4 Normalapproximation der Binomialverteilung

Es sei daran erinnert, daß eine Zufallsgröße X mit der Verteilung

$$P(X=k) = b(k;n,p) = \binom{n}{k} p^k q^{n-k} \quad (q = 1-p) \text{ für } k = 0, 1, \ldots, n$$

binomialverteilt heißt.

Die exakten Werte für $b(k;n,p)$ lassen sich bei festem p allerdings nur für moderat große n und k ($n = 100$ und $k = 50$ ist z.B. schon nicht mehr so leicht) berechnen. Im Falle großer n hilft uns aber eine Version des *Zentralen Grenzwertsatzes*, einer Art Naturgesetz, das die asymptotische Verteilung einer großen Klasse von Variablen angibt.

Die Basis für diese Approximation ist die *Stirlingsche Formel*, die von *James Stirling* (1692-1770) bewiesen wurde:

(4.1) Satz.

$$\lim_{n\to\infty} n!/(\sqrt{2\pi} n^{n+1/2} e^{-n}) = 1.$$

Für einen Beweis: Siehe etwa O. Forster: Analysis 1 §20 Satz 6.

Man bemerke, daß die Stirlingsche Formel nicht bedeutet, daß $|n! - \sqrt{2\pi} n^{n+1/2} e^{-n}|$ gegen 0 konvergiert, im Gegenteil. Es gilt

$$\lim_{n\to\infty} |n! - \sqrt{2\pi} n^{n+1/2} e^{-n}| = \infty.$$

Die erste Frage, die man sich stellen sollte ist die, in welchem Sinne man eigentlich einen Limes von $b(k;n,p)$ sinnvoll definieren kann. Dazu bemerken wir zunächst, daß

$$\frac{b(k+1;n,p)}{b(k;n,p)} = \frac{(n-k)p}{(k+1)(1-p)}$$

ist und daher

$$\frac{b(k+1;n,p)}{b(k;n,p)} < 1 \Leftrightarrow k+1 > (n+1)p.$$

Die Funktion $k \mapsto b(k;n,p)$ nimmt also ihr Maximum genau bei $k = [n+1]p$ an. Nun ist aber mit Hilfe der Stirlingschen Formal sofort klar, daß

$$\begin{aligned} b([n+1(p)];n,p) \simeq b(np;n,p) &\simeq \frac{(\frac{n}{e})^n \sqrt{2\pi n} p^{np} (1-p)^{n-np}}{(\frac{np}{e})^{np} \sqrt{2\pi np} (\frac{n-np}{e})^{n-np} \sqrt{2\pi(n-np)}} \\ &= \sqrt{\frac{1}{2\pi np(1-p)}}, \end{aligned}$$

wobei wir für zwei Folgen a_n und b_n schreiben $a_n \simeq b_n$, falls $\lim_{n\to\infty} \frac{a_n}{b_n} = 1$. Also ist für jedes k und p $\lim_{n\to\infty} b(k;n,p) = 0$. Diese Aussage ist eben so wahr wie unnütz. Im wesentlichen bedeutet sie, daß man, um einen "vernünftigen Grenzwert" zu erhalten, nicht einzelne Wahrscheinlichkeiten $b(k;n,p)$ anschauen sollte, sondern die Wahrscheinlichkeit für ganze Bereiche, also: $\sum_{\alpha_n a+c_n\le k\le \alpha_n b+c_n} b(k;n,p)$, wobei $a,b\in\mathbb{R}$ reelle Zahlen sind und α_n, c_n Funktionen von n sind. Wie aber soll man α_n, c_n wählen? Zunächst ist klar, daß die Binomialverteilung $b(\cdot;n,p)$ den Erwartungswert np hat, es also ratsam ist, $c_n = np$ zu wählen, damit die obige Summe für reelle a,b von einer relevanten Größenordnung ist. Andererseits zeigt die obige Rechnung, daß $\max_k b(k;n,p) \simeq \sqrt{\frac{1}{2\pi np(1-p)}}$ ist. Nimmt man an, daß die Terme $b(k;n,p)$ für k nahe bei np von derselben Ordnung sind, so liegt es nahe $\alpha_n = \sqrt{n}$ oder besser $\alpha_n = p(1-p)\sqrt{n}$ zu wählen (um ein Resultat zu erhalten, das von p nicht abhängt). Letzteres wird in der Tat unsere Wahl sein.

Der erste Schritt zur Herleitung eines Grenzwertsatzes für die Binomialverteilung wird sein, daß wir zunächst die $b(k;n,p)$ einzeln genauer unter die Lupe nehmen. Wir werden sehen, daß diese für relevante k tatsächlich von der Ordnung $1/\sqrt{n}$ sind und darüber hinaus zeigt sich, daß dann $b(k;n,p)$ durch eine schöne Funktion approximiert werden kann. Dazu setzen wir

$$x_k := x_k(n,p) := \frac{k-np}{\sqrt{np(1-p)}}.$$

x_k hängt natürlich von n und p ab, was wir in der Notation jedoch nicht gesondert betonen. Wir kürzen $1-p$ meist durch q ab.

(4.2) Satz. *(lokaler Grenzwertsatz, local limit theorem)* Es seien $0<p<1$, $q=1-p$ und $(a_n)_{n\in\mathbb{N}} > 0$ eine Folge reeller Zahlen mit $\lim_{n\to\infty} a_n^3/\sqrt{n} = 0$. Dann gilt

$$\lim_{n\to\infty} \sup_{k;|x_k|\le a_n} \left| \frac{\sqrt{2\pi npq}\, b(k;n,p)}{e^{-x_k^2/2}} - 1 \right| = 0.$$

(4.3) Bemerkungen.

1. Ist $a_n = A$ eine beliebige, aber feste positive Konstante, so folgt aus dem obigen Satz unmittelbar
$$\lim_{n\to\infty} \sup_{k:|x_k|\le A} \left| \frac{\sqrt{2\pi npq}\, b(k;n,p)}{e^{-x_k^2/2}} - 1 \right| = 0.$$

2. Wir schreiben nachfolgend stets $b(k;n,p) \sim \frac{1}{\sqrt{2\pi npq}} e^{-x_k^2/2}$ für die obige gleichmäßige Konvergenz. Allgemeiner: Sind $\alpha(k,n)$, $\beta(k,n) > 0$ für $n\in\mathbb{N}_0$, $0\le k\le n$, so bedeutet (während des untenstehenden Beweises) $\alpha(k,n)\sim\beta(k,n)$, daß für die obige Folge $(a_n)_{n\in\mathbb{N}} > 0$
$$\lim_{n\to\infty} \sup_{k:|x_k|\le a_n} \left| \frac{\alpha(k,n)}{\beta(k,n)} - 1 \right| = 0$$
gilt.

3. Wir überzeugen uns vom folgenden Sachverhalt, der im Beweis von (4.2) mehrfach verwendet wird:
$$\alpha(k,n)\sim\beta(k,n),\ \ \alpha'(k,n)\sim\beta'(k,n) \ \Rightarrow\ \alpha(k,n)\alpha'(k,n)\sim\beta(k,n)\beta'(k,n).$$

Beweis.

$$\begin{aligned}\left|\frac{\alpha(k,n)\alpha'(k,n)}{\beta(k,n)\beta'(k,n)}-1\right| &\le \frac{\alpha'(k,n)}{\beta'(k,n)}\left|\frac{\alpha(k,n)}{\beta(k,n)}-1\right|+\left|\frac{\alpha'(k,n)}{\beta'(k,n)}-1\right| \\ &\le \left|\frac{\alpha'(k,n)}{\beta'(k,n)}-1\right|\left|\frac{\alpha(k,n)}{\beta(k,n)}-1\right|+\left|\frac{\alpha'(k,n)}{\beta'(k,n)}-1\right|+\left|\frac{\alpha(k,n)}{\beta(k,n)}-1\right|.\end{aligned}$$

Daraus folgt die Aussage sofort. □

Beweis von Satz (4.2). Es gilt

$$k = np + \sqrt{npq}\,x_k, \quad n-k = nq - \sqrt{npq}\,x_k,$$

also

$$k \sim np, \quad n-k \sim nq.$$

Mit Hilfe der Stirlingschen Formel folgt:

$$\begin{aligned}b(k;n,p) &\sim \frac{(\frac{n}{e})^n\sqrt{2\pi n}p^k q^{n-k}}{(\frac{k}{e})^k\sqrt{2\pi k}(\frac{n-k}{e})^{n-k}\sqrt{2\pi(n-k)}} \\ &= \sqrt{\frac{n}{2\pi k(n-k)}}\varphi(n,k) \sim \frac{1}{\sqrt{2\pi npq}}\varphi(n,k),\end{aligned}$$

wobei wir $\varphi(n,k)$ für $(\frac{np}{k})^k(\frac{nq}{n-k})^{n-k}$ schreiben. Es ist nun

$$-\log\varphi(n,k) = nH(k/n|\,p),$$

wobei

$$H(x|\,p) = x\log\Big(\frac{x}{p}\Big) + (1-x)\log\Big(\frac{1-x}{1-p}\Big)$$

(diese Funktion heißt *relative Entropie* von x bezüglich p; sie wird im Rahmen des Studiums der großen Abweichungen (Kapitel 6) eine zentrale Rolle spielen). Wir wollen diese Funktion nun um den Wert p Taylor entwickeln. Es ist $H'(p|\,p) = 0$ und $H''(p|\,p) = 1/p + 1/q = 1/(pq)$. Damit folgt

$$H(x|\,p) = \frac{(x-p)^2}{2pq} + \psi(x-p),$$

wobei ψ das Restglied in der Taylorentwicklung bezeichnet. Insbesondere gilt in jedem endlichen Intervall, das p enthält eine Abschätzung

$$|\psi(x-p)| \le c|x-p|^3$$

mit einer geeigneten Konstanten c. Wir erhalten somit

$$\left|-\log\varphi(n,k) - \frac{n\big(\frac{k}{n}-p\big)^2}{2pq}\right| \le cn\left|\frac{k}{n}-p\right|^3.$$

Aus der Definition der x_k erhält man für eine geeignete Konstante $0 < c' < \infty$ folgt

$$\frac{|k-np|^3}{n^2} = c'\frac{|x_k|^3}{\sqrt{n}}.$$

Wählen wir nun ein k mit $|x_k| \le a_n$, so konvergiert aufgrund der Bedingung an die Folge $(a_n)_{n\in\mathbb{N}}$ die rechte Seite der Ungleichung gegen 0. Da nun aber

$$\frac{n\left(\frac{k}{n}-p\right)^2}{2pq} = \frac{x_k^2}{2},$$

erhalten wir

$$\lim_{n\to\infty} \sup_{k:|x_k|\le a_n} \left|\frac{\varphi(n,k)}{e^{-x_k^2/2}} - 1\right| = 0.$$

Damit ist der Satz gezeigt. □

Ein Rechenbeispiel dazu:
Jemand wirft 1200-mal einen Würfel. Mit welcher Wahrscheinlichkeit hat er genau 200-mal eine 6? Mit welcher Wahrscheinlichkeit 250-mal?

Wir berechnen x_k für $k = 200, 250$, $n = 1200$, $p = 1/6$.

$$x_{200} = 0, \;\; x_{250} = \frac{5\sqrt{6}}{\sqrt{10}} = 3.873$$
$$b(200; 1200, 1/6) \cong 0.0309019$$
$$b(250; 1200, 1/6) \cong 0.0000170913.$$

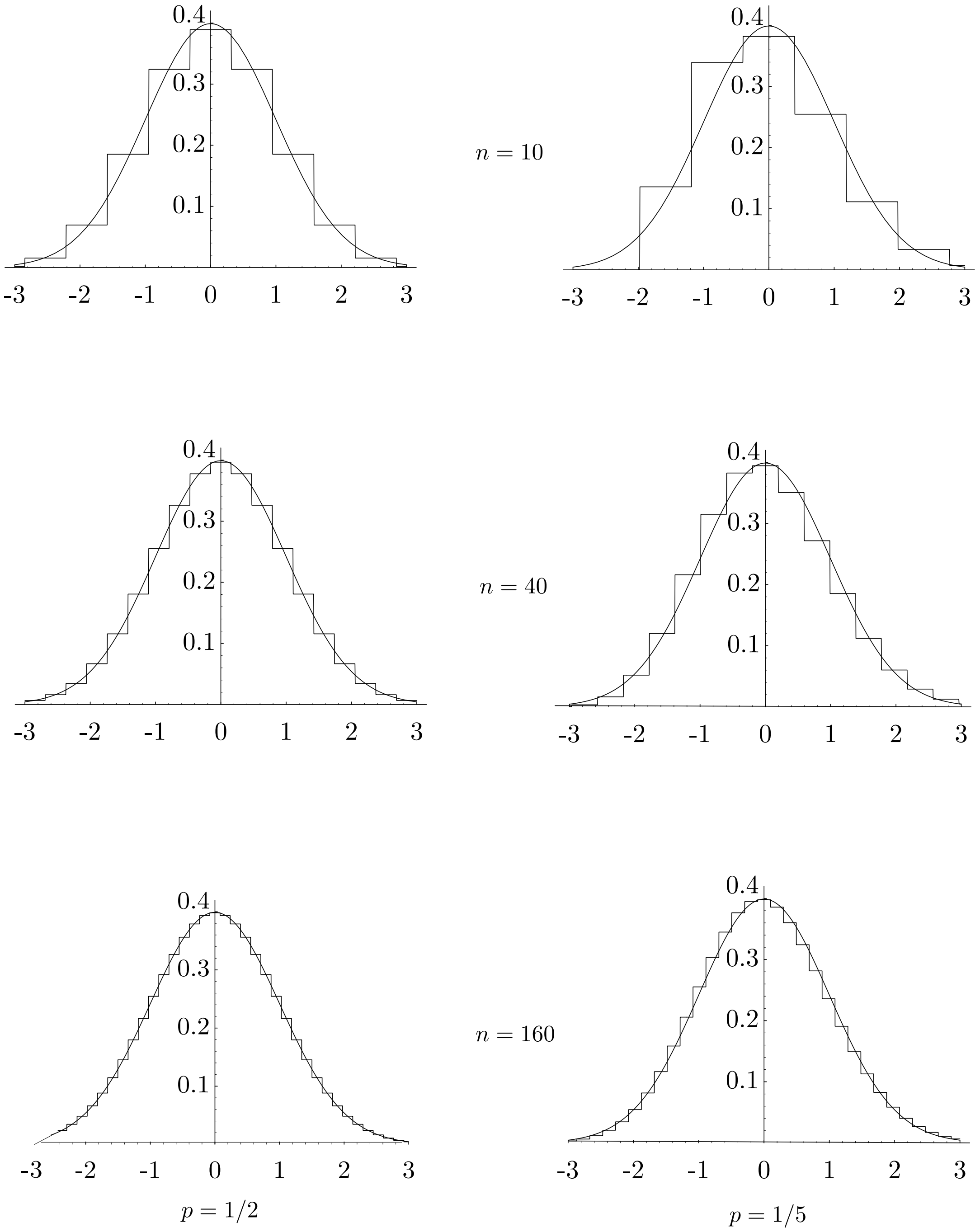
Wie üblich muß hier bemerkt werden, daß ein reines Limesresultat für die Güte einer Approximation wie in obigem Rechenbeispiel zunächst natürlich gar nichts aussagt. Gefragt sind konkrete Abschätzungen des Fehlers. Dies ist ein technisch aufwendiges Feld, in das wir in dieser Vorlesung nicht eintreten werden.

Nachfolgend ist eine numerische Illustration von (4.19) angegeben:

Die sechs Bilder illustrieren die Konvergenz der Binomialverteilung gegen die Funktion $\varphi(x) = (2\pi)^{-1/2}\exp(-x^2/2)$. Hier ist jeweils die Funktion $\varphi(x)$ zusammen mit dem skalierten Histogramm

$$f_{n,p}(x) = \begin{cases} \sqrt{np(1-p)}\,b(k;n,p), & \text{falls } k \in \{0,1,\dots,n\} \text{mit } |x - x_k| < \frac{1}{2\sqrt{np(1-p)}}, \\ 0 & \text{andernfalls,} \end{cases}$$

der Binomialverteilung $b(\cdot\,; n, p)$ gezeichnet; in der linken Spalte der symmetrische Fall mit $p = 1/2$, in der rechten Spalte der asymmetrische Fall $p = 1/5$.



Nun kommen wir dazu, die schon eingangs diskutierten "Bereichswahrscheinlichkeiten" zu approximieren.

(4.4) Satz. *(von de Moivre-Laplace)* Für beliebige reelle Zahlen a und b mit $a < b$ gilt:

$$\lim_{n\to\infty} P\left(a \leq \frac{S_n - np}{\sqrt{npq}} \leq b\right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx. \tag{4.1}$$

Beweis. Die zentrale Idee des Beweises ist es für die einzelnen Summanden der linken Seite von (4.1) die Approximation aus dem lokalen Grenzwertsatz einzusetzen und zu sehen, daß dies eine Riemannsumme für das Integral auf der rechten Seite von (4.1) liefert.

Sei also $k \in \{0, \ldots, n\}$. Dann ist $\{S_n = k\} = \{(S_n - np)/\sqrt{npq} = x_k\}$. Also ist die links stehende Wahrscheinlichkeit gleich

$$\sum_{k:a \leq x_k \leq b} P(S_n = k) = \sum_{k:a \leq x_k \leq b} b(k; n, p).$$

Wir setzen nun für jeden Summanden auf der rechten Seite seinen in Satz (4.2) angegebenen asymptotischen Wert ein und berücksichtigen, daß $x_{k+1} - x_k = \frac{1}{\sqrt{npq}}$ ist. Die Summe dieser Größen nennen wir R_n:

$$R_n = \frac{1}{\sqrt{2\pi}} \sum_{k:a \leq x_k \leq b} e^{-x_k^2/2}(x_{k+1} - x_k).$$

Unter Verwendung der Gleichmäßigkeit der Konvergenz in Satz (4.2) sieht man sofort, daß der Quotient von $P(a \leq \frac{S_n - np}{\sqrt{npq}} \leq b)$ und dem obenstehenden Ausdruck gegen 1 konvergiert, das heißt, es existiert eine Nullfolge $(\varepsilon_n)_{n \in \mathbb{N}}$, $\varepsilon_n > 0$ mit

$$R_n(1 - \varepsilon_n) \leq P\left(a \leq \frac{S_n - np}{\sqrt{npq}} \leq b\right) \leq R_n(1 + \varepsilon_n). \tag{4.2}$$

k und x_k entsprechen einander bikjektiv, und wenn k von 0 bis n läuft, dann variiert x_k im Intervall $[-\sqrt{np/q}, \sqrt{nq/p}]$ mit der Schrittweite $x_{k+1} - x_k = 1/\sqrt{npq}$. Für hinreichend große n umfaßt dieses Intervall das gegebene Intervall $[a, b]$, und die in $[a, b]$ fallenden Punkte x_k teilen dieses in Teilintervalle derselben Länge $1/\sqrt{npq}$. Wenn nun der kleinste und der größte Wert von k mit $a \leq x_k \leq b$ gleich j bzw. l ist, dann ist

$$x_{j-1} < a \leq x_j < x_{j+1} < \ldots < x_{l-1} < x_l \leq b < x_{l+1}$$

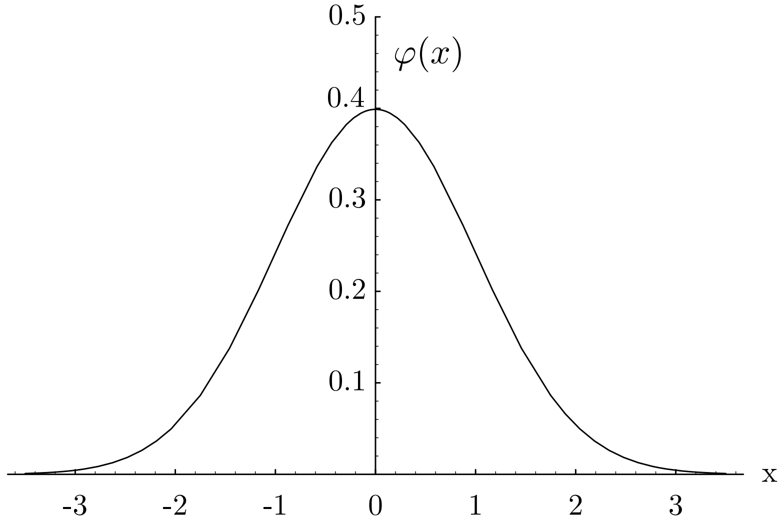
und die obige Summe läßt sich schreiben als

$$\sum_{k=j}^{l} \varphi(x_k)(x_{k+1} - x_k),$$

wobei $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ ist. Das ist eine Riemannsche Summe für das bestimmte Integral $\int_a^b \varphi(x)dx$. Somit konvergiert R_n mit $n \to \infty$ gegen das Integral in der Behauptung des Satzes. Dieser folgt nun sofort mit (4.2). □

Abraham de Moivre (1667–1754) veröffentlichte dieses Ergebnis in seiner „*Doctrine of Chances*“ 1714. *Pierre Simon Marquis de Laplace* (1749-1827) erweiterte das Ergebnis und wies dessen Bedeutung in seiner „*Théorie analytique des probabilités*“ 1812 nach. Es handelt sich um den zuerst bekanntgewordenen Spezialfall des sogenannten *Zentralen Grenzwertsatzes (central limit theorem)*.

Die Funktion $x \to \varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ heißt auch Gaußsche Glockenkurve, wegen des glockenförmigen Verlaufs ihres Graphen.



Die Verteilung, die durch das Integral unter der Glockenkurve gegeben ist, heißt auch Standard–Normalverteilung und wird oft mir $\mathcal{N}(0,1)$ abgekürzt.

Die Integrale $\int_a^b \varphi(x)dx$ sind leider nicht in geschlossener Form mit Hilfe von Polynomen, rationalen Funktionen, Wurzelausdrücken oder elementaren transzendenten Funktionen (wie sin, cos, exp, etc.) darstellbar.

Es gilt offenbar für $a < b$

$$\int_a^b \varphi(x)dx = \int_{-\infty}^b \varphi(x)dx - \int_{-\infty}^a \varphi(x)dx = \Phi(b) - \Phi(a),$$

wobei wir $\Phi(y) := \int_{-\infty}^y \varphi(x)dx$ gesetzt haben. Wie nicht anders zu erwarten ist, gilt

$$\int_{-\infty}^{\infty} \varphi(x)dx = 1. \tag{4.3}$$

Der Beweis, den man üblicherweise in der Analysis für diese Tatsache gibt, benutzt Polarkoordinaten. Wir geben hier einen Beweis, der sich darauf stützt, daß wir den Satz von de-Moivre-Laplace schon kennen: Wir verwenden (4.4) und setzen $S_n^* := \frac{S_n - np}{\sqrt{npq}}$. (Für das Argument hier spielt p keine Rolle; wir können z.B. $p = 1/2$ nehmen.) Sei $a > 0$. Dann ist

$$1 = P(-a \leq S_n^* \leq a) + P(|S_n^*| > a).$$

Nach der Tschebyscheff-Ungleichung gilt:

$$P(|S_n^*| > a) \leq \frac{1}{a^2} \operatorname{Var}(S_n^*) = \frac{1}{a^2}.$$

Nach (4.4) gilt

$$\lim_{n\to\infty} P(-a \leq S_n^* \leq a) = \int_{-a}^a \varphi(x)dx.$$

Demzufolge ist

$$1 - \frac{1}{a^2} \leq \int_{-a}^a \varphi(x)dx \leq 1$$

für jedes $a > 0$, womit (4.6) bewiesen ist.

(4.7) Bemerkung. (a) Wegen $\lim_{n\to\infty} \sup_k P(S_n = k) = 0$ ist es natürlich gleichgültig, ob in der Aussage von (4.21) $\leq$ oder $<$ steht.

(b) Es gilt für $a \in \mathbb{R}$:

$$\begin{aligned}
\lim_{n\to\infty} P\left(\frac{S_n - np}{\sqrt{npq}} \leq a\right) &= \Phi(a) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} e^{-x^2/2} dx, \\
\lim_{n\to\infty} P\left(\frac{S_n - np}{\sqrt{npq}} \geq a\right) &= 1 - \Phi(a).
\end{aligned}$$

Beweis von (b). Wir beweisen die erste Gleichung; die zweite folgt analog. Wegen der Symmetrie von φ und (4.6) gilt:

$$\Phi(x) = \int_{-\infty}^{x} \varphi(u) du = 1 - \int_{x}^{\infty} \varphi(u) du = 1 - \int_{-\infty}^{-x} \varphi(u) du = 1 - \Phi(-x).$$

Wir setzen wieder $S_n^* = \frac{S_n - np}{\sqrt{npq}}$ und wählen $b > 0$ so groß, daß $-b < a$ gilt. Dann ist nach (4.4)

$$\begin{aligned}
\limsup_{n\to\infty} P(S_n^* \leq a) &= \limsup_{n\to\infty} (P(-b \leq S_n^* \leq a) + P(S_n^* < -b)) \\
&= \limsup_{n\to\infty} (P(-b \leq S_n^* \leq a) + (1 - P(S_n^* \geq -b))) \\
&\leq \limsup_{n\to\infty} (P(-b \leq S_n^* \leq a) + (1 - P(-b \leq S_n^* \leq b))) \\
&= \Phi(a) - \Phi(-b) + (1 - \Phi(b) + \Phi(-b)) \\
&= \Phi(a) + \Phi(-b) \\
\liminf_{n\to\infty} P(S_n^* \leq a) &\geq \liminf_{n\to\infty} P(-b \leq S_n^* \leq a) \\
&= \Phi(a) - \Phi(-b).
\end{aligned}$$

Wegen $\Phi(-b) \to 0$ für $b \to \infty$ folgt die gewünschte Aussage. □

Der Satz (4.4) ist eine Präzisierung des Gesetzes der großen Zahlen, welches besagt, daß für jedes $\varepsilon > 0$ $\lim_{n\to\infty} P(|\frac{S_n}{n} - p| \geq \varepsilon) = 0$ ist. Letzteres können wir sofort auch aus (4.4) herleiten:

$$\begin{aligned}
P\left(\left|\frac{S_n}{n} - p\right| \leq \varepsilon\right) &= P\left(-\varepsilon \leq \frac{S_n}{n} - p \leq \varepsilon\right) \\
&= P\left(-\frac{\sqrt{n}\varepsilon}{\sqrt{pq}} \leq \frac{S_n - np}{\sqrt{npq}} \leq \frac{\sqrt{n}\varepsilon}{\sqrt{pq}}\right) \geq P\left(a \leq \frac{S_n - np}{\sqrt{npq}} \leq b\right),
\end{aligned}$$

sofern n so groß ist, daß $\sqrt{n}\varepsilon/\sqrt{pq} \geq b$ und $-\sqrt{n}\varepsilon/\sqrt{pq} \leq a$ sind. Für beliebige Zahlen $a, b \in \mathbb{R}$ ist dies aber für genügend große n der Fall. Somit ist $\lim_{n\to\infty} P(|\frac{S_n}{n} - p| \leq \varepsilon) = 1$ für jedes $\varepsilon > 0$.

Dieser Beweis ist natürlich insgesamt wesentlich aufwendiger als der in Kapitel 3 angegebene. (4.4) ist jedoch sehr viel informativer als das Gesetz der großen Zahlen.

Tabelle der Verteilungsfunktion $\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du$ für $x \geq 0$. Wir hatten bereits gesehen, daß für $x \leq 0$ gilt: $\Phi(x) = 1 - \Phi(-x)$.

x	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7356	0.7389	0.7421	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7703	0.7734	0.7764	0.7793	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8364	0.8389
1.0	0.8413	0.8437	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8687	0.8708	0.8728	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8906	0.8925	0.8943	0.8962	0.8979	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9146	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9278	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9624	0.9633
1.8	0.9641	0.9648	0.9656	0.9664	0.9671	0.9678	0.9685	0.9692	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9761	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9874	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9895	0.9898	0.9901	0.9903	0.9906	0.9908	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9924	0.9926	0.9928	0.9930	0.9932	0.9934	0.9936
2.5	0.9938	0.9939	0.9941	0.9943	0.9944	0.9946	0.9947	0.9949	0.9950	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9958	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9973
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986
3.0	0.9986	0.9987	0.9987	0.9988	0.9988	0.9988	0.9989	0.9990	0.9989	0.9990

Wir wollen nun sehen, dass das Grenzwertverhalten des Satzes von de Moivre/Laplace ein Spezialfall eines viel allgemeineren Phänomens ist, eines Satzes, der neben dem Gesetz der großen Zahlen ein zweites "Naturgesetz" der Stochastik darstellt. Wie wir dies schon im Satz von de Moivre/Laplace kennengelernt haben, befasst sich dieser Satz mit der Konvergenz von Verteilungen $P_n(\bullet) = P[X_n \in \bullet]$ für geeignete Zufallsvariablen X_n. Es läge sicherlich nahe davon zu sprechen, dass eine Folge von Verteilungen P_n gegen eine

Grenzverteilung P_0 konvergiert, falls

$$P_n(\{x\}) \underset{n\to\infty}{\longrightarrow} P_0(\{x\}) \quad \forall\, x \in \mathbb{R}$$

bzw.

$$F_n(x) := P_n((-\infty, x]) = \sum_{y \le x} P_n(y) \to F_0(x)$$

für eine geeignete (Verteilungs-) Funktion F_0 gilt.

Das folgende Beispiel zeigt, dass diese Begriffsbildung nicht das Gewünschte liefert.

(4.8) Beispiel. Seien X_n Zufallsvariablen die im Punkt $\frac{1}{n}$ konzentriert sind, d. h. für alle $n \in \mathbb{N}$ gelte

$$P(X_n = \frac{1}{n}) = 1.$$

Die P_n sind entsprechend Deltafunktionen in $\frac{1}{n}$:

$$P_n(\{x\}) = \delta_{x-\frac{1}{n}}.$$

Es ist anschaulich klar, dass die P_n gegen die Dirac-Verteilung in der 0 konvergieren. Dies würde der obige Konvergenzbegriff aber nicht leisten, denn

$$\lim_{n\to\infty} P_n(\{0\}) = 0 \neq 1 = P_0(\{0\}),$$

wenn P_0 gerade die Dirac-Verteilung in der 0 ist. Entsprechend gilt auch

$$\lim_{n\to\infty} F_n(0) = 0 \neq 1 = F_0(0).$$

Die Schwierigkeit ist hierbei offenbar, dass, der Limes F_0 gerade im Punkt 0 unstetig ist. Um diese Schwierigkeit zu umgehen, verlangt man für den neuen Konvergenzbegriff nur das Folgende:

(4.9) Definition. Eine Folge von Verteilungsfunktionen F_n von Wahrscheilichkeiten P_n auf $\mathbb{R}$ heißt verteilungskonvergent gegen F_0, falls F_0 eine Verteilungsfunktion ist, d. h. falls gilt

a) F_0 ist monoton wachsend;

b) F_0 ist rechtsseitig stetig;

c) $\lim_{x\to-\infty} F(x) = 0$ und $\lim_{x\to\infty} F(x) = 1$

und falls

$$F_n(x) \to F_0(x)$$

für alle x, in denen F_0 stetig ist, gilt. Ist F_0 die Verteilungsfunktion einer Wahrscheinlichkeit P_0 auf $\mathbb{R}$, so schreiben wir

$$P_n \xrightarrow{\mathcal{D}} P_0.$$

(4.10) Beispiel. Für die Funktion F_n, F_0 aus dem Eingangsbeispiel gilt

$$F_n(x) = \begin{cases} 1 & x \geq 1/n \\ 0 & x < 1/n \end{cases} \quad \text{also} \quad \lim_{n\to\infty} F_n(x) = \begin{cases} 1 & x > 0 \\ 0 & x \leq 0 \end{cases}$$

Dies impliziert die Verteilungskonvergenz von F_n gegen F_0.

Es ist interessant, diesen neuen Begriff zu vergleichen mit der Konvergenz von Zufallsvariaben X_n gegen eine Zufallsvariable X_0 in Wahrscheinlichkeit. Letzteres bedeutet, dass analog zum Gesetz der großen Zahlen gilt

$$P(|X_n - X_0| \geq \varepsilon) \to 0 \quad \forall\, \varepsilon > 0.$$

Wir werden sehen, dass der Begriff der Verteilungskonvergenz schwächer ist als der Begriff der Konvergenz in Wahrscheinlichkeit:

(4.11) Satz. Es seien $(X_n)_n$ Zufallsvariablen mit

$$X_n \to X_0 \quad \text{in Wahrscheinlichkeit.}$$

Dann konvergiert P^{X_n}, die Verteilung von X_n, in Verteilung gegen P^{X_0}.

Beweis: Wir schreiben

$$F_n := P^{X_n} \quad \text{bzw.} \quad F_0 := P^{X_0}.$$

Es sei x ein Stetigkeitspunkt von F_0 und $\varepsilon > 0$. Dann gibt es ein $\delta > 0$ mit

$$F_0(x) - \varepsilon \leq F_0(x - \delta) = P(X_0 \leq x - \delta)$$

und

$$F_0(x + \delta) \leq F_0(x) + \varepsilon.$$

Nun gilt aber für alle $n \in \mathbb{N}$:

$$\{X_0 \leq x - \delta\} \subseteq \{X_n < x\} \cup \{|X_n - x| \geq \delta\},$$

da $X_n \geq x$ und $|X_n - X_0| < \delta$ folgt

$$X_0 = (X_0 - X_n) + X_n > x - \delta.$$

Hieraus folgt

$$F_0(x) \leq F_0(x - \delta) + \varepsilon \leq F_n(x) + P(|X_n - X_0| \geq \delta) + \varepsilon.$$

Analog gilt

$$\{X_n \leq x\} \subset \{X_0 < x + \delta\} \cup \{|X_n - X_0| \geq \delta\},$$

also auch

$$F_n(x) \leq F_0(x) + P(|X_n - X_0| \geq \delta) + \varepsilon.$$

Insgesamt erhält man:

$$|F_n(x) - F_0(x)| \leq \varepsilon + P(|X_n - X_0| \geq \delta).$$

Da der letzte Term für $n \to \infty$ verschwindet, folgt die Behauptung. $\square$

(4.12) Bemerkung. Die Umkehrung des vorhergehenden Satzes gilt in der Regel nicht, wie dieses Beispiel zeigt. X sei eine Zufallsvariable mit

$$P(X = 1) = P(X = -1) = \frac{1}{2}.$$

(X_n) sei eine Folge von Zufallsvariablen mit

$$X_{2n} = X \quad \text{und} \quad X_{2n+1} = -X \quad \forall\, n \in \mathbb{N}.$$

Da $P^{X_n} = P^X$ für alle $n \in \mathbb{N}$ gilt, ist X_n natürlich verteilungskovnergent gegen X. Andererseits gilt

$$P(|X_{2n+1} - X| \geq 1) = 1 \quad \forall\, n \in \mathbb{N}.$$

Wir werden diesen Begriff in der Wahrscheinlichkeitstheorie noch genauer betrachten. Für den Moment begnügen wir uns mit einer hinreichenden Bedingung für die Verteilungskonvergenz.

(4.13) Satz. Es seien P_n diskrete Wahrscheinlichkeitsverteilungen über $\mathbb{R}$ und $F_0 : \mathbb{R} \to \mathbb{R}$ differenzierbar, monoton wachsend mit

$$\lim_{x \to -\infty} F_0(x) = 0 \quad \text{und} \quad \lim_{x \to \infty} F_0(x) = 1.$$

Gilt dann

$$\lim_{n \to \infty} \sum_{x \in \mathbb{R}} f(x) P_n(\{x\}) = \int_{-\infty}^{\infty} f(x) F_0'(x) dx,$$

so für alle stetigen Funktionen $f : \mathbb{R} \to \mathbb{R}$ mit existenten Limiten $\lim_{x \to \pm\infty} f(x)$, so ist P_n verteilungskonvergent und es gilt

$$F_n \to F_0 \quad \text{in Verteilung.}$$

Beweis. Sei $g : \mathbb{R} \to \mathbb{R}$ definiert durch

$$g(x) = 1_{(-\infty,0]}(x) + (1 - x) 1_{(0,1)}(x).$$

g ist stetig und es gilt $\lim_{x \to \infty} g(x) = 0$, $\lim_{x \to -\infty} g(x) = 1$. Selbiges gilt für die Funktionen

$$f_k(x) := g(kx).$$

Für die zu P_n gehörigen Verteilungsfunktion F_n gilt dann zum einen für alle $x \in \mathbb{R}$ und $k \in \mathbb{N}$

$$\begin{aligned}
\limsup_{n \to \infty} F_n(x) &= \limsup_{n \to \infty} \sum_{y \leq x} P_n(\{y\}) \\
&\leq \limsup_{n \to \infty} \sum_{y} f_k(y - x) P_n(\{y\}) \\
&= \int_{-\infty}^{\infty} f_k(y - x) F_0'(y) dy \\
&\leq \int_{-\infty}^{x + \frac{1}{k}} F_0'(y) dy = F_0(x + \frac{1}{k}).
\end{aligned}$$

Hierbei haben wir zunächst verwendet, dass g auf $\mathbb{R}^-$ gleich 1 ist, dann die Voraussetzung eingesetzt und schließlich nochmals die Definition von f_k. Andererseits gilt

$$\begin{aligned}
\liminf_{n\to\infty} F_n(x) &= \liminf_{n\to\infty} \sum_{y\le x} P_n(\{y\}) \\
&\ge \liminf_{n\to\infty} \sum_{y} f_k(y-x+\frac{1}{k})P_n(\{y\}) \\
&= \int_{-\infty}^{\infty} f_k(y-x+\frac{1}{k})F_0'(y)dy \\
&\ge \int_{-\infty}^{x-\frac{1}{k}} F_0'(y)dy = F_0(x-\frac{1}{k}).
\end{aligned}$$

Da F_0 insbesondere überall stetig ist, folgt

$$\lim_{k\to\infty} F_0(x+\frac{1}{k}) = \lim_{x\to\infty} F_0(x-\frac{1}{k}) = F_0(x),$$

also insgesamt

$$\lim_{n\to\infty} F_n(x) = F_0(x) \quad \forall\, x \in \mathbb{R}.$$

□

Mit diesem Hilfsmittel an der Hand können wir nun die folgende, allgemeinere Version des Satzes von de Moivre/Laplace beweisen:

(4.14) Satz. (Satz von Lindeberg-Levy/Spezialfall) Es seien für alle n $X_1, X_2, \ldots, X_n$ stochastisch unabhängige Zufallsvariablen, die alle dieselbe diskrete Verteilung besitzen und deren Erwartungswerte $\mathbb{E}X_1$ und Varianzen $\mathbb{V}(X_1) > 0$ existieren. Dann gilt

$$\lim_{n\to\infty} P(-\infty < \frac{\sum_{i=1}^n (X_i - \mathbb{E}X_1)}{\sqrt{n\mathbb{V}(X_1)}} \le x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}dt,$$

d. h. die Variablen $\sum_{i=1}^n (X_i - \mathbb{E}X_1)/\sqrt{n\mathbb{V}(X_1)}$ sind verteiungskonvergent mit Limes

$$F_0(x) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}dt.$$

(4.15) Bemerkungen.

a) Da $e^{-y^2/2}$ schneller fällt als jede Potenz, ist $\int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-y^2/2}dy$ existent.

b) F_0 ist monoton wachsend, stetig und $\lim_{x\to-\infty} F_0(x) = 0$. Außerdem ist F_0 differenzierbar und nach dem Hauptsatz der Infinitesimalrechnung gilt

$$F_0'(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

Schließlich lernt man auch in der Analysis, dass

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-t^2/2}dt = 1 \quad \text{also} \quad \lim_{x\to\infty} F_0(x) = 1$$

gilt.

c) Man beachte, dass die Aussage des Satzes unabhängig ist von der Gestalt der Verteilung von X_1.

Beweis des Satzes. Setze
$$Y_i := \frac{X_i - \mathbb{E}X_1}{\sqrt{\mathbb{V}X_1}}.$$
Die Y_i sind mit den X_i unabhängig und identisch verteilt. Es gilt
$$\mathbb{E}Y_i = 0 \quad \text{und} \quad \mathbb{V}(Y_i) = 1.$$
Setzen wir weiter
$$\bar{Y}_n = \frac{1}{n}\sum_{i=1}^n Y_i,$$
so ist
$$\sqrt{n}\bar{Y}_n = \frac{\sum_{i=1}^n (X_i - \mathbb{E}X_1)}{\sqrt{n\mathbb{V}X_1}}.$$
Mit
$$\varphi(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$$
wollen wir also für jede stetige Funktion $f : \mathbb{R} \to \mathbb{R}$ mit existenten Limiten $\lim_{x\to\pm\infty} f(x)$ beweisen, dass
$$\lim_{n\to\infty}\sum_y f(y)P(\sqrt{n}\bar{Y}_n = y) = \int_{-\infty}^{\infty} f(y)\varphi(y)dy =: I(f).$$
Da man von f immer die Konstante $I(f)$ subtrahieren kann, können wir o.B.d.A. $I(f) = 0$ annehmen. Betrachte
$$h(x) = \frac{1}{\varphi(x)}\int_{-\infty}^{x} f(y)\varphi(y)dy.$$
Da f konstruktionsgemäß gleichmäßig stetig und beschränkt ist, ist h wohldefiniert und als Quotient stetiger Funktionen stetig. Da
$$\varphi'(x) = -\frac{x}{\sqrt{2\pi}}e^{-x^2/2} = -x\varphi(x)$$
gilt, folgt
$$h'(x) = \frac{f(x)\varphi^2(x) - \int_{-\infty}^x f(y)\varphi(y)dy\varphi'(x)}{\varphi^2(x)} = f(x) + xh(x),$$
für alle $x \in \mathbb{R}$. Natürlich ist auch $xh(x)$ stetig und mit l'Hospital folgt
$$\begin{aligned}\lim_{x\to\pm\infty} xh(x) &= \lim_{x\to\pm\infty}\frac{\int_{-\infty}^x f(y)\varphi(y)dy}{\frac{\varphi(x)}{x}}\\ &= \lim_{x\to\pm\infty}\frac{f(x)\varphi(x)}{\frac{-x^2\varphi(x)-\varphi(x)}{x^2}} = \lim_{x\to\pm\infty}\frac{f(x)\varphi(x)}{-\varphi(x)(1+\frac{1}{x^2})}\\ &= -\lim_{x\to\pm\infty} f(x).\end{aligned}$$

Dies wenden wir folgendermaßen an:

$$\begin{aligned}\sum_y f(y)P(\sqrt{n}\bar{Y}_n = y) &= \mathbb{E}[f(\sqrt{n}\bar{Y}_n)] \\ &= \mathbb{E}[h'(\sqrt{n}\bar{Y}_n)] - \mathbb{E}[\sqrt{n}\bar{Y}_n h(\sqrt{n}\bar{Y}_n)] \\ &= \mathbb{E}[h'(\sqrt{n}\bar{Y}_n)] - \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\mathbb{E}[Y_j h(\sqrt{n}\bar{Y}_n)].\end{aligned}$$

Aufgrund der Unabhängigkeit und identischen Verteilung der Y_j ist dies gleich

$$= \mathbb{E}[h'(\sqrt{n}\bar{Y}_n)] - \sqrt{n}\mathbb{E}[Y_1 h(\sqrt{n}\bar{Y}_n)].$$

Nun betrachten wir die Taylor-Entwicklung von h um

$$Z_n := \frac{1}{\sqrt{n}}\sum_{j=2}^{n} Y_j.$$

Dies ergibt:

$$h(\sqrt{n}\bar{Y}_n) = h(Z_n) + h'(Z_n)\frac{Y_1}{\sqrt{n}} + \frac{Y_1}{\sqrt{n}}R_n$$

mit

$$R_n = h'(Z_n + \vartheta\frac{Y_1}{\sqrt{n}}) - h'(Z_n) \quad \text{für ein} \quad \vartheta \in [0,1).$$

Nun sind konstruktionsgemäß Y_1 und Z_n stochastisch unabhängig. Daraus folgt

$$\begin{aligned}\mathbb{E}[Y_1 h(\sqrt{n}\bar{Y}_n)] &= \mathbb{E}(Y_1)\mathbb{E}(h(Z_n)) + \mathbb{E}(Y_1^2)\frac{1}{\sqrt{n}}\mathbb{E}[h'(Z_n)] + \frac{1}{\sqrt{n}}\mathbb{E}[Y_1^2 R_n] \\ &= \frac{\mathbb{E}[h'(Z_n)]}{\sqrt{n}} + \frac{\mathbb{E}[Y_1^2 R_n]}{\sqrt{n}}.\end{aligned}$$

Insgesamt ergibt dies:

$$\mathbb{E}[f(\sqrt{n}\bar{Y}_n)] = \mathbb{E}[h'(Z_n + \frac{Y_1}{\sqrt{n}}) - h'(Z_n)] - \mathbb{E}[Y_1^2 \cdot (h'(Z_n + \frac{\vartheta Y_1}{\sqrt{n}}) - h'(Z_n)].$$

Da h' gleichmäßig stetig ist, konvergieren für festes ω wegen

$$\lim \frac{Y_1(w)}{\sqrt{n}} = 0$$

die Summanden unter beiden Erwartungswerten gegen 0. Da außerdem h' beschränkt ist, konvergieren auch die zugehörigen Erwartungswerte gegen 0. Dies ergibt

$$\lim_{n\to\infty}\mathbb{E}[f(\sqrt{n}\bar{Y}_n)] = \lim_{n\to\infty}\sum f(y)P[\sqrt{n}Y_n = y] = 0.$$

Das war zu zeigen. □

Ein Anwendungsbeispiel (Außersinnliche Wahrnehmung (ASW))
1973 machte C. Tert (Univ. California, Davis) ein Experiment zu ASW. Eine Aquarius genannte Maschine wählte zufällig ein Symbol von A,B,C,D und die Versuchsperson sollte erraten, welches. Tert nahm 15 Personen mit vermuteten “hellseherischen Fähigkeiten” und testete jede 500 Mal. Von den entstandenen 7500 Versuchen waren 2006 Treffer. Bei rein zufälligem Raten wären $7500 : 4 = 1875$ Treffer zu erwarten gewesen. Frage: Können die restlichen $2006 - 1875 = 131$ Treffer durch Zufallsschwankungen erklärt werden ?

Zur Beantwortung dieser Frage bezeichnen wir mit X die Anzahl der Treffer unter der Annahme, daß diese rein zufällig zustande kommen. Wir verwenden den Satz von de Moivre und Laplace mit

$$n = 7500; p = \frac{1}{4}; (1-p) = \frac{3}{4}$$

und erhalten

$$P(X \geq 2006) = P\left(\frac{X - 1875}{\sqrt{7500 \times \frac{1}{4} \times \frac{3}{4}}} \geq \frac{131}{\sqrt{7500 \times \frac{1}{4} \times \frac{3}{4}}}\right).$$

Nach dem Satz von de Moivre und Laplace ist die Größe auf der rechten Seite der Gleichung annähernd normalverteilt, d. h. gemäß $\mathcal{N}(0,1)$. Also

$$P(X \geq 2006) \approx P(X^* \geq 3.5) \approx 0.00023,$$

wobei X^* eine standardnormalverteilte Zufallsvariable bezeichnet. Die Wahrscheinlichkeit dafür, daß die auftretende Differenz das Produkt einer Zufallsschwankung ist, liegt also bei 2.3 Promille und ist damit extrem klein. Trotzdem beweist dieses Experiment nicht mit Sicherheit, daß es ASW gibt, da z.B. im Nachhinein festgestellt wurde, daß der Zufallsgenerator nicht besonders zuverlässig war. (Quellen: C. Tert; Learning to use extrasensory perception, Chicago Univ. Press (1976); M. Gardner; ESP at random, New York book reviews (1977))

Eine weitere Anwendungsmöglichkeit des Satzes von de Moivre und Laplace ist die, auszurechnen, wie groß eine Stichprobe sein muß, um Aussagen über den Parameter p einer Binomialverteilung mit einer gewissen Sicherheit und Genauigkeit machen zu können. Obwohl diese Fragestellung eigentlich in die Statistik gehört, wollen wir uns hierzu schon einmal ein Beispiel anschauen:

Beispiel: In einer Population will man den Anteil an Linkshändern mit 95% Sicherheit auf 1% Genauigkeit bestimmen. Wie viele Personen sollte man dazu (mit Zurücklegen) befragen?

Wir wollen die Wkeit mit Hilfe der Approximation durch die Normalverteilung berechnen. Dazu sei X die Anzahl der Linkshänder in der Stichprobe, $\frac{X}{n}$ ist dann der geschätzte Prozentsatz an Linkshändern in der Gesamtpopulation (warum das eine sinnvolle Schätzung ist, werden wir in dem Kapitel über Statistik diskutieren). Wir wollen, daß

$$|\frac{X}{n} - p| \leq \varepsilon = 0.01$$

und das mit 95% Sicherheit, also

$$P(|\frac{X}{n} - p| \leq 0.01) \geq 0.95. \tag{4.4}$$

Bringt man die Wahrscheinlichkeit auf die Form im Satz von de Moivre und Laplace so ergibt sich:

$$\begin{aligned} P(|\frac{X}{n} - p| \leq 0.01) &= P(-0.01 \leq \frac{X}{n} - p \leq 0.01) \\ &= P\left(\frac{-0.01\sqrt{n}}{\sqrt{p(1-p)}} \leq \frac{X - np}{\sqrt{np(1-p)}} \leq \frac{0.01\sqrt{n}}{\sqrt{p(1-p)}}\right). \end{aligned}$$

Nun kennen wir p dummerweise nicht; aber es gilt stets $p(1-p) \leq \frac{1}{4}$. Setzen wir dies ein, erhalten wir

$$\begin{aligned} & P\left(\frac{-0.01\sqrt{n}}{\sqrt{p(1-p)}} \leq \frac{X}{-np}\sqrt{np(1-p)} \leq \frac{-0.01\sqrt{n}}{\sqrt{p(1-p)}}\right) \\ \geq\ & P\left(-0.01 \times 2\sqrt{n} \leq \frac{X - np}{\sqrt{np(1-p)}} \leq 0.01 \times 2\sqrt{n}\right). \end{aligned}$$

Nach dem Satz von de Moivre und Laplace ergibt sich

$$P\left(-0.01 \times 2\sqrt{n} \leq \frac{X - np}{\sqrt{np(1-p)}} \leq 0.01 \times 2\sqrt{n}\right) \approx \Phi(z) - \Phi(-z) = 2\Phi(z) - 1,$$

da $\Phi(-z) = 1 - \Phi(z)$, wobei

$$z := 0.02\sqrt{n}.$$

Um nun (4.8) zu erfüllen, bestimmen wir aus einer $\mathcal{N}(0,1)$-Tafel z so, daß

$$2\Phi(z) - 1 = 0.95 \Leftrightarrow \Phi(z) = 0.975.$$

Dies ergibt (ungefähr) $z \approx 2$. Setzen wir die Definition von z wieder ein, erhalten wir

$$z = 0.02\sqrt{n} = 2, \text{ d.h.: } n = 10000.$$

Zu bemerken ist noch, daß der benötigte Umfang n der Stichprobe n quadratisch von der Approximationsgenauigkeit ε abhängt. Benötigt man beispielsweise nur eine Genauigkeit von 2% (oder 5%), so genügt eine Stichprobe vom Umfang 2500 (400), um das Ziel mit 95% Sicherheit zu erreichen.

Desweiteren bietet sich noch die Möglichkeit, den Stichprobenumfang durch eine Vorabinformation, wo ungefähr p liegen könnte, zu verkleinern.