

Genomic Best Linear Unbiased Prediction (GBLUP)

Peter von Rohr

12 März 2018

Einführung

- ▶ Schätzung der genomischen Zuchtwerte
 - ▶ Anzahl Parameter (p) grösser als Anzahl Beobachtungen (n), d.h. $n \ll p$
 - ▶ Parameter entsprechen genomischen Zuchtwerten
 - ▶ In gewissen Populationen (USA, CA) stimmt $n < p$ nicht mehr, aber trotzdem können wir Least Squares nicht verwenden, da Kolonnen der Design-Matrix X linear voneinander abhängig
- ▶ Traditionelle Zuchtwertschätzung
 - ▶ Analoge Situation
 - ▶ Anzahl geschätzte Zuchtwerte grösser als Anzahl Beobachtungen
 - ▶ BLUP-basiertes Verfahren für Schätzung
 - ▶ Lineare gemischte Modelle
 - ▶ Mischmodellgleichungen

Traditionelle Zuchtwertschätzung

- ▶ Daten
 - ▶ beobachtete Merkmale und Umweltfaktoren
 - ▶ Pedigree enthält Verwandtschaftsbeziehungen
- ▶ Resultat: alle Tiere im Pedigree erhalten einen Zuchtwert
- ▶ Modell: Lineares gemischtes Modell

$$y = Xb + Za + e$$

- ▶ Parameterschätzung mit Mischmodellgleichungen

$$\begin{bmatrix} X^T X & X^T Z \\ Z^T X & Z^T Z + \lambda A^{-1} \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{a} \end{bmatrix} = \begin{bmatrix} X^T y \\ Z^T y \end{bmatrix}$$

]

Genomische Zuchtwertschätzung

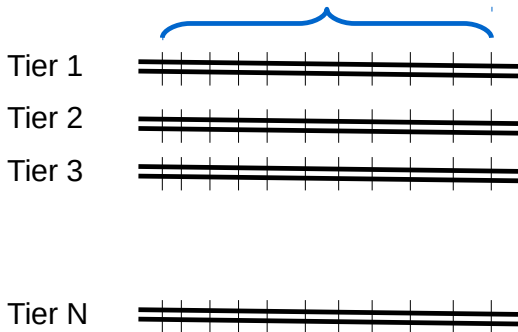
- ▶ Daten
 - ▶ Merkmale und Umweltfaktoren
 - ▶ SNP-Genotypen für typisierte Tiere
- ▶ Kovarianz der zufälligen Effekte
 - ▶ Genomische Verwandtschaftsmatrix G anstatt A
 - ▶ Tiere ohne Genotypen über A berücksichtigt
- ▶ Zwei Ansätze
 - ▶ Ridge Regression (RR - BLUP)
 - ▶ Genomic BLUP (GBLUP)

Ridge Regression BLUP

- ▶ Jeder SNP-Genotyp wird als zufälliger Effekt modelliert
- ▶ Resultierendes Gemischtes lineares Modell

$$y = \mathbf{1}_n\mu + Wq + e$$

SNP1 – SNPq
als Vektor q



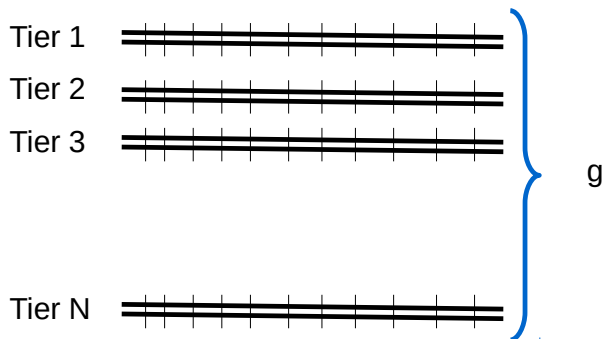
GBLUP

- ▶ Alle SNP-Effekte pro Tier werden als einen zufälligen Effekt g modelliert
- ▶ Analog zum traditionellen BLUP-Tiermodell
- ▶ Kovarianz der zufälligen Effekte

$$\text{var}(g) = G\sigma_g^2$$

wobei G der genomischen Verwandtschaftsmatrix entspricht

GBLUP Graphik



Herleitung von G

- ▶ Aufgrund der Modelle in RR-BLUP und GBLUP gilt, dass

$$g = U * q$$

wobei $U = W - P$ und P von der Codierung der SNP-Genotypen in W abhängt. P wird so gewählt, dass $E[g] = 0$

- ▶ Es soll gelten, dass

$$\text{var}(g) = G * \sigma_g^2$$

- ▶ Somit ist

$$\text{var}(g) = G * \sigma_g^2 = U * \text{var}(q) * U^T = UU^T \sigma_q^2$$

wobei: $\text{var}(q) = I * \sigma_q^2$

Herleitung von G (II)

- ▶ Aus (Gianola et al 2009):

$$\sigma_g^2 = 2 * \sum_{j=1}^m p_j(1 - p_j)\sigma_q^2$$

- ▶ Somit gilt

$$\text{var}(g) = G * 2 * \sum_{j=1}^m p_j(1 - p_j)\sigma_q^2 = UU^T \sigma_q^2$$

- ▶ Und G

$$G = \frac{UU^T}{2 * \sum_{j=1}^m p_j(1 - p_j)}$$