

NORBERT MIELENZ und LUTZ SCHÜLER

Stichprobenumfang zur Schätzung der Dominanzvarianz bei hierarchischer Halb- und Vollgeschwisterstruktur

Abstract

Title of the paper: **Sample size for estimation of dominance variance with hierarchical full- and half-sib structure**

The sampling size for the estimation of the ratio d^2 of the dominance variance to the phenotypic variance is investigated for the case of hierarchical population structures corresponding to the relations in poultry and pig breeding. The standard error of the d^2 -estimation is calculated by using the Maximum-Likelihood method and the concept of parameter transformation. Estimation of the dominance variance requires much larger data sets than estimation of the additive variance. Assuming d^2 -values of 25, 17, 10 and 5% than the estimation of the ratio d^2 required about 6, 10, 20 and 40 times more data than the estimation of the additive variance. For the case $h^2:d^2=0.4:0.1$ the estimation of d^2 with 8320, 14720, 33040 and 58640 records yielded standard errors of 8, 6, 4 and 3%. In the situation of $h^2:d^2=0.225:0.075$ the same accuracies can be reached with records of 4400, 7760, 17520 and 31040 animals.

Key Words: sample size, dominance variance, full- and half-sib families

Zusammenfassung

In dieser Arbeit wird der Stichprobenumfang für die Schätzung des Anteils d^2 der Dominanzvarianz an der phänotypischen Varianz für den Fall hierarchischer Populationsstrukturen untersucht, wie sie bei den Tierarten Huhn und Schwein vorliegen. Zur Ableitung des Standardfehlers für die d^2 -Schätzung werden die Maximum-Likelihood-Methode und das Konzept der Parametertransformation verwendet. Bei vergleichbarer Genauigkeit erfordert die Schätzung der Dominanzvarianz wesentlich größere Datensätze als die Schätzung der additiv-genetischen Varianz. Für hierarchische Halb- und Vollgeschwisterstrukturen steigt für d^2 -Werte von 25, 17, 10 und 5% der Stichprobenumfang beim Übergang vom rein-additiven zum Dominanzmodell um mindestens das 6-, 10-, 20- und 40-fache. Für die Konstellation $h^2:d^2=0,4:0,1$ werden für die Schätzung von d^2 Standardfehler von 8, 6, 4 und 3% realisiert, falls 8320, 14720, 33040 bzw. 58640 Datensätze vorhanden sind. Liegt der Fall $h^2:d^2=0,225:0,075$ vor, so reduziert sich der Stichprobenumfang auf 4400, 7760, 17520 bzw. 31040 Datensätze.

Schlüsselwörter: Stichprobenumfang, Dominanzvarianz, Voll- und Halbgeschwisterfamilien

1. Einleitung

Die Zuchtwertschätzung bei Nutztieren beruht gewöhnlich auf additiven Modellen (SCHÜLER u.a., 2001) wobei vorausgesetzt wird, dass mögliche Interaktionen zwischen verschiedenen Allelen an einem Genort (Dominanz) und Interaktionen zwischen den Allelen verschiedener Loci (Epistasie) ignoriert werden können. Aufgrund der ständigen Weiterentwicklung der statistischen Methoden und Computeralgorithmen insbesondere bei der Inversion der Dominanz-Verwandtschaftsmatrix wurde die genetische Berechnung von nichtadditiven Effekten von großen Datensätzen möglich (HOESCHELE und VAN RADEN, 1991; MISZTAL u.a. 1995). Dabei erfordert die Schätzung der Dominanzvarianz mit Tiermodellen neben einem großen Stichprobenumfang (VAN RADEN u.a., 1992) eine Schätzmethode wie REML (PATTERSON

und THOMPSON, 1971), welche invariant gegenüber Verzerrungen hervorgerufen durch Selektion ist.

Erste Untersuchungen zur Schätzung der Dominanzvarianz mit der REML-Methode bei Unterstellung eines Tiermodells gehen auf WEI und VAN der WERF (1993) zurück. In den letzten Jahren wurden zunehmend additive Tiermodelle, erweitert um Dominanzeffekte, auch für praxisrelevante Tierumfänge bei den Tierarten Schwein und Rind (MISZTAL, 1997, MISZTAL u.a., 1998) angewendet.

Von CHANG (1988) wurde die Anzahl der Familien berechnet, so dass die im Modell enthaltenen genetischen Varianzkomponenten mit Konfidenzniveau $1-\alpha=0,95$ statistisch von Null verschieden geschätzt werden. Die mit Hilfe von stochastischer Simulation erzielten Resultate zeigen, dass für eine Heritabilität im weiteren Sinn (h_w^2) von 0,5 die Schätzung der Dominanzvarianz rund 20 mal mehr und die Schätzung der additiv \times additiv-genetischen Varianz 380 mal mehr Datensätze erfordert als die Schätzung der additiv-genetischen Varianz bei äquivalenter Genauigkeit. Mit Hilfe des Konzeptes von unterschiedlichen Familientypen untersuchten VAN RADEN u.a. (1992) die Genauigkeit der Schätzung von nichtadditiv genetischer Varianz an der U.S. Holstein-Population. Unter Verwendung aller Datensätze von 6 verschiedenen Familientypen (Klone, Vollgeschwister, Elter-Nachkomme, $\frac{3}{4}$ -Geschwister, Halb- und Viertelgeschwister) dieser weltweit größten Nutztierpopulation fanden die Autoren, dass die Dominanzvarianz theoretisch mit einem Standardfehler von 1,4% an der phänotypischen Varianz geschätzt werden kann. Die Ergebnisse zur Versuchsplanung bei VAN RADEN u.a. (1992) beziehen sich hauptsächlich auf die Verhältnisse der Rinderzucht. Es wird vorausgesetzt, dass Tiere aus verschiedenen Familien unverwandt sind und dass alle Tiere innerhalb einer Familie im gleichen Verwandtschaftsverhältnis stehen. In dieser Arbeit wird der Stichprobenumfang für die Schätzung des Anteils d^2 der Dominanzvarianz an der phänotypischen Varianz für den Fall multipaarere Tiere abgeleitet. Das heißt, es werden hierarchische Populationsstrukturen untersucht, wie sie bei den Tierarten Huhn und Schwein vorliegen.

2. Methode

Sei $y_i = (y_{i1}, \dots, y_{i,m_i})'$ mit $i=1, \dots, f$ der Vektor der Leistungen aller Tiere aus Familie i , wobei die phänotypischen Beobachtungen einem Tiermodell erweitert um Dominanz genügen sollen. Vorerst sei lediglich vorausgesetzt, dass sich die Tiere einer Familie durch einheitliche verwandtschaftliche Strukturen auszeichnen. Setzt man Abwesenheit von Inzucht, keine Kovarianz zwischen den additiv-genetischen, den dominanzbedingten und den umweltbedingten Effekten voraus, so besitzen die VC-Matrizen von y_i und $y = (y_1, \dots, y_f)'$ die Darstellungen:

$$(1) \quad \text{Var}(y_i) = V_i = V_{i1} \cdot \sigma_a^2 + V_{i2} \cdot \sigma_d^2 + V_{i3} \cdot \sigma_e^2$$

$$(2) \quad \text{Var}(y) = V = V_1 \cdot \sigma_a^2 + V_2 \cdot \sigma_d^2 + V_3 \cdot \sigma_e^2$$

$$\text{mit } V_{i3} = I_{m_i} \text{ und } V_3 = I_N \quad (N = \sum_i m_i)$$

Hierbei sind σ_a^2 , σ_d^2 und σ_e^2 die additiv-genetische, die dominanzbedingte und die umweltbedingte Varianz. In Beziehung (1) entsprechen die Matrizen V_{i1} bzw. V_{i2} der additiven bzw. der Dominanz-Verwandtschaftmatrix aller Tiere aus Familie i .

Schätzt man die Varianzkomponenten des Vektors $\theta = (\sigma_a^2, \sigma_d^2, \sigma_e^2)'$ mit der Maximum-Likelihood-Methode, so besitzt die VC-Matrix der Schätzungen die Darstellung (SEARLE, 1979):

$$(3) \quad \text{Var}(\hat{\theta}) = 2 \cdot S^{-1}(\theta) \quad \text{mit} \quad S = \{s_{jk}\} = \{Sp(V_j V^{-1} V_k V^{-1})\}; \quad j, k = 1, \dots, 3$$

In Darstellung (3) erhält man die Schätzfehler der REML-Methode, falls die Matrix V durch die Projektionsmatrix $P = V^{-1} - V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}$ ersetzt wird (vgl. PATTERSON and THOMPSON, 1971). Da P jedoch von der Versuchsplanmatrix X der festen Modelleffekte abhängt, wurde im weiterem für die Stichprobenplanung die Beziehung (3) benutzt.

Sei zusätzlich vorausgesetzt, dass Tiere aus verschiedenen Familien unverwandt sind, so lässt sich die numerische Berechnung der Elemente von Matrix S wie folgt vereinfachen:

$$(4) \quad s_{jk} = \sum_{i=1}^f Sp(V_{ij} V_i^{-1} V_{ik} V_i^{-1})$$

Zur Berechnung der Matrizen V_{i1} und V_{i2} in (1) müssen zusätzliche Annahmen an die Verwandtschaftsstruktur innerhalb der Familien gestellt werden. Angenommen es liegen Halbgeschwistergruppen mit b_i - Müttern pro Vater und n_i - Nachkommen pro Mutter vor. Bezeichnen wir mit a_{jk} (bzw. d_{jk}) den additiven (bzw. dominanzbedingten) Effekt eines Tieres aus Halbgeschwistergruppe i . Dann lassen sich die Elemente von V_{i1} und V_{i2} gemäß der genetischen Kovarianz zwischen Voll- und Halbgeschwistern wie folgt berechnen.

$$(5) \quad V_{i1} = \{ \text{cov}(a_{jk}, a_{j'k'}) \}; \quad \text{cov}(a_{jk}, a_{j'k'}) = \begin{cases} 1 & \text{für } j'=j; k'=k \\ 0.5 & \text{für } j'=j; k' \neq k \\ 0.25 & \text{sonst} \end{cases}$$

$$V_{i2} = \{ \text{cov}(d_{jk}, d_{j'k'}) \}; \quad \text{cov}(d_{jk}, d_{j'k'}) = \begin{cases} 1 & \text{für } j'=j; k'=k \\ 0.25 & \text{für } j'=j; k' \neq k \\ 0 & \text{sonst} \end{cases}$$

$$\text{mit } i = 1, \dots, f; \quad j, j' = 1, \dots, b_i \quad \text{und} \quad k, k' = 1, \dots, n_{ij}$$

Bemerkung 1: Die analytische Berechnung der Inversen von Matrix V_i (bei Beachtung von (1) und (5)) gelingt, falls alle Tiere innerhalb einer Familie im gleichen Verwandtschaftsverhältnis stehen, also z.B. nur Voll- oder nur Halbgeschwister sind (vgl. VAN RADEN u.a., 1992). In diesem Spezialfall (mit f - Familien vom Umfang m_i) besitzt V_i die Darstellung:

$$(6) \quad V_i = I_{m_i} \cdot \sigma_w^2 + e_{m_i} e_{m_i}' \cdot \sigma_f^2; \quad \sigma_f^2 = \text{cov}(y_{ij}, y_{ij'}) \quad \text{für } j \neq j'; \quad \sigma_w^2 = \text{Var}(y_{ij}) - \sigma_f^2$$

Hierbei sind σ_w^2 die Varianz innerhalb Familie i und σ_f^2 die Varianz zwischen den Familienmitgliedern.

Bemerkung 2: Bei VAN RADEN u.a. (1992) wurde gezeigt, dass unter den Annahmen von Bemerkung 1 und der zusätzlichen Bedingung $m_i = m$ für alle i gilt:

$$(7) \quad \text{Var}(\hat{\sigma}_f^2) = \frac{2 \cdot [m \cdot \sigma_f^4 + \sigma_w^4 / (m-1) + 2 \cdot \sigma_f^2 \cdot \sigma_w^2]}{m \cdot f}$$

Unter der Voraussetzung $m \cdot f = \text{const}$ wird (7) minimiert für: $m = (\sigma_w^2 / \sigma_f^2) + 1$; Somit folgt z.B. für Vollgeschwisterfamilien $\sigma_f^2 = 0,5 \cdot \sigma_a^2 + 0,25 \cdot \sigma_d^2$, womit sich durch Einsetzen in die Formel für m als optimale Anzahl Tiere pro Vollgeschwisterfamilie ergibt: $m = 4 / (2 \cdot h^2 + d^2)$

Bemerkung 3: Seien neben f_1 - Vollgeschwisterfamilien vom Umfang m_1 zusätzlich f_2 - Halbgeschwisterfamilien vom Umfang m_2 vorhanden, wobei Tiere aus verschiedenen Familien als unverwandt vorausgesetzt werden. Dann kann gemäß der Vorgehensweise von VAN RADEN u. a. (1992) die Dominanzvarianz z.B. wie folgt geschätzt werden:

$$(8) \quad \hat{\sigma}_d^2 = 4 \cdot (\hat{\sigma}_{f_1}^2 - 2 \cdot \hat{\sigma}_{f_2}^2)$$

Bei gegebenen Werten von $N = m_1 \cdot f_1 + m_2 \cdot f_2$, m_1 und m_2 wird der Schätzfehler von (8) minimiert für einen Anteil $\beta = (m_1 \cdot f_1) / N$ von Vollgeschwisterfamilien berechnet gemäß der folgenden Formel:

$$(9) \quad \beta = 1 / (1 + 2 \cdot \sqrt{B_1 / B_2}) \quad \text{mit} \quad B_i = m_i \cdot \sigma_{f_i}^4 + \sigma_{w_i}^4 / (m_i - 1) + 2 \cdot \sigma_{f_i}^2 \cdot \sigma_{w_i}^2 ; i = 1, 2$$

Hierbei sind $\sigma_{f_i}^2$ (bzw. $\sigma_{w_i}^2$) die Varianzen zugehörig zu den 2 Familientypen, d.h. es gilt: $\sigma_{f_1}^2 = 0,5 \cdot \sigma_a^2 + 0,25 \cdot \sigma_d^2$ (bzw. $\sigma_{f_2}^2 = 0,25 \cdot \sigma_a^2$)

Die Berechnung der Varianz-Kovarianzmatrix der zu schätzenden Varianzkomponenten erfolgte gemäß der Formeln (3) bis (5) bei gegebener Anzahl Väter, Mütter pro Vater und Nachkommen pro Mutter mit Hilfe der Matrixprogrammiersprache IML innerhalb des Statistikpaketes SAS. Der Schätzfehler für die Anteile der additiven und der dominanzbedingten Varianz wurden mit Hilfe der Parametertransformation $\gamma_1 = \sigma_a^2 + \sigma_d^2 + \sigma_e^2$; $\gamma_2 = \sigma_a^2 / \gamma_1$ und $\gamma_3 = \sigma_d^2 / \gamma_1$ bestimmt (vgl. LÄUTER und PINCUS, 1989; MIELENZ u.a., 2003).

3. Ergebnisse

Die nachfolgenden Ergebnisse beruhen auf der Annahme von Normalverteilung für die geschätzten Anteile der additiv-genetischen (h^2) und der Dominanzvarianz (d^2) an der Gesamtvarianz, wobei der Standardfehler der Schätzungen unter Verwendung des Formelapparates aus Abschnitt 2 berechnet wurde. In Tabelle 1 ist die minimale Anzahl Väter bei gegebener Anzahl Mütter pro Vater und Nachkommen pro Mutter aufgelistet, so dass die untere Grenze des zugehörigen 95% Konfidenzintervalls für h^2 bzw. d^2 positiv ausfällt. Für die Berechnung des Stichprobenumfangs wurde also die

minimale Vaterzahl ermittelt, so dass der Ausdruck $[h^2 - 1,96 \cdot s(\hat{h}^2)]$ bzw. $[d^2 - 1,96 \cdot s(\hat{d}^2)]$ positiv wird.

Die Ergebnisse aus Tabelle 1 sind wie folgt zu interpretieren. Im rein additiven Fall sind z.B. für 16 Mütter pro Vater mit je 5 Nachkommen bei $h^2=0,25$, $=0,17$, $=0,10$ und $=0,05$ insgesamt 4, 7, 12 und 25 Väter notwendig, um abzusichern dass h^2 signifikant größer Null geschätzt wird.

Tabelle 1

Anzahl der Halbgeschwistergruppen, so dass im additiven Fall die Heritabilität und im Dominanzfall der Anteil d^2 der Dominanzvarianz an der phänotypischen Varianz mit Konfidenzniveau $1-\alpha=0,95$ positiv geschätzt wird (Number of half-sib groups, so that in the additive case the heritability and in the dominance case the ratio d^2 of the dominance variance to the phenotypic variance is estimated greater than zero with confidence level 0.95)

Anzahl		h^2 ($y = \mu + a + u$)				$h^2 : d^2$ ($y = \mu + a + d + u$)			
Mütter	Nachk.	0,25	0,17	0,10	0,05	0,25:0,25	0,33:0,17	0,40:0,10	0,45:0,05
8	5	7	11	21	51	51	133	415	1830
12	5	5	9	15	33	34	94	303	1350
16	5	4	7	12	25	27	77	254	1149
20	5	4	6	10	20	23	67	226	1036
24	5	3	5	9	18	20	61	208	964

Fordert man dagegen, dass d^2 signifikant größer Null geschätzt werden soll, so erhöht sich für die Konstellationen $h^2:d^2=0,25:0,25$, $=0,33:0,17$, $=0,40:0,10$ und $=0,45:0,05$ die Anzahl der Väter auf 27, 77, 254 und 1149. Für die Konstellation mit 16 Müttern erhöht sich also die Anzahl der Datensätze bei $d^2=0,25$ von 320 auf 2160, bei $d^2=0,17$ von 560 auf 6140, bei $d^2=0,10$ von 960 auf 20320 und bei $d^2=0,05$ von 2000 auf 91920. In Abhängigkeit von den 4 Konstellationen für den Anteil der Dominanzvarianz ergibt sich eine Erhöhung des Stichprobenumfanges um mindestens das 6-, 10-, 20- und 40-fache beim Übergang vom rein additiven zum Dominanzmodell.

In Tabelle 2 ist die Anzahl der Datensätze aufgelistet, so dass der Standardfehler einer d^2 -Schätzung vorgegebene Werte von 0,08, 0,06, 0,04 und 0,03 besitzt. Wird Normalverteilung der Schätzungen vorausgesetzt, so ergeben sich als halbe Breite der zugehörigen zweiseitigen Konfidenzintervalle zum Niveau $1-\alpha=0,95$ die Werte 0,157, 0,118, 0,078 und 0,059.

Tabelle 2

Anzahl Datensätze, so dass der Standardfehler für den geschätzten Anteil der Dominanzvarianz an der phänotypischen Varianz bei 8, 6, 4 und 3% liegt (Number of records, so that the ratio of the dominance variance to the phenotypic variance is estimated with a standard error of 8, 6, 4 and 3%)

Varianzverhältnis	$h_w^2 = 0,5$				$h_w^2 = 0,3$			
	8%	6%	4%	3%	8%	6%	4%	3%
$\sigma_d^2 = \sigma_a^2$	5440	9680	21600	38480	3280	5840	13120	23360
$\sigma_d^2 = 0,5 \cdot \sigma_a^2$	6960	12320	27600	49040	3920	6880	15440	27360
$\sigma_d^2 = 0,25 \cdot \sigma_a^2$	8320	14720	33040	58640	4400	7760	17520	31040
$\sigma_d^2 = 1/9 \cdot \sigma_a^2$	9360	16640	37440	66480	4800	8560	27040	34080

Bei der Berechnung des erforderlichen Stichprobenumfanges in Tabelle 2 wurde die Vateranzahl minimiert, so dass der vorgegebene Stichprobenfehler unterschritten wird. Für die Ermittlung der minimalen Vateranzahl wurde generell das Vorhandensein von 16 Müttern pro Vater und 5 Nachkommen je Mutter vorausgesetzt. Während in Ta-

belle 1 die Varianzen im Dominanzmodell so gewählt wurden, dass sich für die Heritabilität im weiteren Sinn ein Wert von 0,5 ergibt ist in Tabelle 2 zusätzlich der Fall $h_w^2 = 0,3$ illustriert. Für den Fall $\sigma_d^2 = \sigma_a^2$ ergibt sich unter der Annahme $h_w^2 = 0,5$ bzw. $=0,3$ bei Verwendung von 5440 bzw. 3280 Datensätzen ein Standardfehler für die d^2 -Schätzung von 8%. Folglich besitzt zum Beispiel für $h_w^2 = 0,5$ zum Konfidenzniveau $1-\alpha=0,95$ das zugehörige Konfidenzintervall die Gestalt $[0,25-1,96 \cdot 0,08; 0,25+1,96 \cdot 0,08]$. Für die gleiche Konstellation liefert eine Schätzung von d^2 mit 21600 Datensätzen einen Standardfehler von 4 und mit 38480 Datensätzen einen Fehler von 3%. Analog ergibt sich für den Fall $\sigma_d^2 : \sigma_a^2 = 0,40 : 0,10$ bei Erhöhung des Stichprobenumfangs von $N=33040$ auf $N=58640$ lediglich eine Reduzierung des Standardfehlers von 0,04 auf 0,03. Unter der Hypothese $h_w^2 = 0,3$ lassen sich im Vergleich zu der Annahme $h_w^2 = 0,5$ gleiche Standardfehler mit geringerem Stichprobenumfang erzielen. Für die Konstellation $\sigma_d^2 = 0,25 \cdot \sigma_a^2$ resultiert ein Standardfehler von 4% bei $h_w^2 = 0,5$ unter Verwendung von 33040 Datensätzen während bei $h_w^2 = 0,3$ lediglich 17520 Datensätze erforderlich sind. Diese Abhängigkeit zwischen Stichprobenumfang und d^2 -Niveau entspricht den Verhältnissen der h^2 -Schätzung mit der Halbgeschwistermethode. Wie z.B. bei HERRENDÖRFER und SCHÜLER (1987) gezeigt, wird bei gleicher Nachkommenzahl pro Vater für kleinere h^2 -Werte ein geringerer Stichprobenumfang benötigt, um vorgegebene Werte für die halbe Breite eines Konfidenzintervalls einzuhalten.

Die bisherigen Ergebnisse zur Versuchsplanung setzen voraus, dass an einen Vater mehrere Mütter hierarchisch angepaart werden und pro Mutter mehr als ein Nachkomme vorhanden ist. Derartige hierarchische Halb- und Vollgeschwisterstrukturen sind einerseits beim Geflügel und beim Schwein typisch aber weichen andererseits von den Verhältnissen in der Rinderzucht ab.

Tabelle 3

Halbe Breite des erwarteten Konfidenzintervalls für den Anteil d^2 der Dominanzvarianz an der phänotypischen Varianz geschätzt mit unverwandten Voll- und Halbgeschwisterfamilien (mit h_w^2 =Heritabilität im weiteren Sinn=0,5) (Half length of the expected significance interval for the ratio d^2 of dominance to the phenotypic variance estimated with unrelated full- and half-sister families, whereby the heritability in the broad sense is 0.5)

records	Vollgeschwister		Halbgeschwister		$1,96 \cdot s(\hat{d}^2)$		
	Familien- größe	Anzahl Familien	Familien- größe	Anzahl Familien	$\sigma_d^2 = \sigma_a^2$	$\sigma_d^2 = 0,5 \cdot \sigma_a^2$	$\sigma_d^2 = 0,25 \cdot \sigma_a^2$
10000	5	200	10	900	0,2590	0,2643	0,2678
	5	400	10	800	0,2036*	0,2101	0,2146
	5	840	10	580	0,1804*	0,1896	0,1962
20000	5	400	10	1800	0,1832*	0,1869	0,1894
	5	800	10	1600	0,1439*	0,1485*	0,1517
	5	1680	10	1160	0,1276*	0,1341*	0,1387
40000	5	800	10	3600	0,1295*	0,1322*	0,1339
	5	1600	10	3200	0,1018*	0,1050*	0,1073
	5	3360	10	2320	0,0930*	0,0948*	0,0981*

*) Für $1-\alpha=0,95$ ergeben sich signifikant positive Schätzwerte für d^2 .

Nachfolgend wird die Vorgehensweise von VAN RADEN u.a. (1992) benutzt und vorausgesetzt, dass mehrere Typen (Gruppen) von Familien vorhanden sind. Den Ergebnissen von Tabelle 3 liegt die Annahme zugrunde, dass neben Voll- zusätzlich

Halbgeschwisterfamilien vorhanden sind, wobei Tiere aus verschiedenen Familien als unverwandt gelten und alle Tiere innerhalb einer Familie eines Typs im gleichen Verwandtschaftsverhältnis stehen. Die in Tabelle 3 für jede der drei Datensatzgrößen aufgeführten Varianten entsprechen einem Anteil von Vollgeschwisterfamilien in der Population von 10, 20 und 42% (vgl. Bemerkung 3).

Die Konstellation $\sigma_d^2 = (1/4) \cdot \sigma_a^2$ entspricht z.B. der Vorgabe $h^2:d^2=0,4:0,1$ in Tabelle 1. Zur Absicherung signifikant positiver Schätzwerte für d^2 werden gemäß Tabelle 1 bei hierarchischer Halb- und Vollgeschwisterstruktur für diese Konstellation rund 20000 Datensätze benötigt. Liegt der Anteil der Vollgeschwisterfamilien nur bei 42%, so muss der Stichprobenumfang in etwa verdoppelt werden um die gleiche Aussage treffen zu können. Stehen z.B. neben 3360 Vollgeschwisterfamilien mit 5 Tieren insgesamt 2320 Halbgeschwisterfamilien mit 10 Tieren pro Familie zur Verfügung, so lassen sich für alle in Tabelle 3 untersuchten Beziehungen zwischen σ_d^2 und σ_a^2 signifikant positive Schätzungen für d^2 gewinnen. Dagegen müssen bei einem Vollgeschwisteranteil von 42% unter Annahme der Konstellation $\sigma_d^2 = (1/2) \cdot \sigma_a^2$ immerhin noch rund 20000 Datensätze vorhanden sein während für den Fall $\sigma_d^2 = \sigma_a^2$ zumindest 10000 Datensätze zur Verfügung stehen müssen.

Betrachtet man in Tabelle 3 den Fall $\sigma_d^2 = 0,5 \cdot \sigma_a^2$, so liefert z.B. für fixierten Stichprobenumfang $N=20000$ und gegebene Werte $m_1=5$ und $m_2=10$ ein Vollgeschwisteranteil von 42,7% den geringsten Standardfehler für die Schätzung von d^2 (vgl. Bemerkung 3 Abschnitt 2).

4. Diskussion

Zur Schätzung der Dominanzvarianz auf dem Genauigkeitsniveau der Varianzschätzung für die additiv-genetische Komponente sind sehr hohe Stichprobenumfänge erforderlich. CHANG (1988) berechnete mit Hilfe von stochastischer Simulation die Anzahl Familien, so dass die additiv-genetische und die dominanzbedingte Varianz jeweils signifikant größer Null geschätzt werden. Für die Konstellation $\sigma_d^2 = 0,5 \cdot \sigma_a^2$ und $h_w^2 = 0,5$ bei 18 Tieren pro Familie erhöhte sich die Familienanzahl von 10 auf 208. Um für die Schätzung der Dominanzvarianz eine zur Schätzung für die additive Varianz äquivalente Genauigkeit zu gewährleisten, muss der Stichprobenumfang mindestens um das 20-fache erhöht werden. Aus den Ergebnissen von CHANG (1988) schlussfolgerten MISZTAL u.a. (1997), dass für die Konstellation $h^2:d^2=0,40:0,15$ rund 65000 Tiere notwendig sind, um für die Schätzung von d^2 einen Schätzfehler von nur 3% abzusichern. Die Schätzgenauigkeit unter Einbeziehung verschiedener Familientypen (Klone, Vollgeschwister, Elter-Nachkomme, $\frac{3}{4}$ -Geschwister, Halb- und Viertelgeschwister) aller verfügbaren Datensätze der U.S. Holstein-Population zur Schätzung von nicht-additiven Varianzen wurden von VAN RADEN u.a. (1992) ermittelt. Für diese weltgrößte Nutztierpopulation fanden die Autoren, dass die Dominanzvarianz mit einem Standardfehler von 1,5% an der phänotypischen Varianz und die Komponenten der Epistasie nur mit einem Standardfehleranteil von 10% geschätzt werden können. Um die Schätzung der Dominanzvarianz mit hoher Genauigkeit, also mit sehr geringem Stichprobenfehler vornehmen zu können, empfehlen MISZTAL u.a. (1995) die Verwendung von 200000 Tieren.

Derartig große Datensätze sind zwangsläufig für die meisten Tierarten nicht verfügbar. Die hier in Tabelle 1 und 2 aus Abschnitt 3 aufgeführten Ergebnisse zeigen, dass derartige Forderungen für hierarchische Zuchtstrukturen weit unterschritten werden können. Zur Absicherung von signifikant von Null verschiedener Schätzungen für d^2 sind bei $h_w^2 = 0,5$ für die Konstellation $\sigma_d^2 = 0,5 \cdot \sigma_a^2$ rund 6000 Datensätze und für die Konstellation $\sigma_d^2 = 0,25 \cdot \sigma_a^2$ allerdings bereits rund 20000 Datensätze erforderlich. Hierbei wurde eine Populationsstruktur mit 5 Nachkommen pro Mutter, 16 Mütter pro Vater und für die Beziehung $h^2:d^2=0,33:0,17$ bzw. $h^2:d^2=0,40:0,15$ das Vorhandensein von 77 bzw. 254 Vätern vorausgesetzt. Die wünschenswerte Anzahl von Datensätzen zur Schätzung der Dominanzvarianz, welche Konfidenzintervalle mit positiver unterer Grenze garantiert, ist folglich neben der Datenstruktur stark abhängig vom Verhältnis der Dominanzvarianz zur additiv-genetischen Varianz. Zum Erreichen äquivalenter Genauigkeiten wird die Schätzung von d^2 für Populationsstrukturen mit hohem Vollgeschwisteranteil, wie sie z.B. beim Geflügel und Schwein vorliegen, weniger Datensätze benötigen als die Schätzung der Dominanzvarianz beim Rind, wo Vollgeschwisterfamilien als Ergebnis von Embryotransfer entstehen. Bei diesem Vergleich muss allerdings vorausgesetzt werden, dass die beim Geflügel und Schwein untersuchten Merkmale nicht durch Effekte beeinflusst werden, die aus einer gemeinsamen Vollgeschwisterumwelt resultieren. In diesem Fall kann sich die Trennung zwischen Dominanz- und Vollgeschwistereffekten als schwierig erweisen (vgl. MISZTAL und BESBES, 2000). Folglich werden die hier abgeleiteten Stichprobengrößen zur Schätzung der Dominanzvarianz für die Verhältnisse beim Schwein nur untere Schranken darstellen. Soll eine Stichprobenplanung zur simultanen Schätzung von Dominanz- und Vollgeschwisterumweltvarianz durchgeführt werden, so müssen die verwandtschaftlichen Beziehungen innerhalb der Familien erweitert werden. So sind hierarchische Voll- und Halbgeschwisterstrukturen denkbar, wobei ein Teil der pro Vater angepaarten Mütter Halbschwestern väterlicherseits sind. In diesem Fall enthält also eine Familie neben Halb- und Vollgeschwistern zusätzlich auch $\frac{3}{4}$ -Geschwister (vgl. Van RADEN u.a., 1992), also Tiere die neben einem gemeinsamen Vater zusätzlich Mütter mit gleichem Großvater besitzen.

Für die Konstellation $h^2:d^2=0,4:0,1$ werden Standardfehler für die d^2 -Schätzung von 8, 6, 4 bzw. 3% realisiert, falls 8320, 14720, 33040 bzw. 58640 Datensätze vorhanden sind. Liegt der Fall $h^2:d^2=0,225:0,075$ vor, so reduziert sich der Stichprobenumfang auf 4400, 7760, 17520 bzw. 31040 Datensätze. Die Schätzung von d^2 mit einem Standardfehler zwischen 6 und 4 Prozent erfordert für hierarchische Zuchtstrukturen Stichprobenumfänge zwischen 10000 und 20000 Datensätze. Derartig hohe Anforderungen an die Stichprobengröße sind nur mit Daten aus mehreren Generationen realisierbar. Da die derzeit angewandte Theorie zur Aufstellung der inversen Dominanz-Verwandtschaftsmatrix die Abwesenheit von Inzucht voraussetzt, muss ein Kompromiss zwischen Inzuchtzuwachs, Veränderung der Allelfrequenzen hervorgerufen durch Selektion und Genauigkeitsforderungen bei der d^2 -Schätzung gefunden werden.

Literatur

CHANG, H.A.:

Studies on estimation of genetic variances under non-additive gene action. Ph.D. dissertation, University of Illinois, Urbana, USA (1988)

HERRENDÖRFER, G.; SCHÜLER, L.:

- Populationsgenetische Grundlagen der gerichteten Selektion. Gustav Fischer Verlag Jena (1987)
- HOESCHELE, I.; VAN RADEN, P.M.:
Rapid inversion of dominance relationship matrices for noninbred populations by including sire by dam subclass effects. *J. Dairy Sci.* **74** (1991) 557-569
- LÄUTER, H.; PINCUS, R.:
Mathematisch-statistische Datenanalyse. Akademie-Verlag Berlin (1989)
- MIELENZ, N.; NURGIARTININGSIH, V.; SCHMUTZ, M.; SCHÜLER, L.:
Schätzung von Varianzkomponenten mit Gruppenmittelwerten am Beispiel von Legehennen, gehalten in Gruppenkäfigen. *Arch. Tierz., Dummerstorf* **46** (2003) 5, 491-498
- MISZTAL, I.; FERNANDO, R.L.; GROSSMANN, M.; LAWLOR, T.J.; LUKASZEWICZ, M.:
Dominance and epistatic effects in genetic evaluation of farm animals. *Animal Science Papers and Reports* **13** (1995) no.4, 251-266
- MISZTAL, I.:
Estimation of variance components with large-scale dominance model. *J. Dairy Sci.* **80** (1997), 965-974
- MISZTAL, I.; VARONA, L.; CULBERTSON, M.; GENGLER, N.; BERTRAND, J.K.; MABRY, J.; LAWLOR, T.J.; VAN TASSELL, C.P.:
Studies on the value of incorporating effect of dominance in genetic evaluations of dairy cattle, beef cattle and swine. *Proc. 6th WCGALP* (1998) Armidahl
- MISZTAL, I.; BESBES, B.:
Estimates of parental-dominance and full-sib permanent environment variances in laying hens. *Animal Science* **71** (2000) 421-426
- PATTERSON, H.D.; THOMPSON, R.:
Recovery of inter-block information when block sizes are unequal. *Biometrika* **58** (1971), 545-554
- SCHÜLER, L.; SWALVE, H.; GÖTZ, K.U.:
Grundlagen der quantitativen Genetik. Verlag Eugen Ulmer, Stuttgart (2001)
- SEARLE, S.R.:
Notes on variance component estimation: a detailed account of maximum likelihood and kindred methodology. Cornell Univ., Ithaca, New York (1979)
- VAN RADEN, P.M.; LAWLOR, T.J.; SHORT, T.H.; HOESCHELE, I.:
Use of reproductive technology to estimate variances and predict effects of gene interactions. *Journal of Dairy Science* **75** (1992), 2892-2901
- WEI, M.; VAN DER WERF, J.H.J.:
Animal model estimation of additive and dominance variances in egg production traits of poultry. *J. Animal Science* **71** (1993), 57-65

Eingegangen: 09.10.2003

Akzeptiert: 10.08.2004

Anschrift der Verfasser

Dr. NORBERT MIELENZ, Univ.-Prof. Dr. habil. LUTZ SCHÜLER
Institut für Tierzucht und Tierhaltung mit Tierklinik der Martin-Luther-Universität
Halle-Wittenberg
Adam-Kuckhoff-Straße 35
D-06108 Halle

E-Mail: mielenz@landw.uni-halle.de