

Züchtungslehre - Eigenschaften von BLUP-Zuchtwerten

Peter von Rohr

2016-11-18

Einleitung

In diesem Kapitel werden die Eigenschaften von BLUP-Zuchtwerten genauer erklärt. Im letzten Kapitel hatten wir gesehen, dass die geschätzten Zuchtwerte eine Funktion der Beobachtungen ist. Tiere ohne phänotypische Beobachtungen erhalten Zuchtwerte über Verknüpfungen in der inversen Verwandtschaftsmatrix. Welche Beobachtungen in einem geschätzten Zuchtwert eine Rolle spielen wollen wir bei der Zerlegung der geschätzten Zuchtwerte analysieren.

Vergleich verschiedener Zuchtwertschätzmethoden

Wir haben bisher eine Reihe von Zuchtwertschätzmethoden kennengelernt. Für den Vergleich sollen Zuchtwerte mit den zu vergleichenden Methoden aufgrund des gleichen Datensatzes geschätzt werden. Folgende Methoden werden miteinander verglichen.

- Zuchtwerte aufgrund von Eigenleistungen
- Zuchtwerte aufgrund von Nachkommenleistungen
- BLUP Zuchtwertschätzung mit Vatermodell
- BLUP Zuchtwertschätzung mit Tiermodell

Daten

Als Datensatz verwenden wir die Zunahmen bis zu Absetzen aus dem letzten Kapitel. Die folgende Tabelle gibt nochmals eine Übersicht über das Datenmaterial.

| Kalb | Geschlecht | Vater | Mutter | WWG |
|------|------------|-------|--------|-----|
| 4 | M | 1 | NA | 4.5 |
| 5 | F | 3 | 2 | 2.9 |
| 6 | F | 1 | 2 | 3.9 |
| 7 | M | 4 | 5 | 3.5 |
| 8 | M | 3 | 6 | 5.0 |

Die Varianzkomponenten σ_e^2 und σ_a^2 sind bekannt und haben die Werte $\sigma_e^2 = 40$ und $\sigma_a^2 = 20$.

Zuchtwerte aufgrund von Eigenleistungen

Diese Methode ist die einfachste um Zuchtwerte zu schätzen. Sie ist aber auch sehr limitiert, da nur Tiere mit einer eigenen beobachteten Leistung einen Zuchtwert bekommen. Für alle anderen Tiere kann kein Zuchtwert geschätzt werden. Für unser Beispiel bedeutet das, dass die Tiere in folgender Tabelle einen geschätzten Zuchtwert erhalten. Der Einfachheit halber nehmen wir den Mittelwert der Beobachtungen als Populationsmittel an.

| Tier | Zuchtwert |
|------|-----------|
| 4 | 0.180 |
| 5 | -0.353 |
| 6 | -0.020 |
| 7 | -0.153 |
| 8 | 0.347 |

Zuchtwerte aufgrund von Nachkommenleistungen

Zuchtwerte aufgrund von Nachkommenleistungen zu schätzen, macht eigentlich nur dann Sinn, wenn ein Tier mehr als ein Nachkommen hat. Dies trifft bei unserem Datensatz für die Tiere 1, 2 und 3 zu. Die geschätzten Zuchtwerte aufgrund von Nachkommendurchschnitten sind definiert als

$$\hat{a} = \frac{2n}{n+k}(\tilde{y} - \mu)$$

wobei n für die Anzahl Nachkommen steht und k als $(4 - h^2)/h^2$ definiert ist. Für unser Beispiel gilt also, dass $k = 11$. Der Durchschnitt der phänotypischen Leistungen der Nachkommen wird mit \tilde{y} bezeichnet. Als Populationsmittel μ verwenden wir wieder das Mittel aller phänotypischen Beobachtungen. Dieses beträgt 3.96. Die folgende Tabelle gibt eine Übersicht über die verwendeten Werte und die geschätzten Zuchtwerte (Kolonne BV).

| Tier | n | y | BV |
|------|---|------|--------|
| 1 | 2 | 4.20 | 0.074 |
| 2 | 2 | 3.40 | -0.172 |
| 3 | 2 | 3.95 | -0.003 |

wobei die Anzahl Nachkommen in Kolonne **n** und die Nachkommendurchschnitte in Kolonne **y** aufgelistet sind.

Zusammenfassend können wir feststellen, dass die geschätzten Zuchtwerte eine tiefere Variabilität aufweisen als die phänotypischen Beobachtungen und die Nachkommendurchschnitte. Vergleichen wir die empirischen Standardabweichungen, wie in folgender Tabelle gezeigt, wird diese Reduktion der Variabilität deutlich.

| Zufallsvariable | Standardabweichung |
|------------------------|--------------------|
| Beobachtungen | 0.823 |
| Nachkommendurchschnitt | 0.409 |
| Zuchtwerte EL | 0.274 |
| Zuchtwerte NL | 0.126 |

Es gilt aber zu beachten, dass diese Standardabweichungen auf sehr wenigen Beobachtungen basieren. Somit ist diese Beobachtung zunächst einmal einfach für diesen Datensatz gültig.

Das Vatermodell

Das Vatermodell gilt als eigentlicher Vorläufer des Tiermodells. Bei besonderen Datenkonstellationen wird dieses Modell aber heute noch in der Praxis - so zum Beispiel in der Fleischrinderzucht - eingesetzt. Beim Vatermodell handelt es sich auch wie beim Tiermodell um ein lineares gemischtes Modell. Die zufälligen Effekte des Vatermodells sind aber die sogenannten Vätereffekte und nicht wie im Tiermodell die Zuchtwerte. Aufgrund

dieser Tatsache bekommen nur männliche Tiere einen Zuchtwert und es werden auch nur die Verwandtschaften über die väterlichen Pfade berücksichtigt. Beim Vatermodell nehmen wir an, dass alle Paarungspartner genetisch ähnlich sind zueinander. Dies kann zu Verzerrungen bei den geschätzten Zuchtwerten führen, falls eine gezielte Paarung praktiziert wird. Der Vorteil des Vatermodells liegt in der reduzierten Anzahl an Gleichungen, welche zu lösen sind.

In Matrixnotation lautet das Vatermodell

$$y = Xb + Zs + e \tag{1}$$

Alle Terme in (1) sind gleich definiert, wie im Tiermodell, ausser dass s jetzt für den Vektor der zufälligen Vatereffekte steht und dass die Inzidenzmatrix Z die Beobachtungen mit den Vatereffekten verknüpft. Die Varianzkomponenten der zufälligen Effekte lauten

$$\text{var}(s) = A \sigma_s^2 \tag{2}$$

wobei A die Verwandtschaftsmatrix zwischen den Vätern ist und $\sigma_s^2 = 0.25\sigma_a^2$ beträgt. Die Mischmodellgleichungen sind gleich, wie beim Tiermodell, ausser, dass

$$\alpha = \frac{\sigma_e^2}{\sigma_s^2} = \frac{4 - h^2}{h^2}$$

Ein Beispiel

Zur Erklärung des Vatermodells verwenden wir den gleichen Datensatz wie für das Tiermodell. Bei den Verwandtschaftsbeziehungen werden die Mütter ignoriert. Der für das Vatermodell modifizierte Datensatz sieht dann wie folgt aus.

Das Ziel hier ist den fixen Effekt des Geschlechts zu schätzen und Zuchtwerte für die Väter 1, 3 und 4 vorauszusagen. Die Parameter seien die gleichen wie im Tiermodell, somit ist $\sigma_s^2 = \sigma_a^2/4 = 5$ und $\sigma_e^2 = 55$ daraus folgt $\alpha = 11$.

| Kalb | Geschlecht | Vater | Mutter | WWG |
|------|------------|-------|--------|-----|
| 4 | M | 1 | NA | 4.5 |
| 5 | F | 3 | NA | 2.9 |
| 6 | F | 1 | NA | 3.9 |
| 7 | M | 4 | NA | 3.5 |
| 8 | M | 3 | NA | 5.0 |

Die Inzidenzmatrix X ist gleich wie beim Tiermodell. Da Z die Beobachtungen zu den Vatereffekten zuordnet, sieht Z nun anders aus als im Tiermodell. Die Inzidenzmatrix Z für die Vatereffekte ist

$$Z = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

Die Komponenten der Mischmodellgleichungen sehen wie folgt aus

$$X^T X = \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix}$$

$$X^T Z = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \end{bmatrix}$$

$$Z^T Z = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Beim Vatermodell werden nur Verwandtschaftsbeziehungen zwischen den Vätern berücksichtigt. Abgesehen davon, dass Vater 4 der Sohn von Vater 1 ist, gibt es keine Abhängigkeiten. Die Inverse A^{-1} sieht somit folgendermassen aus.

$$A^{-1} = \begin{bmatrix} 1.333 & 0.000 & -0.667 \\ 0.000 & 1.000 & 0.000 \\ -0.667 & 0.000 & 1.333 \end{bmatrix}$$

Bei der rechten Handseite sieht $X^T y$ gleich aus wie beim Tiermodell und $Z^T y$ berechnen wir aus dem Produkt von Z^T und y .

$$Z^T y = \begin{bmatrix} 8.4 \\ 7.9 \\ 3.5 \end{bmatrix}$$

Nun haben wir alle Komponenten und können die Mischmodellgleichungen für das Vatermodell aufstellen.

$$\begin{bmatrix} 3.000 & 0.000 & 1.000 & 1.000 & 1.000 \\ 0.000 & 2.000 & 1.000 & 1.000 & 0.000 \\ 1.000 & 1.000 & 16.663 & 0.000 & -7.337 \\ 1.000 & 1.000 & 0.000 & 13.000 & 0.000 \\ 1.000 & 0.000 & -7.337 & 0.000 & 15.663 \end{bmatrix} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{s}_1 \\ \hat{s}_2 \\ \hat{s}_3 \end{bmatrix} = \begin{bmatrix} 13.0 \\ 6.8 \\ 8.4 \\ 7.9 \\ 3.5 \end{bmatrix}$$

Aus den oben gezeigten Mischmodellgleichungen lassen sich die Lösungen für die fixen Effekte und die zufälligen Vätereffekte, welche als vorausgesagte Zuchtwerte der Väter gelten, berechnen.

$$\begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{s}_1 \\ \hat{s}_2 \\ \hat{s}_3 \end{bmatrix} = \begin{bmatrix} 4.336 \\ 3.382 \\ 0.022 \\ 0.014 \\ -0.043 \end{bmatrix}$$

Vergleich zwischen Tiermodell und Vatermodell

Wir haben fixe Effekte geschätzt und Zuchtwerte vorausgesagt einmal mit dem Tiermodell (siehe letztes Kapitel) und soeben mit dem Vatermodell. Wir wollen nun die Resultate der beiden Modelle miteinander vergleichen.

Fixe Effekte

In beiden Modellen waren das Geschlecht als fixer Effekt berücksichtigt. Die folgende Tabelle zeigt die geschätzten Effekte für die zwei Ausprägungen des Geschlechts

| Effekt | Tiermodell | Vatermodell |
|--------|------------|-------------|
| M | 4.36 | 4.34 |
| F | 3.40 | 3.38 |

Bis auf Rundungsfehler sind die Differenzen zwischen den Effektstufen “M” und “F” bei beiden Modellen gleich. Diese Übereinstimmung ist erklärbar, da in beiden Modellen die gleichen fixen Effekte berücksichtigt wurden und die Effekte aufgrund des gleichen Datensatzes geschätzt wurden.

Zuchtwerte

Aufgrund der Modelleigenschaften werden beim Tiermodell Zuchtwerte für alle Tiere vorausgesagt und beim Vatermodell nur für die Väter. Ein Vergleich der Zuchtwerte ist in der folgenden Tabelle gezeigt.

| Tier | Tiermodell | Vatermodell |
|------|------------|-------------|
| 1 | 0.098 | 0.022 |
| 2 | -0.019 | NA |
| 3 | -0.041 | 0.014 |
| 4 | -0.009 | -0.043 |
| 5 | -0.186 | NA |
| 6 | 0.177 | NA |
| 7 | -0.249 | NA |
| 8 | 0.183 | NA |

Die Zuchtwerte und auch paarweise Differenzen zwischen den Zuchtwerten sind verschieden zwischen den beiden Modellen. Auch die Rangierung der Väter ist nicht die gleiche bei den beiden Modellen. Bei Tiermodell werden auch die Beiträge der Paarungspartner und aller Nachkommen berücksichtigt. Beim Vatermodell werden die Paarungspartner alle als identisch betrachtet. Nachkommen werden nur auf der väterlichen Seite berücksichtigt. Diese Unterschiede wirken sich auf die vorausgesagten Zuchtwerte aus.

Zerlegung eines mit dem Tiermodell geschätzten Zuchtwertes

Extrahiert man einzelne Gleichungen aus den Mischmodellgleichungen, so kann gezeigt werden, aus welchen Komponenten sich die vorhergesagten Zuchtwerte zusammensetzen. Diese Technik der Zerlegung erlaubt es weitere Eigenschaften des Tiermodells aufzuzeigen.

Wir gehen von folgendem vereinfachten Modell aus

$$y_i = \mu + a_i + e_i \quad (3)$$

- mit y_i Beobachtung für Tier i
 a_i Zuchtwert von Tier i mit Varianz $(1 + F_i)\sigma_a^2$
 e_i zufälliger Rest mit Varianz σ_e^2
 μ übrige fixe Effekte im Modell

In einem fiktiven angenommenen Datensatz haben alle Tiere nur eine Beobachtung. Tier i hat Eltern s und d und n Nachkommen k_j (wobei $j = 1, \dots, n$) und n Paarungspartner l_j (wobei $j = 1, \dots, n$). Also hat der Nachkomme k_j die Eltern i und l_j . Die Mischmodellgleichungen für das vereinfachte Modell (3) lauten, wie im nachfolgenden Abschnitt gezeigt. Da die fixen Effekte alle in einem gemeinsamen Mittel μ zusammengefasst werden ist die Inzidenzmatrix X einfach ein Vektor mit lauter Einsen. Da jedes Tier eine Beobachtung hat, entspricht die Matrix Z der Einheitsmatrix.

Unter Berücksichtigung der Regeln zur Aufstellung der Inversen Verwandtschaftsmatrix A^{-1} lässt sich die Gleichung für die Beobachtung y_i wie folgt darstellen.

$$y_i = \hat{\mu} + \left[1 + \alpha d^{(i)} + \frac{\alpha}{4} \sum_{j=1}^n d^{(k_j)} \right] \hat{a}_i - \frac{\alpha}{2} d^{(i)} \hat{a}_s - \frac{\alpha}{2} d^{(i)} \hat{a}_d - \frac{\alpha}{2} \sum_{j=1}^n d^{(k_j)} \hat{a}_{k_j} + \frac{\alpha}{4} \sum_{j=1}^n d^{(k_j)} \hat{a}_{l_j} \quad (4)$$

Lösen wir die Gleichung (4) nach \hat{a}_i auf so folgt

$$\hat{a}_i = \frac{1}{1 + \alpha d^{(i)} + \frac{\alpha}{4} \sum_{j=1}^n d^{(k_j)}} \left[y_i - \hat{\mu} + \frac{\alpha}{2} \left\{ d^{(i)} (\hat{a}_s + \hat{a}_d) + \sum_{j=1}^n d^{(k_j)} (\hat{a}_{k_j} - \frac{1}{2} \hat{a}_{l_j}) \right\} \right] \quad (5)$$

Aus dieser Zerlegung ist ersichtlich, dass sich der geschätzte Zuchtwert im Tiermodell aus den folgenden Komponenten zusammensetzt:

- Elternzuchtwerte: \hat{a}_s und \hat{a}_d
- Eigenleistung: y_i
- Zuchtwerte von Nachkommen und Paarungspartnern: \hat{a}_{k_j} und \hat{a}_{l_j}

Ein Beispiel

Zur Illustration nehmen wir ein Beispiel mit $n = 5$ Tieren an.

$$X = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

$$Z = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\begin{aligned}
X^T X &= [5] \\
X^T Z &= [1 \ 1 \ 1 \ 1 \ 1] \\
Z^T Z &= \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}
\end{aligned}$$

$$X^T y = \left[\sum_{j=1}^n y_j \right]$$

$$Z^T y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{bmatrix}$$

Beim Pedigree nehmen wir den einfachsten Fall mit 5 Tieren.

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

Aus dem oben gezeigten Pedigree berechnen wir die folgende Inverse A^{-1} .

$$A^{-1} = \begin{bmatrix} 1.50 & 0.50 & 0.00 & -1.00 & 0.00 \\ 0.50 & 1.50 & 0.00 & -1.00 & 0.00 \\ 0.00 & 0.00 & 1.50 & 0.50 & -1.00 \\ -1.00 & -1.00 & 0.50 & 2.50 & -1.00 \\ 0.00 & 0.00 & -1.00 & -1.00 & 2.00 \end{bmatrix}$$

Wir setzen nun die Komponenten zusammen zu

$$\begin{bmatrix} X^T X & X^T Z \\ Z^T X & Z^T Z + A^{-1} * \alpha \end{bmatrix} \begin{bmatrix} \hat{\mu} \\ \hat{a} \end{bmatrix} = \begin{bmatrix} X^T y \\ Z^T y \end{bmatrix} \quad (6)$$

Setzen wir als Beobachtungen y die Werte von unserem Datensatz ein und verwenden auch den gleichen Wert für $\alpha = 2$, dann sehen die Mischmodellgleichungen in Zahlen wie folgt aus.

$$\begin{bmatrix} 5.00 & 1.00 & 1.00 & 1.00 & 1.00 & 1.00 \\ 1.00 & 4.00 & 1.00 & 0.00 & -2.00 & 0.00 \\ 1.00 & 1.00 & 4.00 & 0.00 & -2.00 & 0.00 \\ 1.00 & 0.00 & 0.00 & 4.00 & 1.00 & -2.00 \\ 1.00 & -2.00 & -2.00 & 1.00 & 6.00 & -2.00 \\ 1.00 & 0.00 & 0.00 & -2.00 & -2.00 & 5.00 \end{bmatrix} \begin{bmatrix} \mu \\ \hat{a}_1 \\ \hat{a}_2 \\ \hat{a}_3 \\ \hat{a}_4 \\ \hat{a}_5 \end{bmatrix} = \begin{bmatrix} 19.8 \\ 4.5 \\ 2.9 \\ 3.9 \\ 3.5 \\ 5.0 \end{bmatrix}$$

Für unser Beispiel betrachten wir nun das Tier 4 mit Eltern 1 und 2. Das Tier 4 hat mit Paarungspartner 3 einen Nachkommen (5).

Aus den Mischmodellgleichungen betrachten wir die zweitletzte Gleichung und können so die Formel (5) verifizieren. Dabei gilt zu berücksichtigen, dass $d^{(i)}$ nur drei Werte annehmen kann.

$$d^{(i)} = \begin{cases} 2 & \text{falls beide Eltern bekannt} \\ \frac{4}{3} & \text{falls ein Elternteil bekannt} \\ 1 & \text{falls beide Eltern unbekannt} \end{cases}$$