 Inverse Verwandtschaftsmatrix

Die Inverse \mathbf{A}^{-1} der Verwandtschaftsmatrix \mathbf{A} kann anhand der LDL-Zerlegung einfacher berechnet werden als mit konventionellen Invertierungsmethoden. Die LDL-Zerlegung für \mathbf{A} lautet

$$\mathbf{A} = \mathbf{L} * \mathbf{D} * \mathbf{L}^T \quad (39)$$

Die Inverse \mathbf{A}^{-1} kann somit berechnet werden als

$$\begin{aligned} \mathbf{A}^{-1} &= (\mathbf{L} * \mathbf{D} * \mathbf{L}^T)^{-1} \\ &= (\mathbf{L}^T)^{-1} * \mathbf{D}^{-1} * \mathbf{L}^{-1} \end{aligned} \quad (40)$$

3.6.1 Matrix \mathbf{D}^{-1}

Die Matrix \mathbf{D}^{-1} ist genau wie die Matrix \mathbf{D} wieder eine Diagonalmatrix, wobei die Diagonalelemente der Inversen \mathbf{D}^{-1} die Werte $1/d_{ii}$ haben. Die Berechnung der Diagonalelemente d_{ii} wurde in Gleichung (38) gezeigt.

3.6.2 Matrix \mathbf{L}^{-1}

Die Struktur der Matrix \mathbf{L}^{-1} kann aufgrund folgender Beziehungen erläutert werden. Zu Beginn dieses Abschnittes hatten wir in Gleichung (27) gesehen, wie der Zuchtwert eines Tieres in je die Hälfte der Zuchtwerte der Eltern plus den Mendelian sampling Effekt zerlegt werden kann. In Matrixschreibweise lautet diese Zerlegung

$$\mathbf{u} = \mathbf{P} * \mathbf{u} + \mathbf{m} \quad (41)$$

Daraus folgt, dass

$$\mathbf{m} = \mathbf{u} - \mathbf{P} * \mathbf{u} = (\mathbf{I} - \mathbf{P}) * \mathbf{u} \quad (42)$$

Aufgrund der rekursiven Zerlegung der Zuchtwerte \mathbf{u} in eine Linearkombination von Mendelian sampling Effekten

$$\mathbf{u} = \mathbf{L} * \mathbf{m} \quad (43)$$

folgt für \mathbf{m}

$$\mathbf{m} = \mathbf{L}^{-1} * \mathbf{u} \quad (44)$$

Durch Gleichsetzen beider Beziehungen für \mathbf{m} können wir \mathbf{L}^{-1} bestimmen.

$$\mathbf{m} = \mathbf{L}^{-1} * \mathbf{u} = (\mathbf{I} - \mathbf{P}) * \mathbf{u} \quad (45)$$

Es gilt also

$$\mathbf{L}^{-1} = \mathbf{I} - \mathbf{P} \quad (46)$$

wobei \mathbf{I} die Einheitsmatrix ist und die Matrix \mathbf{P} einer unteren Dreiecksmatrix entspricht, welche die Zuchtwerte der Tiere mit den Zuchtwerten der Eltern verknüpft (siehe Gleichungen (27) und (28)).

3.7 Berechnung der Inzuchtkoeffizienten

Wie in Abschnitt 3.6.1 und in Gleichung (38) gezeigt braucht es zur Berechnung der Matrix \mathbf{D}^{-1} die Inzuchtkoeffizienten. Die Inzuchtkoeffizienten sind auf der Diagonalen der Verwandtschaftsmatrix \mathbf{A} ersichtlich. Somit müsste zur direkten Invertierung der Verwandtschaftsmatrix \mathbf{A} die gesamte Matrix \mathbf{A} aufgestellt und gespeichert werden. Für grosse Datensätze kann diese Methode sehr ineffizient sein.

3.7.1 Cholesky Zerlegung von \mathbf{A}

Alternativ zum Aufstellen der gesamten Verwandtschaftsmatrix \mathbf{A} können die Inzuchtkoeffizienten aufgrund von einzelnen Zeilen der Matrix aus der Cholesky-Zerlegung von \mathbf{A} berechnet werden. Die Cholesky-Zerlegung von \mathbf{A} sieht wie folgt aus

$$\mathbf{A} = \mathbf{R} * \mathbf{R}^T \quad (47)$$

wobei \mathbf{R} eine linke untere Dreiecksmatrix ist. In der Vorlesung und in den Übungen wurde die Matrix \mathbf{R} als Matrix \mathbf{U} bezeichnet.

Aufgrund der Dreiecksstruktur von Matrix \mathbf{R} können die Diagonalelemente der Matrix \mathbf{A} berechnet werden als

$$a_{ii} = \sum_{n=1}^i r_{in}^2 \quad (48)$$

Schauen wir das bei einem kleinen Beispiel an.

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = \begin{bmatrix} r_{11} & 0 & 0 \\ r_{21} & r_{22} & 0 \\ r_{31} & r_{32} & r_{33} \end{bmatrix} * \begin{bmatrix} r_{11} & r_{21} & r_{31} \\ 0 & r_{22} & r_{32} \\ 0 & 0 & r_{33} \end{bmatrix}$$

Für das gezeigte kleine Beispiel berechnen wir die Diagonalelemente von \mathbf{A} als

$$\begin{aligned} a_{11} &= r_{11}^2 \\ a_{22} &= r_{21}^2 + r_{22}^2 \\ a_{33} &= r_{31}^2 + r_{32}^2 + r_{33}^2 \end{aligned}$$

Die Gleichung (48) zeigt den allgemeinen Fall der Berechnung der Diagonalelemente a_{ii} der Verwandtschaftsmatrix \mathbf{A} . Das bedeutet, dass die Diagonalelemente der Verwandtschaftsmatrix und somit die Inzuchtkoeffizienten aufgrund einer Zeile aus der Matrix \mathbf{R} berechnet werden können. Als nächste Frage stellt sich, wie die Elemente der Matrix \mathbf{R} rekursiv berechnet werden können, ohne dass wir die ganze Matrix \mathbf{A} oder die ganze Matrix \mathbf{R} aufstellen.

3.7.2 Rekursive Berechnung von \mathbf{R}

Der Schlüssel zur Berechnung der Elemente von \mathbf{R} liegt in der Gleichsetzung der Cholesky-Zerlegung der Verwandtschaftsmatrix \mathbf{A} und der LDL-Zerlegung von \mathbf{A} . Daraus lässt sich folgende Beziehung ableiten.

$$\mathbf{A} = \mathbf{R} * \mathbf{R}^T = \mathbf{L} * \mathbf{D} * \mathbf{L}^T \quad (49)$$

Aufgrund von Gleichung (49) wollen wir die Matrix \mathbf{R} in Abhängigkeit der Matrizen \mathbf{L} und \mathbf{D} ausdrücken, dies deshalb, weil wir die Elemente der Matrizen \mathbf{L} und \mathbf{D} einfach berechnen können.

Wir zerlegen die Matrix \mathbf{R} in das Produkt der zwei Matrizen \mathbf{L} und \mathbf{S} . Somit haben wir

$$\mathbf{R} = \mathbf{L} * \mathbf{S} \quad (50)$$

wobei \mathbf{S} eine Diagonalmatrix ist.

Betrachten wir die Zerlegung in Gleichung (50) anhand eines kleinen Beispiels, dann sieht das wie folgt aus

$$\begin{bmatrix} r_{11} & 0 & 0 \\ r_{21} & r_{22} & 0 \\ r_{31} & r_{32} & r_{33} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ l_{21} & 1 & 0 \\ l_{31} & l_{32} & 1 \end{bmatrix} * \begin{bmatrix} s_{11} & 0 & 0 \\ 0 & s_{22} & 0 \\ 0 & 0 & s_{33} \end{bmatrix} \quad (51)$$

Aufgrund der speziellen Struktur der Matrizen \mathbf{L} und \mathbf{S} ist ersichtlich, dass die Diagonalelemente r_{ii} der Matrix \mathbf{R} den Diagonalelementen s_{ii} der Matrix \mathbf{S} entsprechen oder anders formuliert, gilt für alle i , dass

$$r_{ii} = s_{ii} \quad (52)$$

Die Offdiagonalelemente r_{ij} der Matrix \mathbf{R} lassen sich als Produkt von l_{ij} und s_{jj} berechnen. Als Gleichung können wir schreiben, dass

$$r_{ij} = l_{ij} * s_{jj} \quad (53)$$

Im Abschnitt 3.4 haben wir gesehen, dass die Elemente l_{ij} als Durchschnitt von l_{sj} und l_{dj} berechnet werden können. Dabei wurde angenommen, dass s und d die Eltern von i sind. Verwenden wir die Berechnung von l_{ij} für die Berechnung von r_{ij} dann erhalten wir

$$\begin{aligned}
r_{ij} &= l_{ij} * s_{jj} \\
&= \frac{1}{2} (l_{sj} + l_{dj}) * s_{jj} \\
&= \frac{1}{2} (l_{sj} * s_{jj} + l_{dj} * s_{jj}) \\
&= \frac{1}{2} (r_{sj} + r_{dj})
\end{aligned} \tag{54}$$

Somit ist die rekursive Berechnung der Elemente in Matrix \mathbf{R} bestimmt. Wir müssen jetzt nur noch die Elemente der Matrix \mathbf{S} finden. Dazu setzen wir die Zerlegung der Matrix \mathbf{R} in die Cholesky-Zerklagung der Matrix \mathbf{A} ein und setzen diese der LDL-Zerlegung von Matrix \mathbf{A} gleich, dann folgt, dass

$$\begin{aligned}
\mathbf{A} = \mathbf{R} * \mathbf{R}^T &= \mathbf{L} * \mathbf{S} * (\mathbf{L} * \mathbf{S})^T \\
&= \mathbf{L} * \mathbf{S} * \mathbf{S}^T * \mathbf{L}^T \\
&= \mathbf{L} * \mathbf{D} * \mathbf{L}^T
\end{aligned} \tag{55}$$

Da wir auf der linken und auf der rechten Seite des letzten Gleichheitszeichens in (55) von links mit \mathbf{L} und von rechts mit \mathbf{L}^T multiplizieren, können wir diese beiden Matrix-Faktoren auch streichen und es gilt also, dass

$$\mathbf{D} = \mathbf{S} * \mathbf{S}^T$$

Zur Veranschaulichung betrachten wir auch dieses Resultat in unserem kleinen Beispiel.

$$\begin{bmatrix} d11 & 0 & 0 \\ 0 & d22 & 0 \\ 0 & 0 & d33 \end{bmatrix} = \begin{bmatrix} s11 & 0 & 0 \\ 0 & s22 & 0 \\ 0 & 0 & s33 \end{bmatrix} * \begin{bmatrix} s11 & 0 & 0 \\ 0 & s22 & 0 \\ 0 & 0 & s33 \end{bmatrix} \tag{56}$$

Aufgrund der Regeln der Matrixmultiplikation können wir sehen, dass für alle i gilt, dass $d_{ii} = s_{ii}^2$ oder $s_{ii} = \sqrt{d_{ii}}$. Die Elemente d_{ii} können wir aufgrund der Gleichung (38) berechnen als

$$\begin{aligned}
d_{ii} &= 0.5 - 0.25(F_s + F_d) \\
&= 1 - 0.25(a_{ss} + a_{dd})
\end{aligned} \tag{57}$$

Die Diagonalelemente a_{ss} und a_{dd} der Verwandtschaftsmatrix können wir aufgrund der Gleichung (48) als Summen von quadrierten Elementen der Matrix \mathbf{R} ausdrücken. Es gilt also

$$a_{ss} = \sum_{n=1}^s r_{sn}^2$$

$$a_{dd} = \sum_{n=1}^d r_{dn}^2$$

Die Elemente r_{sn} und r_{dn} werden aufgrund der Gleichungen (52) und (54) berechnet. Somit schliesst sich der Kreis und es konnte gezeigt werden, wie die Inzuchtkoeffizienten aufgrund einer Zeile der Matrix \mathbf{R} berechnet werden konnten.

Wichtig für diese Herleitung ist, dass die Tiere gemäss ihres Alters sortiert sind. Konkret heisst das, dass im Pedigree kein Nachkomme vor seinen Eltern auftreten darf. Die gleiche Sortierung gilt auch für die hier verwendeten Matrizen. Nur aufgrund dieser Sortierung ist die ganze Erklärung nicht zirkulär, d.h. die Katze beisst sich nicht in den Schwanz.

A Ableitungen nach Vektoren

Angenommen, wir haben eine Funktion $f(x_1, x_2, x_3)$, welche von drei Variablen x_1 , x_2 und x_3 abhängig ist. Fassen wir die Variablen x_1 , x_2 und x_3 zu einem Vektor \mathbf{x} zusammen, dann ist die Ableitung der Funktion f nach dem Vektor \mathbf{x} definiert als

$$\mathbf{v} = \frac{\partial f}{\partial \mathbf{x}} = \begin{bmatrix} \partial f / \partial x_1 \\ \partial f / \partial x_2 \\ \partial f / \partial x_3 \end{bmatrix} \quad (58)$$

Die Ableitung einer Funktion nach einem Vektor mit drei Komponenten ist wieder ein Vektor (\mathbf{v}) mit drei Komponenten, wobei jede Komponente von \mathbf{v} einer partiellen Ableitung der Funktion f nach einer Komponente von \mathbf{x} ist. Bei der partiellen Ableitung einer Funktion f nach einer Komponente x_1 wird die Funktion f nach x_1 abgeleitet und die anderen Variablen x_2 und x_3 werden als Konstante behandelt. Der Vektor \mathbf{v} wird auch als *Gradient* der Funktion f bezeichnet.

A.1 Beispiel

Wir nehmen an die Funktion f sei definiert als

$$f = 2 * x_1^2 * x_2 + x_2^3 - 5 * x_3 + 1$$

Der Gradient von f , also die Ableitung der Funktion f nach dem Vektor \mathbf{x} lautet dann

$$\begin{aligned} \mathbf{v} = \frac{\partial f}{\partial \mathbf{x}} &= \begin{bmatrix} \partial f / \partial x_1 \\ \partial f / \partial x_2 \\ \partial f / \partial x_3 \end{bmatrix} \\ &= \begin{bmatrix} 4 x_1 x_2 \\ 2 x_1^2 + 3 x_2^2 \\ -5 \end{bmatrix} \end{aligned} \quad (59)$$

B Varianz von Vektoren

Die Varianz eines Vektors \mathbf{x} dessen Komponenten aus Zufallsvariablen x_i bestehen, ist immer eine symmetrische Matrix - nennen wir sie \mathbf{V} . Die Diagonalelemente v_{ii} der Matrix \mathbf{V} entsprechen den Varianzen der einzelnen Komponenten des Vektors \mathbf{x} , wobei gilt, dass

$$v_{ii} = \text{Var}(x_i) = \sigma_i^2$$

Die Offdiagonalelemente v_{ij} der Matrix \mathbf{V} entsprechen den Kovarianzen zwischen den Elementen x_i und x_j . Es gilt also

$$v_{ij} = Cov(x_i, x_j) = \sigma_{ij}$$

Aufgrund der Symmetrie der Matrix \mathbf{V} gilt auch, dass $v_{ij} = v_{ji}$.

B.1 Beispiel

Betrachten wir als Beispiel einen Vektor \mathbf{x} der Länge drei. Unser Vektor \mathbf{x} sieht also wie folgt aus.

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

Die Covarianzmatrix \mathbf{V} des Vektors \mathbf{x} hat dann folgende Struktur

$$\mathbf{V} = Var(\mathbf{x}) = \begin{bmatrix} Var(x_1) & Cov(x_1, x_2) & Cov(x_1, x_3) \\ Cov(x_1, x_2) & Var(x_2) & Cov(x_2, x_3) \\ Cov(x_1, x_3) & Cov(x_2, x_3) & Var(x_3) \end{bmatrix} \quad (60)$$

Für Matrix \mathbf{V} gibt es auch folgende abgekürzte Schreibweise

$$\mathbf{V} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_2^2 & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_3^2 \end{bmatrix} \quad (61)$$

References

- [1] Heidi Signer-Hasler und Hannes Jörg (2015). Quantitative Genetik und Populationsgenetik. Skript Vorlesung 751-6200-00 V. FS2015.
- [2] Hofer, Andreas (1990). Schätzung von Zuchtwerten feldgeprüfter Schweine mit einem Mehrmerkmals-Tiermodell. Diss ETH 9160. <http://dx.doi.org/10.3929/ethz-a-000578289>
- [3] R.A. Mrode (2005). Linear Models For The Prediction Of Animal Breeding Values. Second Edition. CABI Publishing. ISBN 0 85199 000 2