Die Faustregel $np(1-p) \gtrapprox 9$

Herbert Ziezold, Kassel

Zusammenfassung: *Angeregt durch die Arbeit Eichelsbacher [1] wird die didaktische und anwendungsrelevante Bedeutung von Faustregeln zur Beurteilung der Approximierbarkeit von Binomialverteilungen durch Normal- und Poissonverteilungen begründet.*

Schon seit annähernd drei Jahrhunderten haben Mathematiker versucht, eine praktisch verwertbare Fehlerabschätzung für die Differenz zwischen Binomialwahrscheinlichkeiten und ihren numerischen Schätzungen über die (Standard-)Normalverteilung zu finden. Meines Wissens bis heute ohne befriedigenden Erfolg.

In Eichelsbacher [1] hat der Autor einige bedeutsame Stationen dieser Forschung aufgezeigt und damit die Problematik dieser Materie dargelegt. Insbesondere der berühmte Satz von Berry und Esseen ist es wert, von Lehrern in seiner Aussage verarbeitet zu werden, gibt er doch eine bestmögliche, also scharfe, allgemein gültige Abschätzung des Supremums der Differenz zwischen kumulativen Verteilungsfunktionen von standardisierten Summen unabhängiger, identisch verteilter Zufallsvariablen (mit existierendem 3. Moment) und der kumulativen Verteilungsfunktion Φ der Standardnormalverteilung, kurzum der *Supremumsnorm* der Differenz dieser Verteilungsfunktionen.

Angewandt auf die Summe S_n von n unabhängigen Indikatorvariablen X_i mit $P\{X_i = 1\} = p$, $i = 1, 2, \ldots, n$, S_n ist dann definitionsgemäß binomialverteilt mit den Parametern n und p, kurz: $B(n,p)$-verteilt, ergibt sich für die standardisierte Zufallsvariable

$$S_n^* = \frac{S_n - np}{\sqrt{np(1-p)}}$$

die Abschätzung

$$\sup_{x\in\mathbb{R}} |P\{S_n^* \le x\} - \Phi(x)| \le \frac{p^2 + (1-p)^2}{\sqrt{2\pi np(1-p)}}.$$

(Hier ist die von Eichelsbacher [1] angegebene Konstante $\frac{1}{\sqrt{2\pi}}$ statt des von Berry und Esseen ermittelten Werts 6 eingesetzt worden.)

Dies vermittelt insbesondere eine präzise, theoretische Vorstellung von der Konvergenzgeschwindigkeit der Approximation von Binomialverteilungen durch die Normalverteilung und zeigt in aller Deutlichkeit, dass die Güte dieser Approximation in erster Linie von der Standardabweichung $\sigma = \sqrt{np(1-p)}$ der betreffenden $B(n,p)$-Verteilung abhängt.

Wie die Beispiele 1 bis 3 in Eichelsbacher [1] aber zeigen, ist diese Abschätzung für praktische Zwecke zu grob, um nicht zu sagen unbrauchbar. Dies liegt in erster Linie daran, dass der Satz von Berry-Esseen nur eine Aussage über das *Supremum* der Differenz der Verteilungsfunktionen liefert, unabhängig von der gerade interessierenden Stelle einer $B(n,p)$-Verteilung.

So kann man den in Beispiel 3 ausgerechneten maximalen Fehler von 0.026 noch als leidlich klein bezeichnen, wenn er sich auf die Schätzung einer Wahrscheinlichkeit von ca. 0.5 bezieht. Ist aber, wie in diesem Beispiel, von einer Wahrscheinlichkeit von ca. 0.025 die Rede, so ist die Fehlerangabe 0.026 gewiss nicht zufriedenstellend. – Exakte Rechnung zeigt übrigens, dass in diesem Beispiel für $n = 235$ der Fehler etwa 0.0029 und für $n = 236$ etwa 0.0031 ist. Die wahren Fehler liegen also deutlich unter ihrer berechneten oberen Schranke 0.026. (Sollte der Leser kein Programm zur exakten Berechnung von Binomialverteilungen zur Verfügung haben, so lade ich ihn ein, meine Homepage `www.mathematik.uni-kassel.de/~ziezold` zu besuchen und sich von dort ein solches herunter zu laden.)

Auch die folgende Abwandlung der Aufgabe des Beispiels 1 aus Eichelsbacher [1] illustriert, wie grob die Berry-Esseen-Abschätzung sein kann: Eine Fabrik stellt ein Werkstück mit einer Ausschussrate von 10 % her. Wie groß muss n sein, damit bei S_n beobachteten defekten Werkstücken unter n produzierten Artikeln die Schätzung der Wahrscheinlichkeit $P\{S_n/n > 1/8\}$ durch die Normalverteilung höchstens den Fehler 0.005 hat? Im Fall der besagten Arbeit ergibt sich für $n = 400$ der exakte Wert 0.0436 und der approximative Wert 0.0478, also ein tatsächlicher Fehler von 0.0042, der bereits unter der gewünschten Schranke 0.005 liegt, während der Satz von Berry-Esseen den Fehler durch 0.066 nach oben abschätzt. Unter den Voraussetzungen der gestellten Aufgabe müsste nach Berry-Esseen

$$\frac{0.1^2 + 0.9^2}{\sqrt{2\pi n\, 0.1 \cdot 0.9}} \le 0.005$$

gelten, woraus $n \geq 47563$ folgte, also fast das 119fache der obigen Beobachtungsanzahl $n = 400$, für die wir durch exakte Rechnung gerade festgestellt hatten, dass der wahre Fehler bereits unter 0.005 liegt. Und diese Fehlerabschätzung beträgt sogar mehr als 10 % der geschätzten Wahrscheinlichkeit von etwa 0.048, ist also für ein $n \geq 47563$ als unbefriedigend grob anzusehen.

Damit sollte diese Art der Fehleranalyse als praktisch so gut wie nicht brauchbar erkannt sein.

Wie steht es nun um die "Faustregel", dann die Approximation einer $B(n, p)$-Verteilung durch die Normalverteilung als zulässig anzusehen, wenn die Ungleichung $np(1-p) > 9$ erfüllt ist, zuweilen *Laplace-Regel* genannt?

Zunächst einmal ist festzuhalten, dass der Name "Faustregel" schon andeutet, dass es sich nicht um eine mathematisch einwandfrei begründete, notwendige oder hinreichende Bedingung für die Approximation handelt. Hier wird schlichtweg "über den Daumen gepeilt". In Ermangelung einer mathematisch verlässlichen und zugleich praktisch brauchbaren Alternative, ist man darauf angewiesen, Schülern und Anwendern wenigstens einen ungefähren Anhaltspunkt zu liefern, wann sie halbwegs gute Approximationen erwarten können.

Dass hier eine Bedingung über die Größe der Standardabweichung der $B(n, p)$-Verteilung vernünftig ist, ist spätestens seit den Berechnungen von Laplace klar, nach denen die Größenordnung des Approximationsfehlers im Wesentlichen ein Vielfaches von $1/\sqrt{np(1-p)}$ ist. (Siehe Eichelsbacher [1].) Die Zahl 9 ist selbstverständlich willkürlich gewählt, hängt die Wahl einer solchen Konstanten doch von der gewünschten Genauigkeit ab. Als Quadratzahl hat sie den Vorzug vor zum Beispiel 8 oder 10; denn $np(1-p) > 9$ ist gleichbedeutend mit $\sigma = \sqrt{np(1-p)} > 3$, d.h., dass die Standardabweichung der $B(n, p)$-Verteilung größer als 3 ist.

Die Frage, warum statt 3 nicht 2 oder 4 gewählt wird, pflege ich vor Studentinnen und Studenten durch Anschauungsbeispiele zu belegen. Angeregt durch einen Vortrag eines Lehrers auf einer Tagung über Didaktik der Stochastik in Oberwolfach im Jahre 1982, seinen Namen habe ich leider vergessen, präsentiere ich für verschiedene Wertkombinationen von n und p mit Standardabweichungen $\sigma =1$, 3 und 5 auf Folien die Stabdiagramme der $B(n, p)$-Verteilungen und demonstriere damit, dass im Fall $\sigma = 1$ deutliche Abweichungen der Stabdiagramme erkennbar sind, im Falle $\sigma = 3$ für die unterschiedlichsten Wertkombinationen von n und p nur noch relativ geringe Differenzen zu sehen sind und für $\sigma = 5$ mit bloßem Auge nur noch in Randbereichen Unterschiede der Stablängen erkennbar sind, abgesehen von der Lage der Verteilungen, da die Erwartungswerte natürlich verschieden sind. Die Stabdiagramme zu $\sigma = 2$ und $\sigma = 4$ könnte man selbstverständlich zur feineren bildlichen Analyse auch noch heranziehen. Bei den Werten $\sigma =1$, 3 und 5 treten allerdings die qualitativen Unterschiede der Stabdiagramme deutlicher zu tage. (Für $\sigma = 1$ wähle ich die (n, p)-Kombinationen $(4, 0.5)$, $(20, 0.0528)$ und $(100, 0.0101)$, für $\sigma = 3$ die Kombinationen $(36, 0.5)$, $(100, 0.1)$ und $(300, 0.031)$ und für $\sigma = 5$ die Kombinationen $(100, 0.5)$, $(250, 0.1127)$ und $(500, 0.0528)$.)

Hieraus folgt, dass für $\sigma = 1$ keine Hoffnung bestehen kann, die Verteilungen universell approximieren zu können. Dass für $\sigma = 3$ und größer die Umrisse der Stabdiagramme durch geeignete Skalierungen der Gauß'schen Glockenkurve beschrieben werden können, demonstriert man gut mit Hilfe weiterer Folien.

Im Übrigen pflege ich die Laplace'sche Faustregel in die folgende Form zu kleiden:

$$np(1-p) \gtrapprox 9,$$

d.h., $np(1-p)$ sollte größer als 9 oder ungefähr gleich 9 sein, mit anderen Worten: nicht deutlich kleiner als 9 sein. Diese Faustregel ist angemessener als die strikte Forderung $np(1-p) > 9$. Denn warum sollte man eine Binomialverteilung mit $np(1-p) = 9.1$ durch die Normalverteilung approximieren dürfen und im Fall $np(1-p) = 8.9$ nicht? Die Diskussion darüber und entsprechende numerische Beispiele werden Schüler zusätzlich für die Abhängigkeit der Güte der Normalverteilungsapproximation von der Varianz $\sigma^2 = np(1-p)$ sensibel machen können, so dass ihnen nicht bei $np(1-p) = 9$ eine Art Quantensprung oder Phasenübergang suggeriert wird.

Im Fall $np^2 \lessapprox 0.05$ (analog für $n(1-p)^2 \lessapprox 0.05$) erhält man übrigens schon gute Approximationen durch Poisson-Wahrscheinlichkeiten mit Parameter $\lambda = np$, im Folgenden kurz mit $P(\lambda)$-Verteilung bezeichnet. Hinter dieser Faustregel steht die allgemein gültige Abschätzung

$$\sup_{k \in \{0,1,2,\ldots\}} |B(k; n, p) - P(k; np)| \leq 2np^2,$$

wobei $B(k;n,p) = \sum_{i=0}^{k} \binom{n}{i} p^i (1-p)^{n-i}$ der Wert der kumulativen Verteilungsfunktion der $B(n,p)$-Verteilung und $P(k;\lambda) = \sum_{i=0}^{k} \frac{\lambda^i}{i!} e^{-\lambda}$ derjenige der $P(\lambda)$-Verteilung an der Stelle k ist. (Siehe z. B. Krickeberg und Ziezold [3, S. 114f.] oder, in noch allgemeinerer Fassung, Krengel [2, S. 86f.].) Interessant sind in diesem Zusammenhang Wertkombinationen wie $(n,p) = (5000, 0.002)$, für die sowohl $np(1-p) > 9$ als auch $np^2 < 0.05$, so dass beide Approximationen der $B(n,p)$-Verteilung zuverlässige Ergebnisse versprechen, z.B. $B(6;5000,0.002) = 0.1299$, $\Phi(\frac{6-10+0.5}{\sqrt{9.98}}) = 0.1340$ und $P(6;10) = 0.1301$. Im Übrigen liefert die oben genannte allgemein gültige Abschätzung einen Fehler von $2np^2 = 0.04$, also einen wesentlich größeren Wert als der tatsächliche Fehler von etwa 0.0002, so dass auch hier der tatsächliche Fehler stark überschätzt wird.

Schließlich ist noch zu bemerken, dass der Satz von Berry und Esseen nichts über die Genauigkeit der *Korrekturformel* zur Approximation einer $B(n,p)$-Verteilung durch die Normalverteilung aussagt. Diese ist in der Regel deutlich besser als die direkt aus dem Grenzwertsatz von De Moivre und Laplace abgeleitete Formel.

Fazit

Der für den Schulunterricht wohl durchsichtigste Zugang zum Grenzwertsatz von De Moivre und Laplace scheint die über den lokalen Grenzwertsatz begründbare Formel

$$b(k;n,p) \approx \frac{1}{\sqrt{2\pi}} \int_{(k-m-1/2)h}^{(k-m+1/2)h} \mathrm{e}^{-x^2/2} dx \qquad (1)$$

$$= \Phi((k-m+1/2)h) - \Phi((k-m-1/2)h)$$

zu sein. Hierin ist $h = \frac{1}{\sqrt{np(1-p)}}$ und $m = [(n+1)p]$, also die Stelle, an der $k \mapsto b(k;n,p)$ maximal ist. Statt m verwenden wir im Folgenden np, was sowohl zu schlechteren als auch zu besseren Approximationsergebnissen führen kann, aber sich formal an die Standardisierung der binomialverteilten Zufallsvariablen anlehnt. Durch diese Ersetzung wird der Integrationsbereich in (1) nur etwas verschoben, ohne seine Länge zu ändern.

Hieraus folgt dann leicht die Approximation

$$\sum_{k=k_1}^{k_2} b(k;n,p)$$
$$\approx \Phi((k_2 - np + 1/2)h) - \Phi((k_1 - np - 1/2)h))$$
$$= \Phi\left(\frac{k_2 - np + 1/2}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{k_1 - np - 1/2}{\sqrt{np(1-p)}}\right).$$

Sodann demonstriere man mittels Stabdiagrammen und numerischen Beispielen die Güte der Approximation und erwähne an geeigneter Stelle, dass es zwar Fehlerabschätzungen gibt, dass diese aber für praktische Zwecke viel zu hohe Fehler suggerieren und demgemäß nur für theoretische Analysen brauchbar sind. Die Faustregel $np(1-p) \gtrapprox 9$ ist unter diesen Umständen ein willkommener Anhaltspunkt, um beurteilen zu können, wann eine Approximation hinreichend gut sein dürfte.

Ist diese Faustregel verletzt, gilt aber $np^2 \lessapprox 0.05$ oder $n(1-p)^2 \lessapprox 0.05$, so hat man als Ausweichmöglichkeit die Approximation durch die Poisson-Verteilung mit Parameter $\lambda = np$ bzw. $\lambda = n(1-p)$ zur Verfügung. Im letzteren Fall approximiert man, wir benutzen $q = 1-p$,

$$P\{S_n = k\} = P\{n - S_n = n-k\} \approx \frac{(nq)^{n-k}}{(n-k)!} \mathrm{e}^{-nq}$$

und

$$B(k;n,p) \approx 1 - P(n-k-1;nq).$$

Schließlich ist zu bedenken, dass es heutzutage viele schnelle Computerprogramme zur exakten Berechnung von Binomialwahrscheinlichkeiten gibt, die in den meisten Beispielen die Approximation durch die Normalverteilung oder durch die Poissonverteilung unnötig machen. So kann man sich häufig darauf beschränken, diese Approximationsmöglichkeiten nur für grobe Schätzungen von Binomialwahrscheinlichkeiten zu nutzen, wenn die numerische Größe des Fehlers ohne praktische Konsequenzen ist oder wenn jene Schätzung als Ausgangspunkt für weitere, feinere Rechnungen mit Computerprogrammen dienen soll.

Literatur

[1] Eichelsbacher, P.: Eine Diskussion der Faustregel von Laplace. Stochastik in der Schule **21** (2001) 1, S. 22-27

[2] Krengel, U.: *Einführung in die Wahrscheinlichkeitstheorie und Statistik.* 5. Auflage. Vieweg, Braunschweig 2000.

[3] Krickeberg, K.; Ziezold, H.: *Stochastische Methoden.* 4. Auflage. Springer, Heidelberg 1995.

Anschrift des Verfassers:
Prof. Dr. Herbert Ziezold
Fachbereich Mathematik/Informatik
Universität Gesamthochschule Kassel
34109 Kassel
ziezold@mathematik.uni-kassel.de