

Der Satz von de Moivre / Laplace als Aussage über Binomialkoeffizienten

JÖRG MEYER, HAMELN

Zusammenfassung: Bekanntlich gibt es einen engen Zusammenhang zwischen den standardisierten Binomialverteilungen einerseits und der Normalverteilung, insbesondere mit dem Term $e^{-x^2/2}$, andererseits. Hier wird deutlich gemacht, dass sich dieser Zusammenhang schon in einem elementarisierten und stochastik-freien Kontext verstehen lässt. Insbesondere spielt der fragliche Term bei einer Aussage über Binomialkoeffizienten eine Rolle.

1 Einleitung

Der Satz von de Moivre / Laplace und dessen Beweis sind gar nicht so einfach zu verstehen. Es handelt sich um Binomialverteilungen mit der Zufallsvariablen B_n (= Anzahl der Erfolge bei n Versuchen, die jeweils die Erfolgswahrscheinlichkeit p haben sollen) mit

$$\text{prob}(B_n = k) = \binom{n}{k} \cdot p^k \cdot q^{n-k}$$

(wie üblich ist $q = 1 - p$). Man muss die Zufallsvariable standardisieren (vgl. etwa Henze [2000; S. 208]); dadurch bekommt man die Zuordnung

$$\frac{k - \mu_n}{\sigma_n} \mapsto \sigma_n \cdot \binom{n}{k} \cdot p^k \cdot q^{n-k} \quad (1)$$

bzw.

$$x = \frac{k - \mu_n}{\sigma_n} \mapsto \sigma_n \cdot \binom{n}{n \cdot p + \sigma_n \cdot x} \cdot p^{n \cdot p + \sigma_n \cdot x} \cdot q^{n \cdot q - \sigma_n \cdot x} \quad (2)$$

$$=: \varphi_n(x);$$

dabei ist $\mu_n = n \cdot p$ der Erwartungswert und $\sigma_n = \sqrt{n \cdot p \cdot q}$ die Standardabweichung von B_n .

(Im Folgenden wird immer angenommen, dass $n \cdot p + \sigma_n \cdot x$ ganzzahlig ist, damit der Binomialkoeffizient $\binom{n}{n \cdot p + \sigma_n \cdot x}$ sinnvoll bleibt. Hieraus folgt allerdings, dass x in (2) nicht konstant sein kann, sondern für jedes n einen etwas anderen Wert haben wird. Alternativ könnte man den Binomial-

koeffizienten ersetzen durch $\binom{n}{[n \cdot p + \sigma_n \cdot x]}$. Dieser Weg wird hier nicht eingeschlagen.)

Schon die mit (1) und (2) verbundenen Termstrukturen sind für Schüler unübersichtlich. Zum Beweis der Tatsache, dass die durch (1) beschriebenen standardisierten Binomialverteilungen gegen die Normalverteilung konvergieren, muss (1) bzw. (2) geeignet manipuliert werden; kein Wunder, dass hier der Überblick leicht verloren geht.

Aber selbst wenn man den Überblick wieder erlangt hat: Was ist eigentlich das Wesentliche am Beweis (und an der Aussage) des Satzes? Hat das unerwartete Auftreten des Terms $e^{-x^2/2}$ mit der Standardisierung zu tun, oder woher kommt er „eigentlich“?

Hier wäre es wünschenswert, wenn man den Beweis des Satzes von de Moivre / Laplace in kleinere Moduln zerlegen könnte, die möglichst für sich sinnvoll sind und die einzeln verstanden werden können.

Die Aussage des (lokalen) Grenzwertsatzes von de Moivre / Laplace besteht aus drei Teilen:

1. Die Tatsache der Konvergenz der standardisierten Binomialverteilungen. Diese Tatsache ist für Schüler offensichtlich, kann leicht visualisiert werden und wird auf der Schule daher nicht bewiesen.

2. Die Gestalt der Grenzfunktion

$$\varphi(x) = \varphi(0) \cdot e^{-x^2/2}.$$

3. Die Normierung der Grenzfunktion so, dass

$$\int_{-\infty}^{\infty} \varphi(x) \cdot dx = 1 \quad \text{ist, mit dem Ergebnis}$$

$$\varphi(0) = \frac{1}{\sqrt{2 \cdot \pi}}.$$

Hierbei handelt es sich um ein rein analytisches Problem.

Hier soll es allein um Teil 2 gehen (Ausführungen zu den Teilen 1 und 3 findet sich in Meyer [2004]). Dass auch Teil 2 modularisiert und elementarisiert werden kann und wie das geht, ist Inhalt dieses Aufsatzes.

Der Kern der Aussage von Teil 2 wird in einem Kontext erarbeitet, der mit Stochastik gar nichts zu

tun hat. Es wird deutlich werden, dass es sich dabei „nur“ um eine Aussage über Binomialkoeffizienten handelt. Diese Aussage lässt sich dann sehr leicht in den stochastischen Kontext der Binomialverteilungen übertragen.

Vorab ist es sinnvoll, ein analytisches Lemma bereit zu stellen, das auch für andere stochastische Anwendungen heran gezogen werden kann.

$$\begin{aligned} \ln\left(1 + \frac{A}{N} + \frac{B}{N^2}\right)^N &= N \cdot \left(\left(\frac{A}{N} + \frac{B}{N^2}\right) - \frac{\left(\frac{A}{N} + \frac{B}{N^2}\right)^2}{2} + \sum_{i=3}^{\infty} \frac{(-1)^{i+1}}{i} \cdot \left(\frac{A}{N} + \frac{B}{N^2}\right)^i \right) \\ &= \left(A + \frac{B}{N}\right) - \frac{1}{N} \cdot \frac{\left(A + \frac{B}{N}\right)^2}{2} + \frac{1}{N^2} \cdot \sum_{i=3}^{\infty} \frac{1}{N^{i-3}} \cdot \frac{(-1)^{i+1}}{i} \cdot \left(A + \frac{B}{N}\right)^i \end{aligned}$$

Die Reihe $\sum_{i=3}^{\infty} \frac{1}{N^{i-3}} \cdot \frac{(-1)^{i+1}}{i} \cdot \left(A + \frac{B}{N}\right)^i$ lässt sich durch geometrische Reihen leicht abschätzen und ist daher beschränkt. Lemma 1 folgt nun durch Grenzübergang. (Statt den Beweis mit Hilfe der Mercator-Reihe zu führen, kann man auch die Abschätzung $1+x \leq e^x \leq \frac{1}{1-x}$ bzw. $1 - \frac{1}{x} \leq \ln x \leq x - 1$ verwenden.) –

$$\begin{aligned} \left(\prod_{i=1}^{M-1} \left(1 + \frac{i}{N^2}\right)\right)^2 &= \underbrace{\left(1 + \frac{1}{N^2}\right) \cdot \left(1 + \frac{M-1}{N^2}\right)}_{=1 + \frac{M}{N^2} + \frac{1 \cdot (M-1)}{N^4}} \cdot \underbrace{\left(1 + \frac{2}{N^2}\right) \cdot \left(1 + \frac{M-2}{N^2}\right)}_{=1 + \frac{M}{N^2} + \frac{2 \cdot (M-2)}{N^4}} \cdot \dots \cdot \underbrace{\left(1 + \frac{M-1}{N^2}\right) \cdot \left(1 + \frac{1}{N^2}\right)}_{=1 + \frac{M}{N^2} + \frac{(M-1) \cdot 1}{N^4}} \\ &= \prod_{i=1}^{M-1} \left(1 + \frac{M}{N^2} + \frac{i \cdot (M-i)}{N^4}\right) \end{aligned}$$

Nun liegt $i \cdot (M-i)$ stets zwischen 0 und $\frac{M^2}{4}$, also gilt

$$\begin{aligned} \left(1 + \frac{A}{N}\right)^{A \cdot N-1} &\leq \prod_{i=1}^{M-1} \left(1 + \frac{M}{N^2} + \frac{i \cdot (M-i)}{N^4}\right) \\ &\leq \left(1 + \frac{A}{N} + \frac{A^2}{4 \cdot N^2}\right)^{A \cdot N-1} \end{aligned}$$

Wegen Lemma 1 konvergieren die linke und die rechte Schranke beide gegen e^{A^2} , woraus die Behauptung folgt. Die zweite Aussage beweist man analog. (Eine alternative Begründung liefert Henze [32000; S. 72 f.] im Beweis seines Satzes 10.1. Er benutzt nicht Lemma 1, sondern die für die Schule nicht so naheliegende Abschätzung

2 Einige analytische Lemmata

Aus dem Schulunterricht sind die folgenden Resultate nicht vertraut, obgleich die Beweise nicht schwierig sind:

Lemma 1: $\lim_{N \rightarrow \infty} \left(1 + \frac{A}{N} + \frac{B}{N^2}\right)^N = e^A$

Beweis: Mit Hilfe der Mercator-Reihe folgt:

Lemma 2: $\lim_{N \rightarrow \infty} \prod_{i=1}^{A \cdot N} \left(1 \pm \frac{i}{N^2}\right) = e^{\pm A^2/2}$

Beweis: Es sei $M := A \cdot N$; wir führen den Beweis für „+“. Die „Methode des kleinen Gauß“, die natürlichen Zahlen von 1 bis 100 zu summieren, lässt sich – leicht modifiziert – auch hier anwenden. Wir machen im Limes keinen Fehler, wenn wir statt M Faktoren nur $M-1$ Faktoren betrachten; dann ist

$$\exp\left(\frac{x}{1+x}\right) \leq 1+x \text{ .) —}$$

Didaktische Bemerkung: Die Überlegungen zu Lemma 1 und Lemma 2 sind näher an den eigentlichen Zielen des Analysis-Unterrichts als durchorganisierte und schematisierte Kurvendiskussionen. Dies zeigt auch das unterrichtspraktische Dilemma: Gerade weil sich die Überlegungen kaum schematisieren lassen, sind sie für Lernende schwierig. Aber in der Schwierigkeit liegt auch der Gewinn!

3 Exkurs: Lemma 2 und das Geburtstagsproblem

Wir haben eine Urne mit N verschiedenen und unterscheidbaren Kugeln und ziehen n -mal MIT Zurücklegen. Dann gilt:

prob(alle n Kugeln verschieden)

$$= \frac{N}{N} \cdot \frac{N-1}{N} \cdots \frac{N-n+1}{N} = \prod_{i=1}^{n-1} \left(1 - \frac{i}{N}\right).$$

Für jedes N suchen wir n so, dass

$$\prod_{i=1}^{n-1} \left(1 - \frac{i}{N}\right) \approx \frac{1}{2} \quad (3)$$

ist. Beim Geburtstagsproblem ist $N = 365$; das Ergebnis ist $n = 23$. Wie kommt man ohne Ausprobieren auf diesen Wert für n?

Aufgrund von Lemma 2 ist

$$\lim_{N \rightarrow \infty} \prod_{i=1}^{A \cdot \sqrt{N}} \left(1 - \frac{i}{N}\right) = e^{-A^2/2}.$$

Für große N sollte daher $e^{-A^2/2} = \frac{1}{2}$ bzw. $A^2 =$

$\ln 4$ sein, woraus sich $n \approx \sqrt{N \cdot \ln 4}$ ergibt.

Beim Geburtstagsproblem ist $\sqrt{365 \cdot \ln 4} \approx 22,5$, also ein erstaunlich gutes Ergebnis.

Man kann das Geburtstagsproblem als eine wesentliche Anwendung von Lemma 2 auffassen, aber auch als Motivation, sich überhaupt mit einem solchen Produkt wie in (3) zu befassen. Damit sind zwei Möglichkeiten angegeben, Lemma 2 zu unterrichten: Als innermathematische Themensequenz oder aber ausgehend von einer Anwendung wie dem Geburtstagsproblem.

4 Das Verhalten von Binomialkoeffizienten

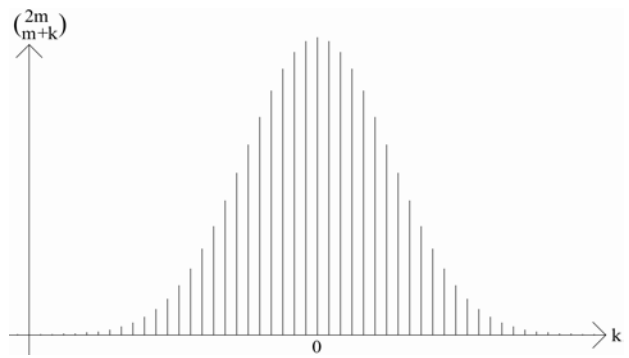
Wir beschäftigen uns zunächst gar nicht mit der Konvergenz der standardisierten Binomialverteilung, sondern mit einem einfacheren und durchsichtigeren Problem, dessen Lösung gleichwohl schon die Lösung des allgemeineren Problems ist.

Das einfachere Problem lautet: Wie verhalten sich Binomialkoeffizienten für große obere Argumente?

Es ist methodisch von Vorteil, dass dieses einfachere Problem keinerlei Standardisierung, ja sogar keinerlei Kenntnis über Binomialverteilungen voraussetzt.

Nun wird für gerade $n = 2 \cdot m$ der größte Wert der Binomialkoeffizienten $\binom{2 \cdot m}{x}$ bei $x = m$ angenommen, so dass es vernünftig ist, die Binomialkoeffizienten als $\binom{2 \cdot m}{m+k}$ zu schreiben.

Die folgende Illustration gibt für $m = 80$ einen ersten Eindruck:



Nun ist es eine glückliche Idee, die Binomialkoeffizienten $\binom{2 \cdot m}{m+k}$ gar nicht direkt zu betrachten, sondern

stattdessen die auf den „Zentralwert“ $\binom{2 \cdot m}{m}$

bezogenen Quotienten $\frac{\binom{2 \cdot m}{m+k}}{\binom{2 \cdot m}{m}}$. Da Zähler und

Nenner ein gemeinsames oberes Argument haben, sind solche Quotienten rechentechnisch gut zugänglich. Wir interessieren uns also für die Größe der Binomialkoeffizienten $\binom{2 \cdot m}{m+k}$ für große m. Eine einfache Rechnung liefert

$$\begin{aligned} \frac{\binom{2 \cdot m}{m+k}}{\binom{2 \cdot m}{m}} &= \frac{m! \cdot m!}{(m+k)! \cdot (m-k)!} \\ &= \frac{m \cdot (m-1) \cdots (m-k+1)}{(m+1) \cdots (m+k)} \\ &= \frac{\prod_{i=0}^{k-1} \left(1 - \frac{i}{m}\right)}{\prod_{i=1}^k \left(1 + \frac{i}{m}\right)}. \end{aligned}$$

Für große m ist das obere Produkt $\prod_{i=0}^{k-1} \left(1 - \frac{i}{m}\right)$ et-

wa so groß wie $\prod_{i=0}^k \left(1 - \frac{i}{m}\right)$ (in dem Sinne, dass der Quotient beider Produkte gegen 1 geht), so dass

$$\frac{\binom{2 \cdot m}{m+k}}{\binom{2 \cdot m}{m}} \approx \frac{\prod_{i=1}^k \left(1 - \frac{i}{m}\right)}{\prod_{i=1}^k \left(1 + \frac{i}{m}\right)}$$

gilt. Wie soll man die beiden Produkte der rechten Seite weiter verarbeiten?

Über ähnliche Produkte wurde in Lemma 2 eine Aussage gemacht. Daher drehen wir den Spieß um und richten die Produkte so ein, dass die Beziehung

$$\lim_{N \rightarrow \infty} \prod_{i=1}^{A \cdot N} \left(1 + \frac{i}{N^2}\right) = e^{A^2/2}$$

bzw.

$$\lim_{N \rightarrow \infty} \prod_{i=1}^{A \cdot N} \left(1 - \frac{i}{N^2}\right) = e^{-A^2/2}$$

anwendbar ist. Dazu muss k von der Gestalt $k = x \cdot \sqrt{m}$ sein; alsdann bekommt man

$$\frac{\binom{2 \cdot m}{m + x \cdot \sqrt{m}}}{\binom{2 \cdot m}{m}} \approx \frac{\prod_{i=1}^{x \cdot \sqrt{m}} \left(1 - \frac{i}{m}\right)}{\prod_{i=1}^{x \cdot \sqrt{m}} \left(1 + \frac{i}{m}\right)} \xrightarrow{m \rightarrow \infty} \frac{e^{-x^2/2}}{e^{x^2/2}} = e^{-x^2}$$

Dies ist eine Aussage über die Größe der Binomialkoeffizienten: Für große m ist demnach

$$\binom{2 \cdot m}{m + k} \approx \binom{2 \cdot m}{m} \cdot e^{-k^2/m}. \quad (4)$$

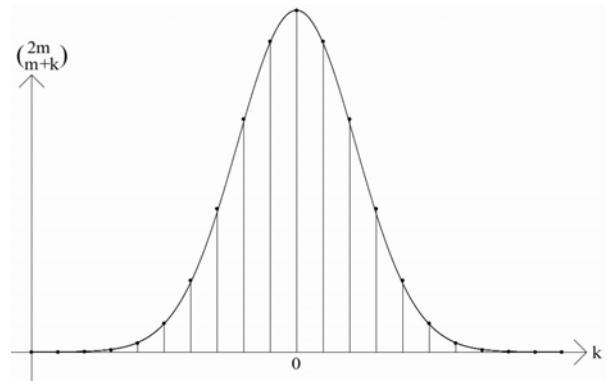
Es ist sehr bemerkenswert, dass der für die Normalverteilung wesentliche Term $e^{-x^2/2}$ im Wesentlichen der Limes eines Quotienten von Binomialkoeffizienten ist! Insofern macht der Satz von de Moivre/ Laplace „nur“ eine Aussage über Binomialkoeffizienten.

Man sollte die Formel (4) am besten „statisch“ auffassen, und zwar als Ungefähr-Aussage für große m und unterschiedliche Werte von k . In diesem Sinne gibt sie auch Aufschluss über die „Kurve der Binomialkoeffizienten“ in Abhängigkeit von k .

Würde man (4) „dynamisch“ auffassen, so wäre m die entscheidende Variable und nicht k . Zudem ergeben sich Interpretationsschwierigkeiten, denn k hängt (wegen $k = x \cdot \sqrt{m}$) von m ab und x ist so einzurichten, dass $x \cdot \sqrt{m}$ ganzzahlig ist.

Hier für $m=10$ eine „optische Illustration“ von (4); die Striche mit den dicken Endpunkten bezeichnen die Länge der Binomialkoeffizienten $\binom{2 \cdot m}{m + k}$, und die Kurve gibt die Approximation

$$\binom{2 \cdot m}{m} \cdot e^{-k^2/m} \text{ an:}$$



Damit hat man eine auf den Zentralwert $\binom{2 \cdot m}{m}$ bezogene gute Übersicht über das Verhalten von $\binom{2 \cdot m}{m + k}$.

Im Schulunterricht habe ich gute Erfahrungen damit gemacht, zunächst diese einfachere Version des Satzes von de Moivre / Laplace zu behandeln, die ja mit Stochastik noch gar nichts zu tun hat und für sich allein hinreichend interessant ist. Wenn dann etwas später die standardisierte Binomialverteilung behandelt wird, so liefert eine leichte Modifikation das anzustrebende Resultat. Eigentlich muss nur noch alter Wein in neue Schläuche gefüllt werden. Dabei erweisen sich die neuen Schläuche für manche Lernenden schon als etwas unübersichtlich, so dass nicht auch noch neuer Wein hinzukommen muss.

5 Quotienten der standardisierten Binomialverteilung für $p = 0,5$

Ziel unserer Überlegungen ist das Grenzwertverhalten von

$$\varphi_n(x) = \sigma_n \cdot \binom{n}{n \cdot p + \sigma_n \cdot x} \cdot p^{n \cdot p + \sigma_n \cdot x} \cdot q^{n \cdot q - \sigma_n \cdot x}$$

für große n bei ganzzahligen $n \cdot p + \sigma_n \cdot x$ (vgl. (2) in Abschnitt 1).

Beginnen wir mit $p = \frac{1}{2}$; hier lässt sich die Vorgehensweise des letzten Abschnitts fast wörtlich übertragen:

Es handelt sich um das Grenzwertverhalten von $\varphi_n(x) = \frac{\sqrt{n}}{2} \cdot \binom{n}{(n + \sqrt{n} \cdot x)/2} \cdot \left(\frac{1}{2}\right)^n$ für große n bei ganzzahligen Werten von $(n + \sqrt{n} \cdot x)/2$.

Wir berechnen $\frac{\varphi_n(x)}{\varphi_n(0)}$, wobei $m := \frac{n}{2}$ ganzzahlig sein muss, damit der Nenner definiert ist. Da auch $(n + \sqrt{n} \cdot x)/2$ ganzzahlig sein muss, ist

$$g := x \cdot \frac{\sqrt{n}}{2} = \frac{x}{\sqrt{2}} \cdot \sqrt{m} \text{ ebenfalls ganzzahlig.}$$

Für $x > 0$ ist (für $x < 0$ kann man analog argumentieren)

$$\frac{\varphi_n(x)}{\varphi_n(0)} = \frac{\binom{2 \cdot m}{m+g}}{\binom{2 \cdot m}{m}} \approx \frac{\prod_{i=1}^{\frac{x}{\sqrt{2}} \cdot \sqrt{m}} \left(1 - \frac{i}{m}\right)}{\prod_{i=1}^{\frac{x}{\sqrt{2}} \cdot \sqrt{m}} \left(1 + \frac{i}{m}\right)}$$

$$\xrightarrow{m \rightarrow \infty} \frac{e^{-x^2/4}}{e^{x^2/4}} = e^{-x^2/2}$$

wegen $\lim_{N \rightarrow \infty} \prod_{i=1}^{A \cdot \sqrt{N}} \left(1 \pm \frac{i}{N}\right) = e^{\pm A^2/2}$.

Leider bekommt man auf diese Art nur eine Aussage über gerade Indizes $2 \cdot m$. Allerdings liefert die dargestellte Argumentation den wesentlichen Grund dafür, dass $\varphi(x) = \varphi(0) \cdot e^{-x^2/2}$ ist, und es ist „anschaulich evident“, dass die zu ungeraden Indizes gehörigen Folgenglieder $\frac{\varphi_{2 \cdot m+1}(x)}{\varphi_{2 \cdot m+1}(0)}$ die Konvergenz nicht stören werden.

6 Quotienten der standardisierten Binomialverteilung für allgemeines p

Für $p \neq \frac{1}{2}$ hat man nur unwesentlich mehr Arbeit:

Es geht um

$$\varphi_n(x) = \sigma_n \cdot \binom{n}{n \cdot p + \sigma_n \cdot x} \cdot p^{n \cdot p + \sigma_n \cdot x} \cdot q^{n \cdot q - \sigma_n \cdot x},$$

und wir berechnen $\frac{\varphi_n(x)}{\varphi_n(0)}$, wobei $n \cdot p$ ganzzahlig sein muss, damit der Nenner definiert ist. Da auch $n \cdot p + \sigma_n \cdot x$ ganzzahlig sein muss, ist $g := \sigma_n \cdot x$ ebenfalls ganzzahlig.

Für $x > 0$ ist

$$\frac{\varphi_n(x)}{\varphi_n(0)} = \frac{\binom{n}{n \cdot p + g}}{\binom{n}{n \cdot p}} \cdot \left(\frac{p}{q}\right)^g$$

$$= \frac{(n \cdot p)! \cdot (n \cdot q)!}{(n \cdot p + g)! \cdot (n \cdot q - g)!} \cdot \left(\frac{p}{q}\right)^g$$

$$= \frac{\prod_{i=0}^{g-1} (n \cdot q - i)}{\prod_{i=1}^g (n \cdot p + i)} \cdot \left(\frac{p}{q}\right)^g = \prod_{i=0}^{g-1} \left(1 - \frac{i}{n \cdot q}\right).$$

Für große n gilt

$$\prod_{i=0}^{g-1} \left(1 - \frac{i}{n \cdot q}\right) \approx \prod_{i=1}^g \left(1 - \frac{i}{n \cdot q}\right),$$

so dass man unter Verwendung von Lemma 2 die angestrebte Grenzaussage

$$\frac{\varphi_n(x)}{\varphi_n(0)} \approx \frac{\prod_{i=1}^g \left(1 - \frac{i}{n \cdot q}\right)}{\prod_{i=1}^g \left(1 + \frac{i}{n \cdot p}\right)} = \frac{x \cdot \sqrt{p \cdot \sqrt{n \cdot q}} \prod_{i=1}^g \left(1 - \frac{i}{n \cdot q}\right)}{x \cdot \sqrt{q \cdot \sqrt{n \cdot p}} \prod_{i=1}^g \left(1 + \frac{i}{n \cdot p}\right)}$$

$$\xrightarrow{n \rightarrow \infty} \frac{e^{-p \cdot x^2/2}}{e^{q \cdot x^2/2}} = e^{-x^2/2}$$

bekommt.

Hald [1998; S. 23] schreibt über diesen Beweis, der im Wesentlichen von Simpson (1740) stammt: „This is the first explicit proof of the approximation in the general case, and it is indeed noteworthy for its simplicity.“ Hätte Simpson so, wie es hier geschehen ist, vorab Lemma 2 isoliert, wäre sein Beweis noch einfacher geworden.

Literatur

Hald, A. [1998]: A history of mathematical statistics from 1750 to 1930. New York usw.: John Wiley.
 Henze, N. [2000]: Stochastik für Einsteiger. Braunschweig usw.: Vieweg Verlag (Original 1997).

Meyer, J. [2004]: Schulnahe Beweise zum zentralen Grenzwertsatz. Hildesheim: Franzbecker Verlag.

Autor

Jörg Meyer
 Schäfertrift 16
 31789 Hameln
 Email: J.M.Meyer@t-online.de