

# Das Leben ist multiplikativ - neue Aspekte zur Verteilung von Daten

E. LIMPERT und W. A. STAHEL

## Zusammenfassung

Quantitative Variation ist für die Genetik, Züchtungsforschung und Züchtung von grundlegender Bedeutung. Unser aktuelles Verständnis - auch in der Wissenschaft generell - beinhaltet allerdings einen fundamentalen Gegensatz, der nicht ohne Auswirkung auf die Verteilung und Analyse von Daten sein kann: Die Lebensgesetze basieren im Wesentlichen auf Multiplikationen, während die Gaußsche Normalverteilung als bekanntester Standard zur Beschreibung der Variation auf Addition beruht.

Ein stehender Begriff in der Züchtungsforschung ist die additive Genwirkung. Sie wird mit normalverteilter Variation in Zusammenhang gebracht, auch wenn es für die additive Wirkung bislang keinen Beweis gibt. Weniger bekannt, wenn auch häufig vorkommend und, wie gezeigt wird, besser nachweisbar ist die multiplikative Genwirkung, die im Einklang mit den Grundprinzipien zu schiefer, log-normalverteilter Variation führt. Aus diesen und weiteren Gründen verdient die Log-Normalverteilung stärkere Beachtung.

Hier wird ein umfassendes Konzept zum Verständnis beider Verteilungen vorgestellt mit einer Reihe neuer und hilfreicher Aspekte. Es beginnt mit mechanischen Modellen, führt zu einer Brücke von der etablierten Normal- zur weit weniger etablierten Log-Normalverteilung und reicht bis hin zu Vorschlägen für eine nahe liegende und einfache Charakterisierung auch der Log-Normalverteilung.

Die Modelle ähneln Glücksspielmaschinen, demonstrieren die Genesis der Verteilungen, und wirken durch den spielerischen Umgang damit einer möglichen Statistik-Phobie entgegen. Ebenso lässt sich ein gewisser Zauber der Statistik daran erkennen, dass mittels unseres Konzepts die bislang schwer zugänglichen und oft missachteten Log-Normal-

verteilungen einfach zu handhaben und selbst per Kopfrechnen zugänglich und vorstellbar werden.

Vorteile und neue Möglichkeiten werden an eingängigen Beispielen aus verschiedenen Bereichen der Wissenschaft demonstriert und mögliche Anwendungen in der Züchtung diskutiert. Für ein verbessertes Verständnis von Variation in der Genetik und ihren Anwendungen, wie auch in vielen Bereichen unseres Lebens, werden so Grundlagen geschaffen, die sowohl für Experten als auch allgemein von Interesse sind.

## Ein wesentlicher Gegensatz

Das Leben ist multiplikativ. Für diese generelle Feststellung im Titel gibt es zahllose Beispiele. Betrachten wir zunächst einmal das Wachstum. Bakterien und Zellen vermehren sich durch Teilung, und die Zweiteilung ist eine Basis des Lebens. Daraus resultiert exponentielles Wachstum, wo der Zuwachs proportional zur vorhandenen Biomasse ist, die sich somit proportional und multiplikativ verändert.

Wesentlich für Wachstum und Entwicklung sind Stofftransport und -mobilität. Im Hagen-Poiseuilleschen Strömungsgesetz sind, als zweites Beispiel, insgesamt sieben Multiplikationsschritte enthalten. Drittens wurde kürzlich für die Permeabilität und Mobilität von Stoffen in pflanzlichem Gewebe nachgewiesen, dass die grundlegenden Prozesse nicht

auf additiven, sondern auf multiplikativ wirkenden Schritten beruhen müssen (BAUR, 1997, sowie ergänzende Auswertung von LIMPERT et al., 2001).

Das grundlegendste der hier genannten Beispiele schließlich betrifft chemische Reaktionen. Wenn A mit B zu C reagiert, so ergibt sich die Reaktionsgeschwindigkeit  $v$  aus der Multiplikation der Konzentrationen von A und B, entsprechend der Formel

$$v \sim [A] \times [B] .$$

Ganz offensichtlich sind ja chemische Reaktionen die Grundlage allen Lebens. Ein eindrucksvolles Bild der Stoffwechselwege mit ihren vielfachen Vernetzungen liefern z.B. die Biochemischen Wandkarten von BOEHRINGER, auch wenn die Komplexität des gesamten Stoffwechsels im Leben noch um ein Vielfaches größer ist. Ganz generell basieren die Natur- und Lebensgesetze auf Multiplikationen, und das gilt natürlich auch für die grundlegende Biologie in der Pflanzenzüchtung.

*Wie ist es im Vergleich dazu mit der Variation im Leben?* Variation gilt bislang überwiegend als Gauß- oder normalverteilt. Die Normalverteilung wird oft mit dem bekannten Zentralen Grenzwertsatz begründet, der sagt, dass viele kleine additive Effekte näherungsweise zu einer normalverteilten Summe führen. Die genannten multiplikativen Kräfte stehen damit im Widerspruch. (Auf die wenig bekannte multiplikative Version des zentralen Grenzwertsatzes kommen wir noch zurück.)

In Genetik und Züchtung wurden additive und multiplikative Genwirkungen oft

**Tabelle 1: Unterschiede additiver und multiplikativer Genwirkung. In dem Zahlenbeispiel sind die Mittel der Eltern und der F1 dargestellt sowie die erwartete Verteilung um das jeweilige F1-Mittel in der F2. Näheres siehe Text.**

Mittelwerte, Mediane		
Eltern		P1: 10, P2: 90
Erwartet in der F1	additiv	50
	multiplikativ	30
Verteilung in der F2	additiv	Normalverteilung
	multiplikativ	Log-Normalverteilung

**Autoren:** Dr. Eckhard LIMPERT, Scheuchzerstraße 210, CH-8057 ZÜRICH, Dr. Werner A. STAHEL, Seminar für Statistik, Eidgenössische Technische Hochschule Zürich, CH-8092 ZÜRICH



beschrieben und diskutiert, wenn auch nicht immer nachgewiesen. Was für ein Unterschied ist bei diesen Genwirkungen zu erwarten? Das kann am Zahlenbeispiel (siehe *Tabelle 1*) zu Ertrag oder Blütenlänge im Prinzip dargestellt werden:

Bei additiver Genwirkung liegt die F1 beim *arithmetischen* Mittel  $\bar{x}$  von 50 - mit einer Differenz von 40 zu beiden Eltern - und die F2 verteilt sich symmetrisch und *normal* um das Mittel. Bei multiplikativer Genwirkung liegt die F1 beim *geometrischen* Mittel  $\bar{x}^*$  von 30 ( $10 \times 90$ )<sup>1/2</sup> - mit einem multiplikativen Unterschied vom Faktor 3 zu beiden Eltern - und die F2 verteilt sich asymmetrisch und *log-normal* um dieses Mittel (Beispiele s.u.).

## Normal- und Log-Normalverteilung

Bevor wir nach dieser prinzipiellen Darstellung auf die genannten Fragen zurückkommen, wollen wir uns zunächst den additiven und multiplikativen Prinzipien bei der Verteilung von Daten zuwenden. Neben verschiedenen bekannten Verteilungen im Leben (z.B. Exponential-, Gamma-, Weibull-) sind die *Normal-* und *Log-Normalverteilung* weiterhin am wichtigsten. Deshalb macht es Sinn, sich angesichts des Tagungsthemas darauf noch einmal zu konzentrieren. Die Normalverteilung ist seit langem etabliert und generell gut bekannt, im Gegensatz zur Log-Normalverteilung, die die Wirkung von multiplikativen Effekten adäquat beschreibt.

Angesichts unserer Prägung auf Symmetrie und Normalverteiltes weisen Log-Normalverteilungen auf der Ebene der ursprünglichen Daten eine unerwünschte Schiefe auf. Als Folge kann es zu Fehlern bei der Auswertung kommen. Nun weist zwar der Name darauf hin, dass die Daten nach Logarithmierung normalverteilt sind, aber dann wiederum sind die Logarithmen unhandlich und störend.

So wird Multiplikatives in der Wissenschaft oft als additiv und symmetrisch dargestellt. Selbst in begutachteten Zeitschriften kann z.B. die Sensitivität von Pilz-Isolaten gegen Chemikalien beschrieben werden als „log-normal, mit einer mittleren EC<sub>50</sub> von 4,1±3.7“ (STEHMANN und DE WAARD, 1996). Diese für Normalverteilungen übliche Darstellung mit Mittelwert und Standardabweichung

macht für schiefe Verteilungen wenig Sinn. Die Verteilung der Daten ist auf diese Weise schlecht vorstellbar, und es ist dann schwierig, auf die wichtige Frage nach dem genetischen Anteil der Variation überhaupt zu kommen. Auf eine sinnvolle Charakterisierung dieser Verteilung kommen wir noch zurück.

Beispiele multiplikativer Vererbung sind bei quantitativen Merkmalen seit langem bekannt: Die Blütengröße (Corolla-Länge) beim Tabak (EAST, 1913), die Fruchtgröße bei Tomate (MACARTHUR und BUTLER, 1938), die Resistenz bei Tomate gegen Hohlfrüchtigkeit (PALEVITCH und KEDAR, 1970). Resistenz gegen Mehltau, Braunrost, Rhynchosporium und Lager entsprechen ebenfalls multiplikativer Vererbung mehr als additiver (G. BACKES, pers. Mitt.). KOCH (1966) schließlich beschreibt u.a. auch an Beispielen aus der Züchtungsforschung Mechanismen, die die Log-Normalverteilung exakt erzeugen.

In der Pflanzen- und Tiergenetik sind multiplikative Effekte als Erklärung von u.a. Dominanz, Heterosis und Epistasie gut analysiert (WRIGHT, 1922; RICHHEY, 1942; POWERS, 1944; COCKERHAM, 1959; WILLIAMS, 1959; GEIGER und WAHLE, 1978; BECKER, 1984; GEIGER, 1988, SCHNELL und COCKERHAM, 1992). Die Wuchshöhe z.B. ist das Produkt aus Anzahl und Länge von Internodien. Bei komplexen Merkmalen lässt sich schließlich feststellen, dass die relativen Effekte in Teilbereichen weniger stark ausgeprägt sind als beim Gesamt-Merkmal. Dabei entsprechen die quantitativen Ergebnisse einer multiplikativen Wirkung (WRIGHT, 1922; DICKERSON, 1955).

Insofern gibt es hier Zusammenhänge, die von den multiplikativen Effekten in den Naturgesetzen und den ermittelten Ergebnissen bis hin zum statistischen Modell der Log-Normalverteilung reichen.

Von diesen Fragen ausgehend sollen zunächst die additiven und multiplikativen Effekte als Grundlage von Normal- und Log-Normalverteilungen näher betrachtet werden. Dabei ermöglicht es unser Konzept, beide Verteilungen auf einem sehr einfachen Niveau und dennoch wissenschaftlich korrekt zu behandeln. Mechanische Modelle erlauben sogar das

spielerische Erlernen statistischer Grundlagen. So ist es eines unserer Ziele, mit diesem Beitrag nicht nur Experten sondern selbst Schüler anzusprechen.

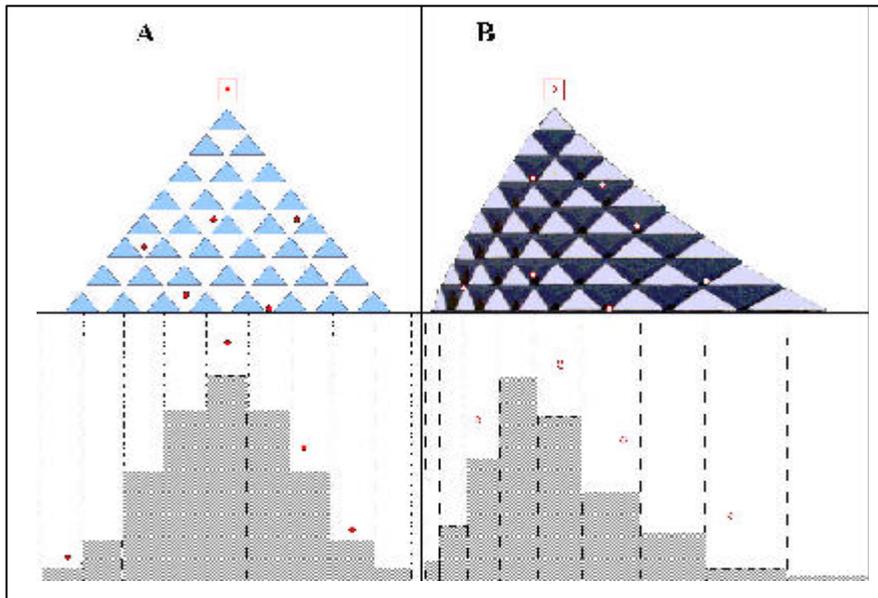
Ausgehend vom etablierten Wissen über die Normalverteilung werden dann Brücken zum Verständnis der weniger etablierten Log-Normalverteilung gebaut. So lässt sich erkennen, dass *Log-Normales* ähnlich einfach zu verstehen und zu handhaben ist wie *Normales*. Schließlich werden die sich daraus ergebenden Möglichkeiten an Beispielen quantitativer Variation dargestellt und diskutiert.

Im Vergleich zu bisherigen Publikationen zur Genetik, Züchtungsforschung und Züchtung wird hier offensichtlich, dass die wissenschaftlichen Argumente unseres Beitrags von einer viel generelleren Betrachtungsweise kommen. Diese ist in vielen Fällen sowohl für die Analyse von Daten aus diesen Gebieten, wie auch weit darüber hinaus sinnvoll und wichtig.

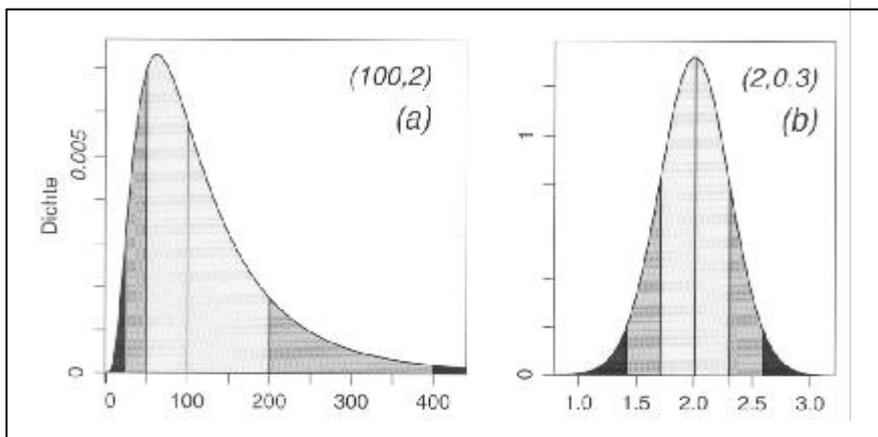
## Modelle zur quantitativen Variation

Der Normal- und der Log-Normalverteilung ist gemeinsam, dass sie durch die Wirkung vieler kleiner, unabhängiger Effekte entstehen. Das entspricht dem Zentralen Grenzwertsatz der Mathematik. Additive Effekte führen dabei zur Normalverteilung, während multiplikative Effekte zur Log-Normalverteilung führen. Während der Additive Zentrale Grenzwertsatz gut bekannt ist, ist die multiplikative Form (AITCHISON und BROWN, 1957) bislang selbst unter Mathematikern wenig geläufig und verdient stärkere Beachtung.

Beide Verteilungen lassen sich an Hand mechanischer Modelle im wahrsten Sinne des Wortes *begreifen* und somit besser verstehen (*Abbildung 1*). Das Modell der Normalverteilung ist eine Weiterentwicklung des ursprünglich von GALTON (1889) publizierten und nach ihm benannten GALTON-Brettes und verkörpert den additiven Zentralen Grenzwertsatz. Auf diesem und dem KAPTEYN-Brett aufbauend (KAPTEYN, 1903) wurde kürzlich von LIMPERT ein neues Modell für die Log-Normalverteilung entwickelt als Verkörperung der multiplikativen Variante des Zentralen Grenzwertsatzes (LIMPERT et al., 2001).



**Abbildung 1: Mechanische Modelle zur Entstehung der Normal- und der Log-Normalverteilung.** In beiden Modellen treffen Partikel von einem Ausgangspunkt auf die Spitzen von Dreiecken, wo sie nach rechts und links mit gleicher Wahrscheinlichkeit abgelenkt werden und schließlich in Auffangbehältern landen. Wenn die Spitze des obersten Dreiecks an der horizontalen Position 100 ist, können Spitzen weiterer Dreiecke nach rechts unten, für die Normalverteilung (A) an den Positionen 110, 120, 130 usw. sein. Die Veränderung entspricht dann jeweils einem additiven Wert von 10. Für die Log-Normalverteilung (B) im Vergleich, können die Spitzen der weiteren Dreiecke wiederum nach rechts unten an den Positionen 110, 121, 133 usw. sein - entsprechend einem multiplikativen Faktor von 1,1. Die Spitzen weiterer Dreiecke nach links unten sind dann an den Positionen 90, 80 usw. für die Normalverteilung, und 91, 83 usw. für die Log-Normalverteilung. So verändert sich der Partikelstrom in einer additiven ( $\pm$ , plus/minus) bzw. multiplikativen ( $\cdot$ , mal/geteilt durch) Weise. Entsprechend können also diese Geschwister-Verteilungen als **additive** bzw. **multiplikative Normalverteilung** bezeichnet werden.



**Abbildung 2: Eine Log-Normalverteilung und eine handliche Art der Charakterisierung.** Beide Grafiken zeigen dieselbe Verteilung, auf dem Niveau der ursprünglichen Daten (a) und nach logarithmischer Transformation (b). Die Charakterisierung von Log-Normalverteilungen auf dem Niveau der ursprünglichen Daten durch  $\bar{x}^* = 100$  und  $s^* = 2$  (a) ermöglicht Kopfrechnen und Schätzen (s. Text): Eine gute Basis für wissenschaftliches Denken.

GALTON (1889) spricht im Zusammenhang mit quantitativer Variation und seinem Brett auch vom Zauber der Statistik, *the charms of statistics*, der mit solchen

„Glücksspielmaschinen“ besser erfahren werden kann. Für sein Modell verwendete er Nägel, wo die aktuellen Modelle didaktisch vorteilhafte Dreiecke aufwei-

sen. So sind diese Modelle für Lehrzwecke u.a. auf dem Internet (<http://stat.ethz.ch/vis/lognormal>) so beliebt, dass man sie seit langem in der Suchmaschine Google unter dem Suchbegriff „log-normal“ an erster Stelle findet.

### Eine Brücke vom Etablierten zum weniger Etablierten

Auf diesen Modellen von Normal- und Log-Normalverteilung baut ein Konzept auf, das auf dem Niveau der ursprünglichen Daten eine verblüffend einfache Handhabung auch der Log-Normalverteilung erlaubt. Man bekommt damit die sonst so unangenehme Schiefe leicht in den Griff, und selbst Kopfrechnen wird wieder möglich (Abbildung 2).

Die Log-Normalverteilung ist ja dadurch charakterisiert, dass die logarithmus-transformierte Größe normalverteilt ist. Daraus lassen sich die grundlegenden Eigenschaften der Log-Normalverteilung mathematisch recht einfach ableiten. Die übliche Art, eine Log-Normalverteilung durch Parameter festzulegen, benutzt den Mittelwert  $\bar{x}$  und Standardabweichung  $s$  der logarithmierten Beobachtungen. Die Rück-Transformation dieser Werte führt zum geometrischen Mittel  $\bar{x}^*$  und der multiplikativen Standardabweichung  $s^*$ . Das erlaubt nun eine handliche Charakterisierung auf dem Niveau der ursprünglichen Daten: Aus  $\bar{x}^* = 100$  und  $s^* = 2$  ergibt sich per Kopfrechnen der „1s-Bereich“ der Daten zwischen 50 und 200 ( $= 100 \cdot \frac{1}{2}$ ) und der „2s-Bereich“ zwischen 25 und 400 ( $= 100 \cdot \frac{1}{2^2}$ ).

Zusammenfassend ergibt sich die einfache Brücke von der etablierten Normalverteilung zu der so sinnvollen und zu wenig bekannten Log-Normalverteilung wie in Tabelle 2 dargestellt.

Interessanterweise ist die Log-Normalverteilung in einem Punkt sogar einfacher: Während der Variationskoeffizient als Dispersionsmaß für die Normalverteilung extra berechnet werden muss, ist das geeignete Maß für die Log-Normalverteilung einfach gleich  $s^*$ . Ein Gefühl für  $s^*$ -Werte und die zugehörigen Formen von Log-Normalverteilungen lässt sich aus Abbildung 3 gewinnen. Diese Verteilungen sind sehr vielfältig. Bislang macht man sich sehr selten klar, dass Log-Normalverteilungen fast symmetrisch

Tabelle 2: Eine Brücke von der Normal- zur Log-Normalverteilung

	Normalverteilung oder Additive Normalverteilung	Log-Normalverteilung oder Multiplikative Normalverteilung
Wirkung der Faktoren	additiv	multiplikativ
Charakterisierung durch Mittelwert,	$\bar{x}$ , arithmetisch	$\bar{x}^*$ , geometrisch
Standardabweichung	s, additiv	$s^*$ , multiplikativ
Dispersionsmaß	$cv = s / \bar{x}$	$s^*$
Vertrauensintervalle		
68.3 % = 1s	$\bar{x} \pm s$	$\bar{x}^* \times/ s^*$
95.5 % = 2s	$\bar{x} \pm 2s$	$\bar{x}^* \times/ (s^*)^2$
99.7 % = 3s	$\bar{x} \pm 3s$	$\bar{x}^* \times/ (s^*)^3$

cv Variationskoeffizient,  $\times/$  mal/geteilt durch, entsprechend plus/minus für  $\pm$

sein können und damit statistisch nicht unterscheidbar sind von Normalverteilungen. Ebenso können sie extrem schief sein: Der mittlere 95%-Bereich der schiefsten hier dargestellten Kurve reicht z.B. von 0,592 - 16'900 ( $100 \times/ 13^2$ ), und es gibt weit schiefere Verteilungen, z.B. in der Ökologie (Tabelle 3).

Im Gegensatz dazu kann man die uns so geläufigen Normalverteilungen in ihrem Anwendungsbereich für Original-Daten sogar als ziemlich limitiert und speziell bezeichnen: Bei passenden Datensätzen liegen die Variationskoeffizienten normalerweise unter 10%. Wegen der zu Grunde liegenden Symmetrie kann selbst im Extremfall kein Wert mehr als 100% nach beiden Seiten schwanken. Für viele An-

wendungen mag der „3s-Bereich“ ausreichen, was den Variationskoeffizienten auf 33% limitiert. Für Selektionsprozesse in einer Pilzpopulation jedoch können die dann außerhalb liegenden 0,3% noch sehr wichtig sein: Bei Populationsgrößen von nahezu  $10^{20}$  und annähernd 20 Generationen jährlich, wie das z.B. beim Getreidemehltau in Europa der Fall ist (DEISING et al, 2002; LIMPERT, Daten unveröffentlicht), wären für Modellrechnungen selbst „4-und-mehr-s-Bereiche“ noch sehr interessant, was den Variationskoeffizienten noch weiter limitiert. Auch wenn nicht davon auszugehen ist, dass für so extreme Werte ein einfaches Modell noch passt, wird die Log-Normalverteilung doch zu sinnvolleren Vorstellungen führen als die Normalverteilung.

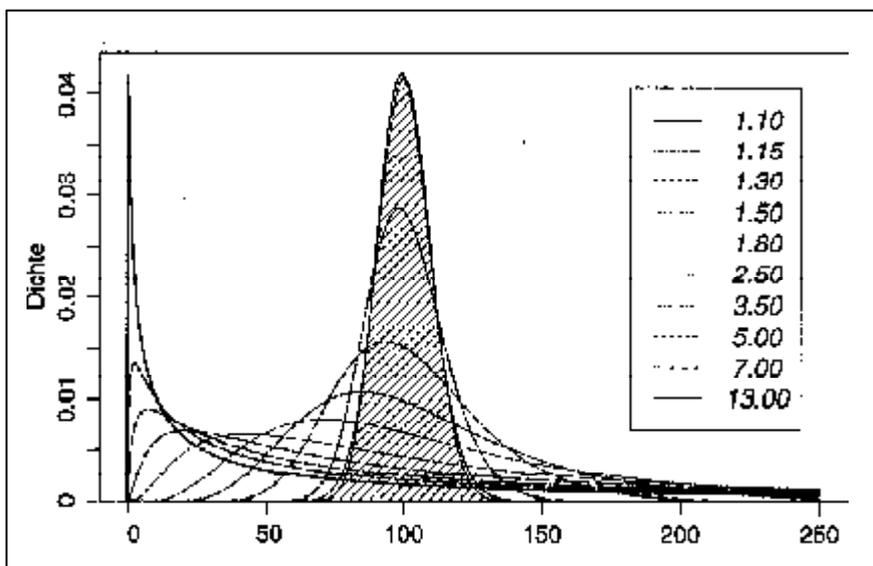


Abbildung 3: Multiplikative (oder Log-) Normalverteilungen sind vielfältig und einfach. Angegeben sind die Dichtefunktionen von zehn Verteilungen mit  $\bar{x}^* = 100$  und ausgewählten  $s^*$ -Werten, im Vergleich zu einer Normalverteilung des selben Mittels. Es ist interessant zu sehen, dass Log-Normalverteilungen ein sehr weites Spektrum abdecken von fast symmetrisch bis extrem schief, während Normalverteilungen, die zu empirischen Daten gut passen auf einen engen Bereich beschränkt sind, mit Variationskoeffizienten von zumeist weniger als dem hier angegebenen von 10% (grau, schraffiert).

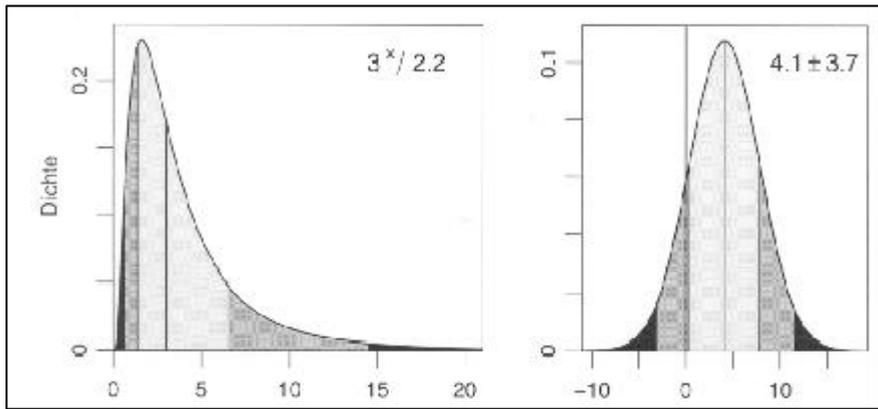
Das eingangs genannte Problem der Charakterisierung einer Log-Normalverteilung mit den Werten  $4,1 \pm 3,7$  wird aus *Abbildung 4* (rechts) deutlich, ebenso wie die Lösung des Problems (links). Die beiden Parameter ( $\bar{x}^* = 3,0$  und  $s^* = 2,2$ ) geben ein passendes Bild und sind handlich. Mit dem korrekten Erkennen der multiplikativen Zusammenhänge ist dann der Schritt naheliegend, den Varianz-Ursachen durch Varianzanalyse der logarithmierten Werte nachzugehen. Nur dann können sich für den genetischen Anteil der Variation sinnvolle Schätzungen ergeben.

### Beispiele und Vergleiche quantitativer Variation

Log-normalverteilte Daten sind ubiquitär. Bei Ergebnissen aus der Literatur liegen die  $s^*$ -Werte im Vergleich zu den Verteilungen in *Abbildung 3* zwischen 1,03 und >33 (LIMPERT et al., 2001). Einen allgemein interessanten Überblick dazu gibt *Tabelle 3*.

Größere Zusammenhänge ergeben sich aus der Betrachtung zusammenfassender Daten verschiedener Untersuchungen eines Merkmals oder von Merkmalsgruppen (siehe oberer Teil von *Tabelle 3*). Der Beginn der Alzheimer-Krankheit lag im Mittel bei ca. 60 Jahren, mit einer relativ geringen Variation von  $s^* = 1,16$ . Bei den Infektionskrankheiten fällt auf, wie stark die mittleren Latenzzeiten  $\bar{x}^*$  schwanken (von Stunden bei Nahrungsmittelvergiftung bis zu Monaten und Jahren bei Serumhepatitis, AIDS und BSE), während die Variation mit einem  $s^*$  um 1,5 relativ typisch ist. Da die Infizierten von ihrem Schicksal im Prinzip ja nichts wissen, entspricht dies vermutlich der biologischen Variation im gegebenen Umfeld.

Bei der Überlebenszeit nach Krebsdiagnose kommen zu der biologischen weitere Variationsursachen hinzu, wie die des Diagnosezeitpunkts und der Therapie. Also dürfte  $s^*$  hier größer sein, was tatsächlich der Fall und mit einem Wert von 2,8 sehr deutlich ist. Während die mittleren 95% der Daten bei Infektionskrankheiten also um einen Faktor von ca. 5 streuen ( $1,5^4$ ), streuen die Überlebenszeiten entsprechend um einen Faktor von >60. Stärker noch variiert die Regenmenge, sowie die Besiedelung ökologischer Räume und Nischen, die sich in der Indi-



**Abbildung 4: Charakterisierung einer multiplikativen (oder Log-) Normalverteilung auf dem Niveau der ursprünglichen Daten - Sinnvolles und weniger Sinnvolles. Die beiden Parameter ( $\bar{x}^* = 3,0$  und  $s^* = 2,2$ ) geben ein passendes Bild der Verteilung (links). Eine Charakterisierung durch  $\bar{x} = 4,1$  und  $s = 3,7$  (rechts) wäre zwar formal korrekt, die entsprechende Normalverteilung passt jedoch nicht zu den Daten. Sie beinhaltet u.a. Werte unter Null. Aus  $\bar{x}$  und  $s$ , die bei solchen Ergebnissen häufig angegeben sind, lassen sich  $\bar{x}^*$  und  $s^*$  einfach berechnen (LIMPERT et al., 2001).**

viduenzahl pro biologischer Art widerspiegelt.  $s^*$  kann somit ein Maß für die Komplexität des Merkmals bzw. die Stärke der beteiligten Faktoren sein.

Bei den Beispielen aus dem Alltagsleben (einzelne Untersuchungen, unterer Tabeleinteil) liegt  $s^*$  nahe dem der Infektionskrankheiten mit 1,5. Das ist der Wert, der bei der vorhandenen Auswahl an Ergebnissen am häufigsten war. Zur Ermittlung der Wortlängen-Variation wurden englische Telefongespräche analysiert. Der Einkommensverteilung in der Schweiz im Jahre 1990 lagen  $1,7 \times 10^6$  Haushalte zu Grunde, und schließlich mag interessant sein, dass selbst beim Genuss von Eiscreme offenbar Log-Normalverteiltes eine Rolle spielt.

Auf Grund dieser Erkenntnisse wäre es nun interessant, die umfangreiche Literatur zur quantitativen Vererbung einmal unter solchen Aspekten zu re-evaluieren.

In welchem Bereich bewegen sich die  $s^*$ -Werte in der Züchtung und Züchtungsforschung? Und gibt es hier für einzelne Bereiche ebenfalls typische  $s^*$ -Werte, wie z.B. bei Blüten- und Fruchtgrößen, Gehalt an Inhaltsstoffen, Krankheitsresistenz, Lager, Qualitätseigenschaften, Anzahl Borsten bei Drosophila und QTL-Effekte (EAST, 1913; MACARTHUR und BUTLER, 1938; PALEVITCH und KEDAR, 1970; LYMAN et al., 1996; G. BACKES, pers. Mitt.). Es kann sein, dass sich die Variation zum Zeitpunkt des Alterschiebens als eines von vielen Wachstumsmerkmalen in verschiedenen Aspekten ähnlich verhält wie die in dem genannten Beispiel zu den Latenzzeiten.

Es wäre auch interessant zu sehen, ob sich aus Re-Evaluierungen neue Erkenntnisse über quantitative Zusammenhänge bei Dominanz, Heterosis und Epistasie gewinnen lassen (WRIGHT 1922; RICH-

EY, 1942; POWERS, 1944; COCKERHAM, 1959; WILLIAMS, 1959; GEIGER und WAHLE, 1978; BECKER, 1984; GEIGER, 1988, SCHNELL und COCKERHAM, 1992). Und wenn bei komplexen Merkmalen festgestellt wurde, dass die Effekte in Teilbereichen weniger stark ausgeprägt waren als beim Gesamt-Merkmal (WRIGHT, 1922; DICKERSON, 1955), wie wären dann die zugehörigen  $s^*$ -Werte?

Für die Resistenzzüchtung und Krankheitsentwicklung spielt je nach Epidemiologie die Konzentration an Inokulum auf der Blattoberfläche und in der Luft - auch in Ihren saisonalen Schwankungen - oder bei Spritzwasser-Verbreitung die Verteilung der Tröpfchengröße eine Rolle. Für diese Kriterien wurden Log-Normalverteilungen als zutreffend befunden, und die Daten zeigen wiederum aufschlussreiche Muster von Mittelwerten und Streuung (LIMPERT et al., 2001 und unveröffentlicht).

### Geschichtlicher Seitenblick

Bei der Betrachtung der Normal- und der Log-Normalverteilung und ihrer Berücksichtigung für die Grundlagen der Genetik ist ein geschichtlicher Seitenblick aufschlussreich. die Geschichte der Normalverteilung begann vor knapp 300 Jahren, als der Hugenotte Abraham DE MOIVRE (1738) in London Ratschläge für Glücksspieler gab. Die der Log-Normalverteilung ist nur etwa halb so alt (WEBER, 1834; FECHNER, 1860; GALTON, 1879; MCALISTER, 1879) und setzte damit zu einem Zeitpunkt ein, als die Gaußsche Glockenkurve bereits auf dem Weg zur Popularität war.

GAUSS war Mathematiker und Astronom und hat sich mit der Variation biologischer Daten offenbar kaum befasst. Auch zur Zeit R.A. FISHER'S (1918) und seiner klassischen Arbeit über genetische Variation und Korrelation war der multiplikative Zentrale Grenzwertsatz (AITCHISON und BROWN, 1957) nicht beschrieben.

So wurde von additiver Genwirkung ausgegangen (FISHER, 1918). Aufbauend z.B. auf einem Effekt von 2 für das Allel A und 1 für Allel B gelangt man so zu einem additiven Modell (ALLARD, 1960, *Tabelle 10-1*), das eingängig und weit verbreitet ist. Ein entsprechendes

**Tabelle 3: Ausgewählte Beispiele multiplikativer (oder Log-) Normalverteilungen. Die Ergebnisse sind auf dem Niveau der ursprünglichen Daten charakterisiert (LIMPERT et al., 2001).  $\bar{x}^*$  ist das geometrische Mittel oder der Median,  $s^*$  das multiplikative Streuungsmaß und der Form-Parameter der Verteilung. Interessant ist der Vergleich der  $s^*$ -Werte hier mit denen der Kurven in *Abbildung 3*.**

Disziplin/Bereich	Beispiel	$\bar{x}^*$	$s^*$
Medizin	- Beginn von Alzheimer	~ 60 Jahre	1,16
	- Latenzzeiten von Infektionskrankheiten	Stunden - Monate	1,5
	- Überlebenszeit nach Krebsdiagnose	Monate - Jahre	2,8
Umwelt	- Regen, beimpft und unbeimpft	80 - 200 ( $\times 10^3$ m <sup>3</sup> )	4-5
	- Ökologie: Anzahl Individuen pro Art	-	6-33
Alltagsleben	- Sprache, Anzahl Buchstaben bei 76'054 Worten, und den 738 unterschiedlichen Worten	3,12 5,05	1,65 1,47
	- Einkommen unselbstständig-erwerbender Haushalte	6'726 SFR	1,54
	- Nahrungsmittel: Kristallgrößen von Eiscreme	15µm	1.50

multiplikatives Modell lässt sich ebenso entwickeln, ausgehend z.B. von einem mittleren Ertrag von 100 und multiplikativen Effekten von 1,02 für Allel A und 1,01 für Allel B. Mit Erträgen von z.B. 106,13 ( $1,02^2 \times 1,01^2 \times 100$ ) für den Genotyp AABB und entsprechend von 94,22 für den Genotyp aabb ist es dem additiven Modell sehr ähnlich und auch statistisch kaum davon zu unterscheiden.

Angesichts des geschichtlichen Hintergrunds anfangs des letzten Jahrhunderts ist das Folgende um so interessanter (COCHRAN, 1980). Obwohl die Varianzanalyse mit additiven Modellen eng verbunden ist, verwarf FISHER in seiner ersten Varianzanalyse das additive Modell zu Gunsten des multiplikativen (FISHER und MACKENZIE, 1923). Der Rest des mittleren Abweichungsquadrats des Fehlers von  $836,3/55 = 15,39$  war dort leicht besser als der von 17,8 des additiven Modells, wie FISHER erwartet zu haben scheint. Überraschenderweise beruht dann jedoch praktisch die gesamte spätere Arbeit FISHERS zur Varianzanalyse auf dem additiven Modell, ohne Hinweis warum das multiplikative fallengelassen wurde.

## Schlussfolgerungen und Ausblick

Log-Normalverteilungen und Normalverteilungen sind von wesentlichem Interesse für Pflanzenzüchtung und Züchtungsforschung. Die hier vorgestellten mechanischen und virtuellen Modelle ermöglichen es, auf spielerische Art einen wissenschaftlich exakten Zugang zu den Grundlagen dieser Verteilungen zu gewinnen, wobei die nicht seltenen Statistik-Ängste reduziert werden.

Der häufige Rat, Daten für die Auswertung zu logarithmieren, wird durch unser Konzept wesentlich untermauert. Unsere Überlegungen ermöglichen es, als Ergänzung zum Denken in logarithmierten Werten, die Variation der ursprünglichen Daten mit geeigneten Begriffen zu erfassen. Ausgehend von der Normalverteilung bietet das Konzept eine Brücke zur Log-Normalverteilung und führt zu einer gut vorstellbaren Art der Charakterisierung multiplikativer Variation. Die Charakterisierung ist handlich, nutzerfreundlich, und erfolgt auf der Ebene der ursprünglichen Daten. Die Möglichkeit zum Kopfrechnen und Schätzen macht das Konzept *alltagstauglich*, und die

Hinweise auf weiterführende Auswertungen und Vergleiche können zu Lichtblicken und *wissenschaftlichen Sonn- und Feiertagen* führen.

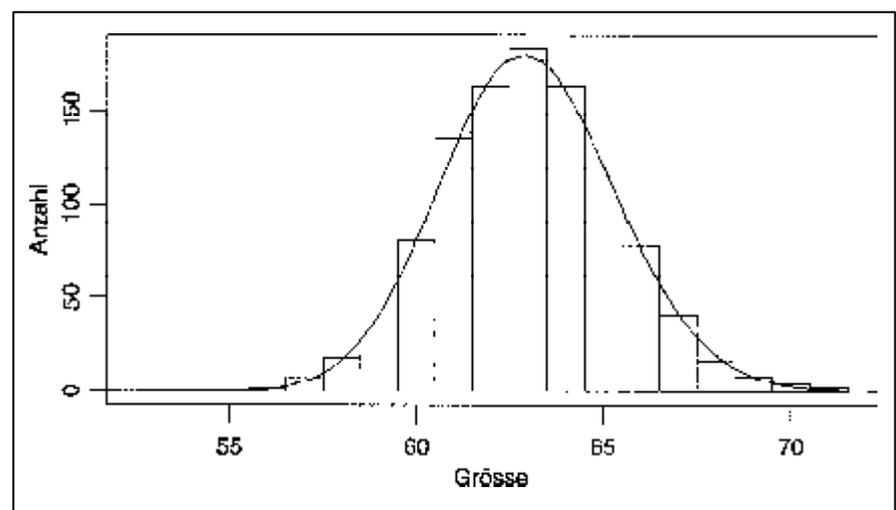
Von der Zeitschrift *BioScience* waren wir um Beispiele normalverteilter Original-Daten gebeten worden, die nicht ebenso gut zur Log-Normalverteilung passen. Erstaunlicherweise konnten wir bislang keine solchen Daten finden (LIMPERS et al., 2001) was auch weiterhin zutrifft. Auch das klassische Beispiel der Verteilung der Körpergröße, das in Lehrbüchern zur Demonstration der Normalverteilung hergenommen wird, passt ebenso gut zur Log-Normalverteilung. *Abbildung 5* bringt dies auf den Punkt, und bereits 1930 argumentiert CAMP, warum für die Körpergröße und die meisten weiteren anthropometrischen Messwerte die Log-Normalverteilung besser passt. Somit erscheint die Log-Normalverteilung als das vorherrschende, vereinheitlichende und ubiquitär zu findende Modell, das im Einklang ist mit den prinzipiell multiplikativen Funktionsweisen in der Natur und im Leben.

Die Normalverteilung ist vermutlich für die Anwendung auf Original-Daten in den Naturwissenschaften derzeit stark überbewertet. Auch ist man sich z.B. bei logarithmisch gegliederten Boniturskalen und normalverteilten Werten der zu Grunde liegenden Log-Normalverteilung nicht immer bewusst.

Das Standardwerk zur Genetik und Analyse quantitativer Merkmale (LYNCH und WALSH, 1998) baut ebenfalls auf

der Normalverteilung, dem additiven Zentralen Grenzwertsatz und additiven Effekten auf, und die Log-Normalverteilung wird schließlich eher als Frage des Maßstabs behandelt, mit dem man wieder zur Normalverteilung kommen kann. Von besonderem Interesse erscheint in dem Zusammenhang noch ein näherer Blick auf das grundlegende Kapitel 2 in dem insgesamt beeindruckenden Werk von LYNCH und WALSH (1998), wo es um die Eigenschaften von Verteilungen geht und um die Normalverteilung. Diese wird mit dem Gewicht von 1003 weiblichen Ratten sowie der Wurfgröße bei deren Nachkommen demonstriert (s. dort *Tabelle 2.1*; *Abbildung 2.1*).

Als Verteilung für die Wurfgröße kommt die Poissonverteilung in Frage. Da die Erwartungswerte klein sind, passt die Normalverteilung nicht, aber auch nicht die Log-Normalverteilung. Dreidimensionale Merkmale wie das Volumen oder logischerweise auch das Körpergewicht führen bereits nach KAPTEYN (1903) zu schiefen Verteilungen, zu denen nach AITCHISON und BROWN (1957, weitere Literatur s. dort) eine Log-Normalverteilung besser passt als die Normalverteilung. Unsere statistische Re-Evaluierung der genannten Körpergewichte führt dann auch zu p-Werten von 0,039 für die Normalverteilung und 0,09 für die empfohlene drei-parametrische Log-Normalverteilung. Zwar ist zu berücksichtigen, dass bei so großem Stichprobenumfang die Annahme einer Normalverteilung auch bei dem genannten p-Wert un-



**Abbildung 5: Eine multiplikative (oder Log-) Normalverteilung. Dargestellt ist die Körpergröße von 1052 Frauen (in inch), die zu der Log-Normalverteilung ( $62,48 \times 1,039$ ) mit einem p-Wert von 0,74 passt. Bislang wurde mit diesen Daten die Normalverteilung demonstriert (mit  $62,54 \pm 2,38$ ; SNEDECOR und COCHRAN, 1989), die mit einem p-Wert von 0,75 ebenso passt.**

ter 0,05 nicht unbedingt abgelehnt werden muss, aber offensichtlich passt die Log-Normalverteilung doch besser. Durch diese genauere Betrachtung werden also die Zweifel an der Normalverteilung als geeignetes Modell für die quantitative Genetik noch verstärkt.

Des Weiteren stellt sich auch angesichts der nachweisbaren multiplikativen Effekte in der Züchtungsforschung einerseits und der umfangreichen Literatur zur „additiven“ Genwirkung andererseits (ALLARD, 1960; FALCONER und McKAY, 1996; HENNIG, 2002; LYNCH und WALSH, 1998) jetzt die grundlegende Frage: Inwiefern können überhaupt additive Genwirkungen und Effekte nachgewiesen werden?

Die Antwort darauf wäre auch wichtig für die Sekundärliteratur, der - nebenbei gesagt - ja ebenfalls eine multiplikative Wirkung nachgesagt wird. Dort wird z.B. die additive Genwirkung als Begründung für die Bedeutung der Normalverteilung hergenommen (NEUHÄUSER, 2000) - ein *circulus vitiosus!* In besonderen Fällen, bei Gen-Duplikationen z.B., mögen ja additive Effekte vorkommen, aber auch da sollte wieder berücksichtigt werden, dass diese in den weiteren komplexen Stoffwechsel mit seinen multiplikativen Wirkungen eingebettet sind.

Beim Stoffwechsel kommt es auf Moleküle und deren Konzentrationen an und, wie eingangs gesagt, auf die Produkte der Konzentrationen. Deshalb sollte die Genwirkung in erster Näherung prinzipiell multiplikativ sein. Den Züchtungspraktiker schließlich interessieren vor allem Fragen der Selektion, wobei anzunehmen ist, dass sich durch dieses Konzept am bisherigen Vorgehen kaum etwas ändert. Denkbar ist allerdings, dass Prognosen über den Selektionsgewinn bei einigen Merkmalen verbessert werden können. Wir haben den Eindruck, Carl Friedrich GAUSS wäre nicht angetan von den vielen Versuchen in der Wissenschaft, seine symmetrische Kurve auf Daten anzuwenden, die im Prinzip asymmetrisch und log-normal sind. Der empfohlene Schritt vom normalen zum log-normalen Denken ist dabei ja nicht weit, wenn die „logarithmische Transitstrecke“ wieder über die so schön symmetrische Gaußsche Glockenkurve führt, wo wiederum die Additivität zu Grunde liegt (SCHNELL und COCKERHAM, 1992). Meldungen über sensationell große Früchte zur Erntezeit

wären dann zwar im Endeffekt weniger überraschend, wohl aber ebenso interessant und erfreulich.

Unter diesem und verschiedenen weiteren Aspekten erscheint jetzt eine Re-Evaluierung der umfassenden Literatur zur quantitativen Variation in der Züchtung, Züchtungsforschung und Genetik sehr sinnvoll. Auch wenn es diverse weitere Verteilungen von Interesse gibt, wird sich vermutlich die grundlegende Vorstellung bestätigen, dass das Leben und die Genwirkung *multiplikativ* sind, mit log-normaler Variation.

Wir danken COST in Bern und Brüssel für finanzielle Unterstützung, sowie H. F. UTZ, H. H. GEIGER, F.-M. GUMPERT, F. LAIDIG und G. FISCHBECK für wertvolle Diskussionen.

Unser Beitrag ist Herrn Professor F. W. SCHNELL zum 90. Geburtstag gewidmet.

## Literatur

- AITCHISON J., J.A.C. BROWN, 1957: The Log-normal Distribution. Cambridge University Press.
- ALLARD, R.W., 1960: Principles of Plant Breeding. Wiley, London.
- BAUR, P., 1997: Lognormal distribution of water permeability and organic solute mobility in plant cuticles. *Plant, Cell and Environment* 20: 167-177.
- BECKER, H.C., 1984: Theoretische Überlegungen und experimentelle Untersuchungen zur genetischen Basis der Heterosis - eine Literaturübersicht. *Vortr. Pflanzenzücht.* 5: 23-42.
- COCHRAN, W.G., 1980: Fisher and the analysis of variance. In: Fienberg, S.E., Hinkley D. V. (eds.), R.A. Fisher: An Appreciation. Springer Berlin, 17-34.
- CAMP, B.H., 1930: Notes on the distribution of geometric mean. *Ann. Math. Statist.* 9, 221.
- COCKERHAM, C.C., 1959: Partitions of hereditary variance for various genetic models. *Genetics* 44: 1141-1148.
- DEISING, H.B., S. REIMANN, A. PEIL and W.E. WEBER, 2002: Disease management of rusts and powdery mildews. In Kempken, F. (ed.) *The Mycota XI. Application in Agriculture*. Springer, Berlin, 243-269.
- DE MOIVRE, A., 1738: The doctrine of chances. 2nd Ed. (reprinted 1967, Frank Cass, London).
- DICKERSON, G.E., 1955: Genetic slippage in response to selection for multiple objectives. *Cold Spring Harbour Symp. Quant. Biol.* 20, 213-224.
- EAST, E.M., 1913: Inheritance of flower size in crosses between species of *Nicotiana*. *Bot. Gaz.*, 55: 177-188.
- FALCONER, D.S., T.F.C. MCKAY, 1996: Introduction to quantitative genetics. 4th Ed. Longman Sci. and Tech., Harlow, UK.
- FECHNER, G.T., 1860: Elemente der Psychophysik: Leipzig: Breitkopf und Härtel.
- FISHER, R.A., 1918: The correlation between relatives on the supposition of Medelian inheritance. *Trans Roy Soc Edinburgh* 52, 399-433.
- FISHER, R.A. and W.A. MACKENZIE, 1923: Studies in crop variation. II. The manurial response of different potato varieties. *J. Agric. Sci.* 13, 311-320.
- GALTON, F., 1879: The geometric mean, in vital and social statistics. *Proceedings of the Royal Society* 29: 365-367.
- GEIGER, H.H., 1988: Epistasis and Heterosis. In: Weir S, Eisen EJ, Goodman MM, Namkoong G (eds) *Proc 2nd Intl Conf on Quantitative Genetics*. Sinauer Assoc, Inc, Pbl Sunderland, MA, USA, 395-399.
- GEIGER, H.H. and G. WAHLE, 1978: Struktur der Heterosis von Komplexmerkmalen bei Winterroggen-Einfachhybriden. *Z. Pflanzenzücht.* 80: 198-210.
- HENNIG, W., 2002: Genetik. 3. Aufl. Springer, Berlin.
- KOCH, A.L., 1966: The logarithm in biology. I. Mechanisms generating the log-normal distribution exactly. *Journal of Theoretical Biology* 23: 276-290.
- KAPTEYN, J.C., 1903. *Skew Frequency Curves in Biology and Statistics*. Astronomical Laboratory, Groningen: Noordhoff.
- LIMPERT, E., W.A. STAHEL and M. ABBT, 2001: Log-normal distributions across the sciences - keys and clues. *BioScience* 51, 341-352.
- LYMAN, R.F., F. LAWRENCE, S.V. NUZHIDIN and T.F.C. MCKAY, 1996: Effects of single P element insertions on bristle number and viability in *Drosophila melanogaster*. *Genetics* 143, 277-292.
- LYNCH, M. and B. WALSH, 1998: Genetics and analysis of quantitative traits. Sinauer, Sunderland, Mass.
- MACARTHUR, J.W., and L. BUTLER, 1938: Size inheritance and geometric growth processes in the tomato fruit. *Genetics*, 23: 253-268.
- MCALISTER, D., 1879: The law of the geometric mean. *Proceedings of the Royal Society* 29: 367-376.
- NEUHÄUSER, M., 2000: Statistik formelfrei. Fischer, Aachen, 79pp.
- PALEVITCH, D., and N. KEDAR, 1970: Resistance of tomato cultivars and progenies to fruit hollowiness. *Euphytica*, 19: 253-260.
- RICHEY, F.D., 1942: Mock-dominance and hybrid vigor. *Science* 96, 280-281.
- SCHNELL, F.W. and C.C. COCKERHAM, 1992: Multiplicative vs. arbitrary gene action in heterosis. *Genetics* 131: 461-469.
- STEHMANN, C. and M.A. De WAARD, 1996: Sensitivity of populations of *Botrytis cinerea* to triazoles, benomyl and vinclozolin. *European Journal of Plant Pathology* 102: 171-180.
- WILLIAMS, W., 1959: Heterosis and the genetics of complex characters. *Nature* 184, 527-530.
- WRIGHT, S., 1922: The effects of inbreeding and crossbreeding on guinea pigs. III. Crosses between highly inbred families. *Tech Bull US Dept Agric* 1121, 60 pp.

