

Züchtungslehre - Repetition Quantitative Genetik

Peter von Rohr

2016-09-30

Einleitung

Hier soll das Konzept des **Zuchtwertes** wie er in der quantitativen Genetik für einen Genort eingeführt wurde, kurz repetiert werden. Der Sinn dieser Repetition ist, dass wir diesen Begriff des Zuchtwertes von einem Genort auf sehr viele (theoretisch gesehen unendlich viele) Genorte erweitern wollen. Diese Erweiterung wird uns dann in einem späteren Kapitel zur heute immer noch verwendeten BLUP-Zuchtwertschätzung führen.

Begriffe

Der **Phänotyp** oder die phänotypische Beobachtung ist eine messbare oder beobachtbare Eigenschaft oder Leistung eines Tieres. Beispiele dafür sind Milchleistung, Zuwachs, Körpergrösse, usw. Das **Genom** bezeichnet die totale genetische Information eines Individuums. Ein **Genort** oder **Locus** steht für eine definierte Position auf dem Genom. Der **Genotyp** steht für eine bestimmte Kombination von Genvarianten an einem Locus. Ein **Allel** bezeichnet eine mögliche Genvariante an einem Locus. Sind an einem bestimmten Genort die beiden Allele gleich, so spricht man von einem **homozygoten** Locus. Sind die beiden Allele verschieden wird der Locus als **heterozygot** bezeichnet.

Zuchtwert

Das fundamentale Problem in der Tierzucht besteht in der Auswahl der Elterntiere für eine kommende Generation. Diese Auswahl geschieht immer mit einer Idealvorstellung eines optimalen Tieres im Hintergrund. Diese Idealvorstellung wird auch als **Zuchtziel** bezeichnet. Konkret sollen die Elterntiere der kommenden Generation so ausgewählt werden, dass die zukünftigen Nachkommen näher am Zuchtziel sind im Vergleich zur aktuellen Population, aus der wir die Eltern auswählen.

Vererbung

Eltern vererben ihren Nachkommen keine Phänotypen und keine Genotypen sondern eine zufällige Auswahl jeweils einer Variante jedes Genorts.

Informationsquellen

Heute ist es zwar möglich ganze Genome zu sequenzieren und somit die komplette Information der gesamten Erbmasse eines Individuums zu bekommen. Leider ist die Entschlüsselung des Genoms, das heisst die Zuordnung der Genorte zu wichtigen Merkmalsausprägungen noch nicht sehr fortgeschritten. Somit können wir unsere Zuchtentscheide noch nicht auf das ganze Genom abstützen.

In der **genomischen Selektion** wird eine für die wichtigsten Merkmale repräsentative Stichprobe von Genorten bestimmt. Die Varianten an diesen Genorten werden mit erwünschten Merkmalsausprägungen in Verbindung gebracht. Diese Assoziation zwischen Genort und Merkmalsausprägung liefern wertvolle Hinweise bei der Auswahl der Elterntiere.

Bei den **traditionellen Zuchtverfahren** ohne Berücksichtigung der genomischen Information stehen phänotypische Leistungen und Verwandtschaftsbeziehungen zwischen Tieren zur Verfügung. Dabei werden phänotypische Beobachtungen und das genetische Potential über das folgende einfache Modell miteinander verknüpft.

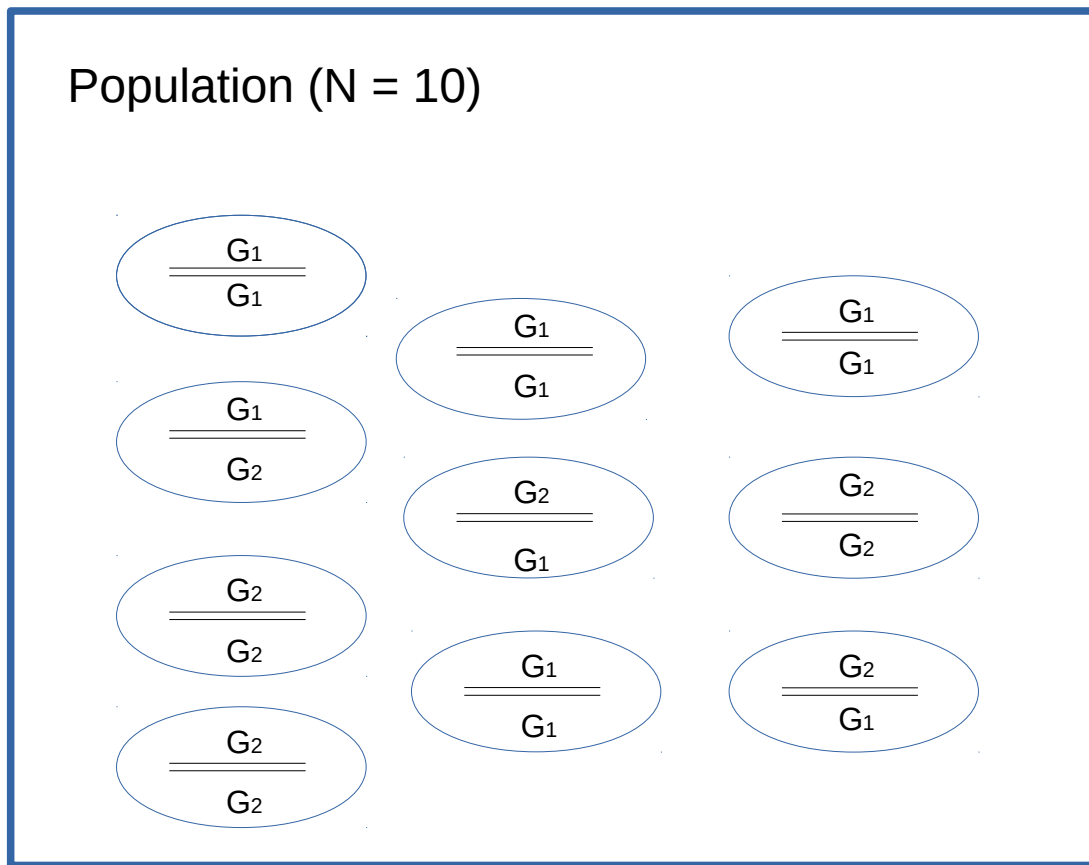
$$y = g + e \quad (1)$$

wobei: y für die phänotypische Beobachtung steht, g für den genetischen Anteil und e für den nicht-erklärbaren Restanteil.

Das Modell (1) ist so parametrisiert, dass die nicht-erklärbaren Resteffekte e sich im Mittel aufheben und somit ist der Erwartungswert $E[e] = 0$. Daraus folgt, dass im Mittel die genetischen Effekte g den mittleren phänotypischen Leistungen y entsprechen, d.h. $E[y] = E[g]$. Bei fehlender Genotyp-Umweltinteraktion sind die Varianzen additiv und somit folgt, dass

$$Var(y) = Var(g) + Var(e)$$

Das Konzept der Modellierung der phänotypischen Leistung mittels genetischem Anteil und umweltbedingtem Restanteil kann mit folgendem Schema veranschaulicht werden.



Aus diesem einfachen Schema lassen sich sehr einfach die Frequenzen $f(G_1G_1)$, $f(G_1G_2)$ und $f(G_2G_2)$ der Genotypen G_1G_1 , G_1G_2 und G_2G_2 ableiten. In der gezeigten Population befinden sich vier Individuen mit dem Genotyp G_1G_1 , und je drei Individuen mit den Genotypen G_1G_2 und G_2G_2 . Somit sind die Frequenzen

$$\begin{aligned}
 f(G_1G_1) &= \frac{4}{10} = 0.4 \\
 f(G_1G_2) &= \frac{3}{10} = 0.3 \\
 f(G_2G_2) &= \frac{3}{10} = 0.3
 \end{aligned}$$

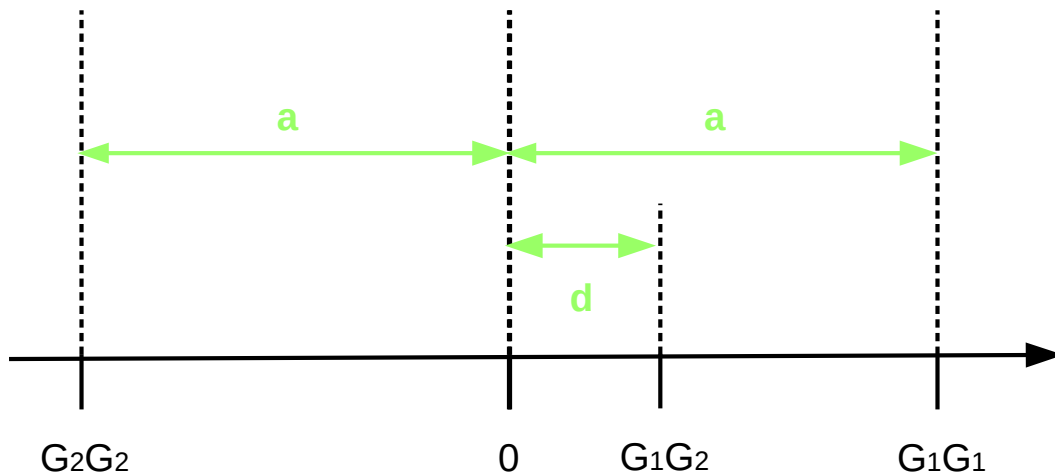
Die Allelefrequenzen $f(G_1)$ und $f(G_2)$ lassen sich aufgrund der Genotypfrequenzen berechnen.

$$\begin{aligned}
 f(G_1) &= f(G_1G_1) + 1/2f(G_1G_2) = 0.4 + 0.15 = 0.55 \\
 f(G_2) &= f(G_2G_2) + 1/2f(G_1G_2) = 0.3 + 0.15 = 0.45
 \end{aligned}$$

Genotypischer Wert

Der genotypische Wert V_{ij} für einen bestimmten Genotypen G_iG_j ist der mittlere Wert aller Individuen mit dem gleichen Genotyp G_iG_j in der gleichen Umwelt. Zunächst nehmen wir an, dass wir nur einen Genort (1 Gen) mit 2 Allelen (Genvarianten) betrachten. Die Population befinde sich im Hardy-Weinberg Gleichgewicht, welches besagt, dass sich bei zufälliger Paarung und unendlicher Populationsgrösse die Frequenz $f(G_iG_j)$ der Genotypen G_iG_j von einer Generation zur nächsten konstant bleibt.

Die Skala, auf welcher wir die genotypischen Werte V_{ij} messen wird der Einfachheit halber so gelegt, dass der Nullpunkt zwischen den beiden homozygoten Genotypen G_2G_2 und G_1G_1 liegt. Das nachfolgende Diagramm zeigt die genotypischen Werte der drei Genotypen.



Aus dem oben gezeigten Diagramm erkennen wir, dass G_1 das vorteilhafte Allel ist. Das heisst mit jedem G_1 steigt der genotypische Wert. Die folgende Tabelle fasst die genotypischen Werte für die drei Genotypen zusammen.

Genotyp	genotypischer Wert
G_1G_1	$V_{11} = a$
G_1G_2	$V_{12} = d$
G_2G_2	$V_{22} = -a$

Populationsmittel

Eine Kenngrösse der gezeigten genotypischen Werte, welche wir später brauchen werden ist das Populationsmittel μ . Dies entspricht dem Erwartungswert $E[V]$ der genotypischen Werte. Allgemein ist der Erwartungswert $E[X]$ für eine diskrete Zufallsvariable X definiert als

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

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

Wenden wir diese Definition auf die genotypischen Werte an, dann beträgt die Menge aller möglichen Genotypen $\mathcal{X} = \{V_{11}, V_{12}, V_{22}\}$ und somit beträgt das Populationsmittel.

$$\mu = E[V] = f(G_1G_1) * V_{11} + f(G_1G_2) * V_{12} + f(G_2G_2) * V_{22} \quad (2)$$

Die Frequenzen $f(G_1)$ und $f(G_2)$ der Allele G_1 und G_2 werden häufig mit den Variablen $p = f(G_1)$ und $q = f(G_2) = 1 - p$ bezeichnet. Aufgrund des Hardy-Weinberg Gleichgewichts resultieren die folgenden Genotypfrequenzen.

Genotyp	Frequenzen
G_1G_1	$p * p = p^2$
G_1G_2	$p * q + q * p = 2pq$
G_2G_2	$q * q = q^2$

Setzen wir die Frequenzen aus obiger Tabelle in die Gleichung (2) für das Populationsmittel ein, dann erhalten wir

$$\begin{aligned} \mu &= p^2 * a + 2pq * d - q^2 * a \\ &= (p^2 - q^2)a + 2pqd \\ &= (p + q)(p - q)a + 2pqd \\ &= (p - q)a + 2pqd \end{aligned} \quad (3)$$

Das Populationsmittel ist von den Allelfrequenzen (p und q) und somit von der Selektion abhängig.

Definition des Zuchtwertes

Für die Zucht interessieren uns die genetischen Anlagen und deren Wert, welche potentielle Elterntiere an ihre Nachkommen weitergeben. Bei der Vererbung geben Eltern keine Genotypen, sondern eine zufällige Stichprobe der Allele an ihre Nachkommen weiter. Für die Auswahl von potentiellen Elterntieren aus einer Population geht es darum die erwartete Leistung von Nachkommen eines bestimmten Elterngenotyps zu bestimmen.

Der **Zuchtwert** eines bestimmten Tieres i ist definiert als die doppelte Abweichung des erwarteten Mittelwertes der Nachkommen von Tier i vom Populationsmittel.

Herleitung des Zuchtwertes ZW_{11} für den Genotyp G_1G_1

Zur Herleitung des Zuchtwertes ZW_{11} müssen wir die erwartete mittlere Leistung der Nachkommen eines Tieres mit Genotyp G_1G_1 bestimmen. Die mittlere Leistung der Nachkommen eines Tieres mit Genotyp G_1G_1 ist als Erwartungswert definiert. Zur Berechnung des Erwartungswertes brauchen wir die Genotypfrequenzen der Nachkommen des Tieres mit Genotyp G_1G_1 . Diese Frequenzen werden in der folgenden Tabelle bestimmt.

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

Die Frequenzen der Nachkommen ist symmetrisch betreffend der Eltern. In der oben gezeigten Tabelle wird eine Mutter mit Genotyp G_1G_1 zufällig an Väter angepaart. Bei der Mutter kommt nur das G_1 Allel vor. Somit betragen die Allelfrequenzen $f(G_1) = 1$ und $f(G_2) = 0$, wobei das G_2 Allel bei der Mutter in der obigen Frequenz-Tabelle weggelassen wurde. Durch die Annahme der zufälligen Anpaarung der Mutter an die Väter entsprechen die Allelfrequenzen bei den Vätern dem Populationsdurchschnitt und somit ist $f(G_1) = p$ und $f(G_2) = q$.

Der erwartete Mittelwert der genotypischen Werte der Nachkommen der Mutter mit Genotyp G_1G_1 ist somit

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

Gemäss Definition können wir den Zuchtwert ZW_{11} für die Mutter mit Genotyp G_1G_1 berechnen als

$$\begin{aligned}
 ZW_{11} &= 2 * (\mu_{11} - \mu) \\
 &= 2 (pa + qd - [(p - q)a + 2pqd]) \\
 &= 2 (pa + qd - (p - q)a - 2pqd) \\
 &= 2 (qd + qa - 2pqd) \\
 &= 2 (qa + qd(1 - 2p)) \\
 &= 2q (a + d(1 - 2p)) \\
 &= 2q (a + (q - p)d)
 \end{aligned} \tag{4}$$

Zuchtwerte für die Genotypen G_2G_2 und G_1G_2 können analog hergeleitet werden.

Herleitung des Zuchtwertes ZW_{22} für den Genotyp G_2G_2

Als erstes bestimmen wir wieder die Frequenzen der Nachkommen einer Mutter mit Genotyp G_2G_2

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

Der erwartete Mittelwert der genotypischen Werte der Nachkommen der Mutter mit Genotyp G_2G_2 ist somit

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

Der Zuchtwert ZW_{22} ist somit

$$\begin{aligned}
 ZW_{22} &= 2 * (\mu_{22} - \mu) \\
 &= 2 (pd - qa - [(p - q)a + 2pqd]) \\
 &= 2 (pd - qa - (p - q)a - 2pqd) \\
 &= 2 (pd - pa - 2pqd) \\
 &= 2 (-pa + p(1 - 2q)d) \\
 &= -2p(a + (q - p)d)
 \end{aligned} \tag{5}$$

Herleitung des Zuchtwertes ZW_{12} für den Genotyp G_1G_2

Als erstes bestimmen wir wieder die Frequenzen der Nachkommen einer Mutter mit Genotyp G_1G_2

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

Der erwartete Mittelwert der genotypischen Werte der Nachkommen der Mutter mit Genotyp G_1G_2 ist somit

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

Der Zuchtwert ZW_{22} ist somit

$$\begin{aligned}
ZW_{12} &= 2 * (\mu_{12} - \mu) \\
&= 2 (0.5(p - q)a + 0.5d - [(p - q)a + 2pqd]) \\
&= 2 (pd - qa - (p - q)a - 2pqd) \\
&= (q - p)a + (1 - 4pq)d \\
&= (q - p) [a + (q - p)d]
\end{aligned} \tag{6}$$

Zusammenfassung der Zuchtwerte

In den Resultaten der Zuchtwerte kommt der Term $a + (q - p)d$ vor. Wir bezeichnen diesen Term mit α und fassen die Zuchtwerte in folgender Tabelle zusammen.

Genotyp	Zuchtwert
G_1G_1	$2q\alpha$
G_1G_2	$(q - p)\alpha$
G_2G_2	$-2p\alpha$

Allelsubstitution

Vergleicht man die beiden Genotypen G_2G_2 und G_1G_2 , dann kann der erste Genotyp in den zweiten verwandelt werden, indem eines der G_2 Allele vom ersten Genotypen durch ein G_1 -Allel ersetzt wird. Dies wird als **Allelsubstitution** bezeichnet. Der Unterschied der Zuchtwerte ZW_{12} und ZW_{22} beträgt

$$\begin{aligned}
ZW_{12} - ZW_{22} &= (q - p)\alpha - (-2p\alpha) \\
&= (q - p)\alpha + 2p\alpha \\
&= (q - p + 2p)\alpha \\
&= (q + p)\alpha \\
&= \alpha
\end{aligned} \tag{7}$$

Ersetzt man im Genotyp G_1G_2 das verbleibende G_2 -Allel durch ein G_1 -Allel so resultiert der Genotyp G_1G_1 . Die Differenz der zugehörigen Zuchtwerte ZW_{11} und ZW_{12} beträgt somit

$$\begin{aligned}
ZW_{11} - ZW_{12} &= 2q\alpha - (q - p)\alpha \\
&= (2q - (q - p))\alpha \\
&= \alpha
\end{aligned} \tag{8}$$

Die Zuchtwerte sind abhängig von den Allelfrequenzen. Die Differenzen der Zuchtwerte sind additiv, d.h. die Abhängigkeit der Differenzen zur Anzahl der G_1 -Allele ist linear.

Dominanzabweichung

Aufgrund der Definition des Zuchtwerts sind Abweichungen zwischen dem genotypischen Wert V_{ij} und dem Zuchtwert ZW_{ij} eines Tieres zu erwarten. Betrachtet man diese Abweichungen zwischen genotypischem Wert und Zuchtwert für die drei Genotypen G_1G_1 , G_1G_2 und G_2G_2 allgemein so erhalten wir folgende Resultate

$$\begin{aligned}
 V_{11} - ZW_{11} &= a - 2q\alpha \\
 &= a - 2q[a + (q - p)d] \\
 &= a - 2qa - 2q(q - p)d \\
 &= a(1 - 2q) - 2q^2d + 2pqd \\
 &= [(p - q)a + 2pqd] - 2q^2d \\
 &= \mu + D_{11}
 \end{aligned} \tag{9}$$

$$\begin{aligned}
 V_{12} - ZW_{12} &= d - (q - p)\alpha \\
 &= d - (q - p)[a + (q - p)d] \\
 &= [(p - q)a + 2pqd] + 2pqd \\
 &= \mu + D_{12}
 \end{aligned} \tag{10}$$

$$\begin{aligned}
 V_{22} - ZW_{22} &= -a - (-2p\alpha) \\
 &= -a + 2p[a + (q - p)d] \\
 &= [(p - q)a + 2pqd] - 2p^2d \\
 &= \mu + D_{22}
 \end{aligned} \tag{11}$$

Zusammenfassung: Genotypische Werte, Zuchtwerte und Dominanzabweichungen

Die folgende Tabelle gibt eine Übersicht über Genotypische Werte, Zuchtwerte und Dominanzabweichungen

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	$-a$	$-2p\alpha$	$-2p^2d$

Aufgrund der Definition der Dominanzabweichung als Differenz vom um das Populationsmittel korrigierte genotypische Wert minus der Zuchtwert folgt die in Gleichung (12) gezeigte Zerlegung des genotypischen Wertes.

$$V_{ij} = \mu + ZW_{ij} + D_{ij} \tag{12}$$

Varianzen

Genotypische Werte, Populationsmittel und Zuchtwerte wurden als Erwartungswerte definiert. Diese sind Lageparameter und geben Informationen für quantitative Vergleiche zwischen Tieren. Da die hier beschriebenen Prozesse der Vererbung nicht deterministisch sind, sondern immer auch eine stochastische oder zufällige Komponente haben, spielt die Streuung der verwendeten Modellparameter immer auch eine wichtige Rolle.

Für eine diskrete Zufallsvariable X ist die Varianz $Var[X]$ definiert als

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

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

Wenden wir diese Definition auf die genotypischen Werte V_{ij} an, dann erhalten wir den folgenden Ausdruck für die genetische Varianz.

$$\begin{aligned} \sigma_G^2 = Var[V] &= (V_{11} - \mu)^2 * f(G_1 G_1) \\ &+ (V_{12} - \mu)^2 * f(G_1 G_2) \\ &+ (V_{22} - \mu)^2 * f(G_2 G_2) \end{aligned} \quad (13)$$

wobei $\mu = (p - q)a + 2pqd$ dem Populationsmittel entspricht.

Gemäss dem Modell aus Gleichung (12) können wir die Abweichungen der genotypischen Werte vom Populationsmittel ersetzen durch $V_{ij} - \mu = ZW_{ij} + D_{ij}$. Die Genotypfrequenzen ersetzen wir durch die aus dem Hardy-Weinberg bekannten Terme. Somit können wir die genetische Varianz schreiben als

$$\begin{aligned} \sigma_G^2 &= (ZW_{11} + D_{11})^2 * p^2 \\ &+ (ZW_{12} + D_{12})^2 * 2pq \\ &+ (ZW_{22} + D_{22})^2 * q^2 \end{aligned} \quad (14)$$

Für die Zuchtwerte ZW_{ij} und die Dominanzabweichungen D_{ij} können wir nun die bekannten Definitionen einsetzen. Danach lässt sich der Ausdruck für die genetische Varianz vereinfachen. Diese Vereinfachung ist im Anhang (siehe Gleichung (24)) gezeigt. Das Resultat für die genetische Varianz lautet

$$\sigma_G^2 = 2pq\alpha^2 + (2pqd)^2 \quad (15)$$

Wie im Anhang in den Gleichungen (25) und (26) gezeigt wurde, entspricht der Term $2pq\alpha^2$ der mit den Genotypfrequenz gewichteten Summe der quadrierten Zuchtwerte. Der Term $(2pqd)^2$ kann aus der mit den Genotypfrequenzen gewichteten Summe der quadrierten Dominanzabweichungen berechnet werden.

Aufgrund der Beziehungen zwischen Zuchtwerten, welche wir in den Gleichungen (7) und (8) gezeigt haben, werden Zuchtwerte als **additiv** bezeichnet. Da der Term $2pq\alpha^2$ von den Zuchtwerten abhängt wird dieser auch als **genetisch additiv** Varianz σ_A^2 bezeichnet. Analog dazu bezeichnen wir den Term $(2pqd)^2$ als Dominanzvarianz σ_D^2 . Somit lässt sich die genetische Varianz in einen additiv genetischen Teil und einen Dominanz-Teil zerlegen. Formell können wir schreiben

$$\begin{aligned} \sigma_G^2 &= 2pq\alpha^2 + (2pqd)^2 \\ &= \sigma_A^2 + \sigma_D^2 \end{aligned} \quad (16)$$

Erweiterung auf mehrere Genorte

Bis hierher haben wir uns mit der Modellierung des Einflusses eines Genortes auf ein bestimmtes kontinuierliches Merkmal. Viele wirtschaftlich interessante Merkmale sind aber von sehr vielen verschiedenen Genorten beeinflusst. Falls verschiedene Genorte auf dem gleichen Chromosom zu finden sind, dann ist deren Vererbung und deren Wirkung nicht mehr unabhängig voneinander.

Abhängigkeit zwischen Genorten

Mit Unabhängigkeit meinen wir hier, dass die Frequenz der Genotypen an zwei Loci nicht dem Produkt der Frequenzen der Genotypen an einem Locus entspricht. Für zwei Loci A und B mit je zwei Allelen sieht das beispielsweise so aus, dass die Frequenz $f(A_1A_1B_1B_1)$ nicht dem Produkt aus $f(A_1A_1)$ mal $f(B_1B_1)$ entsprechen.

$$f(A_1A_1B_1B_1) \neq f(A_1A_1) * f(B_1B_1) \quad (17)$$

Interaktionen

Falls die Wirkung mehrerer Genorte auf ein Merkmal nicht der Summe der Wirkungen der einzelnen Gene entspricht, so spricht man von Interaktionen. Betrachten wir wieder die beiden Genorte A und B mit genotypischen Werten $V_{ij}^{(A)}$ und $V_{kl}^{(B)}$, dann entspricht der totale genotypische Wert g nicht der Summe zwischen den genotypischen Werten an den einzelnen Loci, sondern es kommt noch ein sogenannter **Epistasieeffekt** hinzu.

$$\begin{aligned} g + \mu &= V_{ij}^{(A)} + V_{kl}^{(B)} + I_{ijkl}^{(A) \cdot (B)} \\ &= \mu^{(A)} ZW_{ij}^{(A)} + D_{ij}^{(A)} + \mu^{(B)} + ZW_{kl}^{(B)} + D_{kl}^{(B)} + I_{ijkl}^{(A) \cdot (B)} \end{aligned} \quad (18)$$

wobei $\mu^{(A)}$ und $\mu^{(B)}$ die Populationsmittel aufgrund der Wirkungen der Loci A und B sind. Mit μ bezeichnen wir die Summe aus den Populationsmitteln, d.h. $\mu = \mu^{(A)} + \mu^{(B)}$. Aufgrund der Interaktionen lassen sich die genotypischen Werte nicht mehr in additive Zuchtwerte plus Dominanzabweichung zerlegen. Hinzu kommt neu die Komponente der Epistasie.

Polygenes Modell

Betrachten wir anstelle von nur zwei Loci sehr viele (theoretisch unendlich viele) Genorte, dann sprechen wir von einem **polygenen Modell**. Für dieses Modell ist die Schreibweise aus Gleichung (18) nicht mehr praktikabel. Als Verbesserung fassen wir alle additiven Komponenten zusammen zu einem einzigen Effekt a , alle Dominanzeffekte werden im Effekt d zusammengefasst und alle Interaktionen werden mit dem Effekt i bezeichnet. Somit können wir den genetischen Wert zerlegen in

$$g = a + d + i \quad (19)$$

Da wir die Effekte a , d und i als Abweichungen vom Populationsmittel definiert haben sind deren Erwartungswerte alle gleich Null. Es gilt also, dass $E[a] = E[d] = E[i] = 0$. Somit resultiert der Erwartungswert für g als

$$E[g] = E[a] + E[d] + E[i] = 0$$

Auch bei der genetische Varianz σ_G^2 kommt aufgrund von Interaktionen zwischen Genorten eine zusätzliche Varianzkomponente, nämlich die epistatische Varianz σ_I^2 hinzu.

$$\sigma_G^2 = \sigma_A^2 + \sigma_D^2 + \sigma_I^2 \quad (20)$$

Zuchtwert als additiv genetischer Effekt

Bei polygene Merkmalen gehen wir von sehr vielen Genorten, wobei jeder Genort nur eine sehr kleine Wirkung hat, aus. Da jeder Elternteil eine Stichprobe seiner Allele weitergibt, sind für die Züchtung vor allem die additiven Effekte a von grosser Bedeutung. Diesen additiv genetische Effekt a bezeichnen wir als **Zuchtwert**. Dominanz- und Epistasie-Effekte kommen erst durch die spezielle Kombinationen von Allelen und Genorten zustande und sind deshalb sehr schwer zu quantifizieren. Aus diesem Grund werden Dominanz und Epistasie oft vernachlässigt.

Der zentrale Grenzwertsatz sagt, dass die Verteilung der Summe aus sehr vielen Zufallsvariablen mit sehr geringer Wirkung sich einer Normalverteilung nähert. Das Argument des zentralen Grenzwertsatzes und die Annahme, dass sich der additiv genetische Effekt aus Beiträgen von vielen Genorten zusammensetzt, rechtfertigen die Wahl der **Gausschen Normalverteilung** als Dichteverteilung für die Zuchtwerte.

Phänotypisch Beobachtung

Am Anfang dieses Dokuments hatten wir schon gesehen, dass wir bei den traditionellen Zuchtverfahren die genetischen Werte und somit die Zuchtwerte nicht beobachten können. Als Informationsquellen stehen uns phänotypische Beobachtungen und die Verwandtschaftsbeziehungen zwischen den Tieren in der Population zur Verfügung. Wir brauchen also ein Modell, welches uns die phänotypischen Informationen mit genotypischen Werten verknüpft. Dieses Modell hatten wir schon in der Gleichung (1) gesehen. Da wir alle genetischen Effekte als Abweichungen vom Populationsmittel definiert haben, müssen wir das Modell (1) noch leicht modifizieren, indem wir das Populationsmittel μ aus dem g -Effekt herausziehen und es explizit als Faktor hinschreiben. Somit haben wir für die phänotypische Beobachtung y folgendes Modell

$$y = \mu + g + e \quad (21)$$

Da auch der umweltbedingte Resteffekt e des Modells als Abweichung von μ definiert wurde, ist der Erwartungswert der phänotypischen Beobachtung $E[y] = \mu$. Die Varianz der phänotypischen Beobachtung kann aus dem Modell (21) abgeleitet werden.

$$Var[y] = \sigma_p^2 = \sigma_g^2 + \sigma_e^2 + 2Cov[g, e] \quad (22)$$

Der Ausdruck $Cov[g, e]$ im Modell (22) tritt nur dann auf, falls Genotyp-Umweltinteraktionen bestehen. Falls solche Interaktionen zwischen Genotypen und der Umwelt fehlen, dann können wir $Cov[g, e] = 0$ setzen und somit vereinfacht sich das Modell (22) zu

$$\sigma_p^2 = \sigma_g^2 + \sigma_e^2 \quad (23)$$

Anhang

Hier wird die Vereinfachung der genetischen Varianz gezeigt.

$$\begin{aligned}
\sigma_G^2 &= (ZW_{11} + D_{11})^2 * p^2 \\
&\quad + (ZW_{12} + D_{12})^2 * 2pq \\
&\quad + (ZW_{22} + D_{22})^2 * q^2 \\
&= (2q\alpha - 2q^2d)^2 * p^2 \\
&\quad + ((q-p)\alpha + 2pqd)^2 * 2pq \\
&\quad + (-2p\alpha - 2p^2d)^2 * q^2 \\
&= (4q^2\alpha^2 - 8q^3d\alpha + 4q^4d^2) * p^2 \\
&\quad + (q^2\alpha^2 - 2pq\alpha^2 + p^2\alpha^2 - 4(q-p)pqd\alpha + 4p^2q^2d^2) * 2pq \\
&\quad + (4p^2\alpha^2 + 8p^3d\alpha + 4p^4\alpha^2) * q^2 \\
&= 4p^2q^2\alpha^2 - 8p^2q^3d\alpha + 4p^2q^4d^2 \\
&\quad + 2pq^3\alpha^2 - 4p^2q^2\alpha^2 + 2p^3q\alpha^2 \\
&\quad - 8p^3q^2d\alpha + 8p^2q^3d\alpha + 8p^3q^3d^2 \\
&\quad + 4p^2q^2\alpha^2 + 8p^3q^2d\alpha + 4p^4q^2d^2 \\
&= 4p^2q^2\alpha^2 + 4p^2q^4d^2 \\
&\quad + 2pq^3\alpha^2 + 2p^3q\alpha^2 \\
&\quad + 8p^3q^3d^2 \\
&\quad + 4p^4q^2d^2 \\
&= 2pq\alpha^2 (p^2 + 2pq + q^2) \\
&\quad + (2pqd)^2 (p^2 + 2pq + q^2) \\
&= 2pq\alpha^2 + (2pqd)^2
\end{aligned} \tag{24}$$

Der Term $2pq\alpha^2$ entspricht der Summe der mit den Genotypfrequenzen gewichteten quadrierten Zuchtwerte. Konkret heisst das

$$\begin{aligned}
2pq\alpha^2 &= p^2ZW_{11}^2 + 2pqZW_{12}^2 + q^2ZW_{22}^2 \\
&= p^2(2q\alpha)^2 + 2pq((q-p)\alpha)^2 + q^2(-2p\alpha)^2 \\
&= 4p^2q^2\alpha^2 + 2pq(q^2\alpha^2 - 2pq\alpha^2 + p^2\alpha^2) + 4p^2q^2\alpha^2 \\
&= 4p^2q^2\alpha^2 + 2pq^3\alpha^2 - 4p^2q^2\alpha^2 + 2p^3q\alpha^2 + 4p^2q^2\alpha^2 \\
&= (4p^2q^2 + 2pq^3 + 2p^3q)\alpha^2 \\
&= 2pq(2pq + q^2 + p^2)\alpha^2 \\
&= 2pq\alpha^2
\end{aligned} \tag{25}$$

Den Term $(2pqd)^2$ setzen wir der Summe der mit den Genotypenfrequenzen gewichteten quadrierten Dominanzabweichungen gleich. Die entsprechende Ableitung ist nachfolgend gezeigt.

$$\begin{aligned}
(2pqd)^2 &= p^2 D_{11}^2 + 2pq D_{12}^2 + q^2 D_{22}^2 \\
&= p^2 (-2q^2 d)^2 + 2pq (2pqd)^2 + q^2 (-2p^2 d)^2 \\
&= 4p^2 q^4 d^2 + 8p^3 q^3 d^2 + 4p^4 q^2 d^2 \\
&= 4p^2 q^2 d^2 (q^2 + 2pq + p^2) \\
&= 4p^2 q^2 d^2
\end{aligned} \tag{26}$$