Kapitel 4

Grenzwertsätze

Dieses Kapitel ist gewissermaßen der Höhepunkt der bisher entwickelten Theorie und bietet gleichzeitig einen Übergang zum Einsatz der W'keitstheorie in der Praxis. Bislang waren wir immer von Modellvorgaben ausgegangen: "Sei X so-und-so-verteilt mit dem/der/den Erwartungswert/Streuung/Eigenschaften etc." In der praktischen Anwendung ist es kaum realistisch, von konkreten Verteilungen mit bekannten Parametern auszugehen. Stattdessen wird man eher einen Satz von Messdaten vorliegen haben, die stochastisch zu interpretieren sind (Statistik). Die folgenden Grenzwertsätze sind hilfreich, statistische Daten in das Gerüst der W'keitstheorie einzuordnen.

4.1 Das (schwache) Gesetz der großen Zahl

Gegeben sei eine unfaire Münze. Um $P(„\text{Kopf}“)$ zu bestimmen, wird man sie häufig werfen und dann

$$P(„\text{Kopf}“) \approx \frac{\text{Anzahl der geworfenen Köpfe}}{\text{Anzahl aller Würfe}}$$

setzen. Wieso (bzw., in welchem Sinne) liefert dies eine vernünftige Schätzung für $P(„\text{Kopf}“)$?

Satz 4.1: Das (schwache) Gesetz der großen Zahl

Seien $X_1, \ldots, X_n$ unabhängige Zufallsvariable, alle mit dem selben Erwartungswert $\mu = \mathrm{E}(X_1) = \cdots = \mathrm{E}(X_n)$ und der selben Streuung $\sigma = \sigma(X_1) = \cdots = \sigma(X_n)$. Sei

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$$

die "Mittelwert"–Variable. Für jedes $\epsilon > 0$ gilt

$$P(|\bar{X}_n - \mu| \geq \epsilon) \leq \frac{\sigma^2}{\epsilon^2\, n} \ ,$$

bzw., äquivalenterweise

$$P(|\bar{X}_n - \mu| < \epsilon) \geq 1 - \frac{\sigma^2}{\epsilon^2\, n} \ .$$

Bew: Wegen der Unabhängigkeit ist nicht nur der Erwartungswert, sondern auch die Varianz linear (siehe Abschnitt 2.5.2):

$$\mathrm{E}\,(\bar{X}_n) = \frac{1}{n} \sum_{i=1}^{n} \mathrm{E}\,(X_i) = \frac{1}{n} \sum_{i=1}^{n} \mu = \mu \ ,$$

$$\mathrm{Var}\,(\bar{X}_n) = \mathrm{Var}\Big(\frac{1}{n} \sum_{i=1}^{n} X_i\Big) = \frac{1}{n^2} \sum_{i=1}^{n} \mathrm{Var}\,(X_i) = \frac{1}{n^2} \sum_{i=1}^{n} \sigma^2 = \frac{\sigma^2}{n} \ .$$

Die Chebyshev-Ungleichung 2.36 liefert sofort:

$$P(|\bar{X}_n - \mathrm{E}\,(\bar{X}_n)| \geq \epsilon) = P(|\bar{X}_n - \mu| \geq \epsilon) \leq \frac{\mathrm{Var}\,(\bar{X}_n)}{\epsilon^2} = \frac{\sigma^2}{\epsilon^2\, n} \ .$$

Q.E.D.

25.6.57↓ **Interpretation 1, 4.2:**

Man betrachte den Spezialfall, dass alle Variablen $X_1, \ldots, X_n$ die selbe Verteilung haben, also das selbe Zufallsexperiment darstellen: es geht also um n unabhängige Wiederholungen des selben Experiments.

> Die Variablen $X_1, \ldots, X_n$ werden als unabhängige Wiederholungen ein und des selben Zufallsexperiments X interpretiert. Die Mittelwertvariable $\bar{X}_n$ entspricht dann dem Zufallsexperiment: „führe n unabhängige Einzelmessungen von X durch und betrachte den Mittelwert“.

Zu jedem noch so kleinen vorgegebenen $\epsilon > 0$ wird die Wahrscheinlichkeit, dass der Mittelwert um mehr als ϵ vom (in der Regel unbekannten) Erwartungswert des Einzelexperiments X abweicht, mit wachsendem n immer kleiner (also, je öfter man das Einzelexperiment wiederholt). Komplementär: die Wahrscheinlichkeit, dass der Mittelwert bis auf ein ϵ den Erwartungswert liefert, liegt bei genügend großem n praktisch bei 1.

In dieser Interpretation liefert das Gesetz der großen Zahl überhaupt erst die Grundlage, die mathematische W'keitstheorie auf praktische Fragestellungen anwenden zu können (Statistik). Die Parameter einer Verteilung wie z.B. der Erwartungswert sind in praktischen Experimenten unbekannt, man muss sie durch Messungen abschätzen:

Man kann den Erwartungwert $\mu = \mathrm{E}\,(X)$ eines Einzelexperiments X durch empirische Mittelwerte $\bar{X}_n$ messen! Wenn die "Stichprobengröße" n nur groß genug ist, d.h., wird nur genügend oft wiederholt, so approximiert der Mittelwert mit großer W'keit den Erwartungswert. Für jedes feste $\epsilon > 0$ gilt:

$$\lim_{n\to\infty} P(|\bar{X}_n - \mu| < \epsilon) = 1 \ .$$

Umgekehrt liefert diese Aussage die Interpretation des Erwartungswertes als Kennzeichen einer Zufallsvariable, der ja in Definition 2.22 als rein mathematisches Objekt eingeführt wurde:

Der Erwartungswert einer Zufallsvariable entspricht dem Wert, den man nach häufiger unabhängiger Wiederholung des Experiments durch Mittelung erhält.

Daher der Sprachgebrauch 2.23:

„*X nimmt* im Mittel *den Erwartungswert* $\mathrm{E}\,(X)$ *an*".

Interpretation 2, 4.3:

Wir lassen nun einen gewissen Paradigmen-Wechsel zu. Wie interpretiert man eine W'keitsaussage wie

$$P(|\bar{X}_n - \mu| < \epsilon) \geq 1 - \frac{\sigma^2}{\epsilon^2\, n}$$

in einer praktischen Anwendung? Bislang war unsere Interpretation: sei das Modell (der Stichprobenraum, die Verteilung, die Parameter wie z.B. der Erwartungswert μ etc.) vorgegeben. Dann wurde die obige W'keit folgendermaßen interpretiert:

> Führen wir (in Zukunft) ein reales Zufallsexperiment durch, so werden wir dabei mit der berechneten bzw. abgeschätzten W'keit konkrete Werte $X_1 = x_1$, $X_2 = x_2$, $\ldots$, $X_n = x_n$ ermitteln, deren Mittelwert $\bar{x}_n = \frac{1}{n}\sum_{i=1}^n x_i$ bis auf eine absolute Genauigkeit ϵ Werte in der Nähe des (vorgegebenen) Erwartungswerts μ annimmt.

D.h., eine praktische Interpretation beinhaltet, Zufallsvariable X_i (mathematisch ein "Abstraktum": Funktionen vom Stichprobenraum nach $\mathbb{R}$) durch konkrete Zahlen x_i ("Messwerte eines tatsächlich durchgeführten Zufallsexperiments") zu ersetzen.

Von nun an erlauben wir auch folgende Interpretation: Das tatsächliche Zufallsexperiment sei bereits durchgeführt, es liegen Zahlenwerte ("Messungen") $X_1 = x_1, , \ldots , X_n = x_n$ mit dem Mittelwert $\bar{x}_n = \frac{1}{n}\sum_{i=1}^n x_i$ vor. Wo bleibt nun in der Aussage

$$P(|\bar{X}_n - \mu| < \epsilon) \geq 1 - \frac{\sigma^2}{\epsilon^2\, n}$$

das intuitive Konzept von "W'keit", wenn $\bar{X}_n$ einen gemessenen Zahlenwert $\bar{x}_n$ angenommen hat? Nun, man mache sich klar, was ein Modell (im Sinne der philosophischen Erkenntnistheorie) prinzipiell ist: es kann nicht (und soll nicht) ein mathematisches Abbild einer "absoluten Wahrheit" sein. Konkret heißt dies hier, weder konkrete Verteilungen noch Parameter sind in praktischen Anwendungen wirklich bekannt.

Ersetzen wir in $P(|\bar{X}_n - \mu| < \epsilon)$ die Zufallsvariable $\bar{X}_n$ durch einen Messwert $\bar{x}_n$, so interpretieren wir

$$P(|\bar{x}_n - \mu| < \epsilon) \geq 1 - \frac{\sigma^2}{\epsilon^2\, n}$$

folgendermaßen:

> $P(|\bar{x}_n - \mu| < \epsilon)$ gibt die „**Sicherheit**“ an, dass der (unbekannte) Erwartungsert μ bis auf eine absolute Genauigkeit ϵ in der Nähe des gemessenen Mittelwerts $\bar{x}_n$ liegt.

Diese Interpretation lässt sich nicht sauber in das strenge Konzept der W'keitstheorie einordnen (dazu müssten wir μ in irgendeinem Sinne als eine Zufallsvariable interpretieren). Daher sprechen wir bei $P(|\bar{x}_n - \mu| < \epsilon)$ nicht von einer W'keit, sondern von einer (nur intuitiv zu interpretierenden) Sicherheit.

In der Tat ist die obige Interpretation ein arg einfaches Weltbild, das zu Proteststürmen ausgemachter Stochastiker führen dürfte, die in der „Schätztheorie“ eine saubere Interpretation im Sinne des W'keitsmaßes der unterliegenden Modelle benutzen.

Das Interval $[\bar{x}_n - \epsilon, \bar{x}_n + \epsilon]$ wird auch **Konfidenzinterval** *für den Parameter μ zur Sicherheit $P(|\bar{X}_n - \mu| < \epsilon)$ genannt.*

Bemerkung 4.4: *Es gibt auch ein sogenanntes <u>starkes</u> Gesetz der großen Zahl. Dies ist die (stärkere) Aussage*

$$P(\lim_{n\to\infty} |\bar{X}_n - \mu| = 0) = 1 \ ,$$

wobei mit $P(\lim_{n\to\infty} |\bar{X}_n - \mu| = 0)$ gemeint ist $P(\{\omega \in \Omega;\ \lim_{n\to\infty} \bar{X}_n(\omega) = \mu\})$. Siehe Krengel, Kapitel §12.

Beispiel 4.5: Zurück zum Problem, für eine unfaire Münze $P(„\text{Kopf}“)$ durch empirische Messung zu bestimmen. Allgemein, sei ein Bernoulli-Experiment $X : \Omega \mapsto \{0, 1\}$ mit unbekannter Erfolgsw'keit $p = P(X = 1)$ (und $q = P(X = 0) = 1 - p$) gegeben. Nach Beispiel 2.38 (mit $n = 1$) gilt

$$\mathrm{E}\,(X) = p\ ,\ \ \mathrm{Var}\,(X) = p\,q\ .$$

Seien $X_1, \ldots, X_n$ unabhängige Wiederholungen von X. Damit hat der Mittelwert die Interpretation

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{\text{Anzahl der Erfolge}}{\text{Anzahl der Versuche}}\ .$$

Das Gesetz der großen Zahl liefert für jedes $\epsilon > 0$:

$$P(\bar{X}_n - \epsilon < p < \bar{X}_n + \epsilon)\ \geq\ 1 - \frac{p\,q}{n\,\epsilon^2}\ \geq\ 1 - \frac{1}{4\,n\,\epsilon^2}\ .$$

(Beachte: für jedes Bernoulli-Experiment gilt $\mathrm{Var}\,(X) = p\,q = p\,(1 - p) \leq 1/4$. Dies ist hier hilfreich, da p ja nicht bekannt ist, sondern erst noch abgeschätzt werden soll.) Angenommen, eine unfaire Münze wurde $n = 1000$ mal geworfen, worden, es sei dabei 600 mal „Kopf“ eingetreten, also: das Ereignis $\bar{X}_n = \bar{x}_n = 0.6$ ist eingetreten. Mit verschiedenen Werten von ϵ ergeben sich die Sicherheiten

$$\begin{aligned}
\epsilon = \frac{1}{10}: &\quad P(0.6 - 0.1\ \ < p < 0.6 + 0.1\ \ )\ \geq\ 1 - \frac{1}{4 \cdot 1000 \cdot (1/10)^2} = 0.975\ ,\\
\epsilon = \frac{1}{20}: &\quad P(0.6 - 0.05 < p < 0.6 + 0.05)\ \geq\ 1 - \frac{1}{4 \cdot 1000 \cdot (1/20)^2} = 0.9\ ,\\
\epsilon = \frac{1}{50}: &\quad P(0.6 - 0.02 < p < 0.6 + 0.02)\ \geq\ 1 - \frac{1}{4 \cdot 1000 \cdot (1/50)^2} = 0.375\ .
\end{aligned}$$

Also: nach dem Experiment kann ich mit mindestens 97.5%-iger Sicherheit sagen, dass p im Intervall $(0.5, 0.7)$ liegt. Mit mindestens 90%-iger Sicherheit liegt p im Intervall $(0.55, 0.65)$. Mit mindesten 37.5%-iger Sicherheit liegt p im Intervall $(0.58, 0.62)$.

Anmerkung: eine sehr ähnliche Diskussion haben wir schon einmal in Beispiel 2.39 geführt. Die Variable $S_n = n\bar{X}_n = \sum_i X_i =$ „Anzahl der Erfolge eines Bernoulli-Experiments“ ist bekannterweise binomial-verteilt, also

$$P(|\bar{X}_n - p| < \epsilon) = P(|n\,\bar{X}_n - n\,p| < n\,\epsilon) = P(|S_n - \mathrm{E}\,(S_n)| < n\,\epsilon)$$

$$= \sum_{\substack{k=0,\ldots,n\\ |k-n\,p|<n\,\epsilon}} P(S_n = k)\ \ = \sum_{\substack{k>n\,(p-\epsilon)\\ k<n\,(p+\epsilon)}} \binom{n}{k}\, p^k\, q^{n-k}\ .$$

Da wir p nicht kennen, können wir diese W'keiten zwar nicht exakt berechnen, aber mit $p \approx 0.6$, $n = 1000$ grob abschätzen:

$$P\Big(|\bar{X}_n - p| < \frac{1}{10}\Big) \approx \sum_{k=n\,(0.6-0.1)+1}^{n\,(0.6+0.1)-1} \binom{n}{k}\, 0.6^k\, 0.4^{n-k} \approx 1 - \frac{1.33}{10^{10}} ,$$

$$P\Big(|\bar{X}_n - p| < \frac{1}{20}\Big) \approx \sum_{k=n\,(0.6-0.05)+1}^{n\,(0.6+0.05)-1} \binom{n}{k}\, 0.6^k\, 0.4^{n-k} \approx 0.9986 ,$$

$$P\Big(|\bar{X}_n - p| < \frac{1}{50}\Big) \approx \sum_{k=n\,(0.6-0.02)+1}^{n\,(0.6+0.02)-1} \binom{n}{k}\, 0.6^k\, 0.4^{n-k} \approx 0.792 .$$

Die vom Gesetz der großen Zahl gelieferten Sicherheiten sind also arg untertrieben (kein Wunder, denn das Gesetz der großen Zahl beruht auf Chebyshev 2.36).

Beispiel 4.6: Bei Würfen mit einem evtl. manipulierten Würfel soll die W'keit p, eine „Sechs" zu würfeln, empirisch ermittelt werden. Wie oft muss man werfen, um p mit einer Sicherheit von mindestens 97.5% auf eine absolute Genauigkeit von $\epsilon = 1/1000$ festlegen zu können?

Sei $\bar{X}_n$ wieder die Mittelwertvariable $\frac{1}{n}\times$„Anzahl der geworfenen Sechsen bei n Würfen" (Erwartungswert $\mathrm{E}\,(\bar{X}_n) = p$):

$$P(\bar{X}_n - \epsilon < p < \bar{X}_n + \epsilon) \;\geq\; 0.975 \;=: S .$$

Gemäß Beispiel 4.5 gilt

$$P(\bar{X}_n - \epsilon < p < \bar{X}_n + \epsilon) \;\geq\; 1 - \frac{p\,q}{n\,\epsilon^2} \;\geq\; 1 - \frac{1}{4\,n\,\epsilon^2} .$$

Also muss n so groß sein, dass $1 - 1/(4\,n\,\epsilon^2) \geq S$ gilt, also

$$n \geq \frac{1}{4\,(1-S)\,\epsilon^2} = \frac{10^6}{4 \cdot 0.025} = 10\,000\,000 .$$

Dies ist wiederum eine zwar sichere, aber ziemlich grobe Abschätzung. Setzt man analog zu Beispiel 4.5 die Approximation $p \approx 1/6$ in die exakte Binomial-Verteilung der Variable $S_n =$ „Anzahl der Sechsen" ein, so ergibt sich mit $n = 10\,000\,000$ für eine gemessene relative Häufigkeit $\bar{x}_n$ der Sechsen die wesentlich näher bei 1 liegende Sicherheit

$$P\Big(|\bar{x}_n - p| < \frac{1}{1000}\Big) \approx \sum_{k=\lceil n\,(1/6-1/1000)\rceil}^{\lfloor n\,(1/6+1/1000)\rfloor} \binom{n}{k} \Big(\frac{1}{6}\Big)^k \Big(\frac{5}{6}\Big)^{n-k} \approx 1 - \frac{2.155...}{10^{17}} .$$

Probiert man einige Werte von n für $p = 1/6$ mit der obigen exakten Formel aus, so zeigt sich, dass bereits $n \approx 700\,000$ Versuche für die Sicherheit 0.975 reichen:

$$P\Big(|\bar{x}_n - p| < \frac{1}{1000}\Big) \approx \begin{cases} 0.9624 & \text{für } n = 600\,000 , \\ 0.9752 & \text{für } n = 700\,000 , \\ 0.9836 & \text{für } n = 800\,000 . \end{cases}$$

Wir werden diese Rechnung in Beispiel 4.13 noch einmal aufgreifen.

Allgemein gilt, dass das Gesetz der großen Zahl (also i.W. die Chebyshevsche Ungleichung) zwar sehr wichtig für die Philosophie der W'keitsrechnung ist, für die Praxis aber nur schlechte Abschätzungen liefert (siehe die letzten Beispiele). Der folgende Abschnitt liefert wesentlich genauere Abschätzungen für die Binomial-Verteilung.

4.2 Die Normalverteilung als Grenzwert der Binomial-Verteilung: der Satz von Moivre-Laplace

In Abschnitt 2.4.2 wurde mit der Poisson–Verteilung bereits eine Grenzverteilung der Binomial–Verteilung vorgestellt. Zur Erinnerung: betrachtet man $n \to \infty$ und gleichzeitig $p \to 0$, so dass $\lambda = n\,p$ konstant ist, so gilt

$$\lim_{n\to\infty} \binom{n}{k}\, p^k\, (1-p)^{n-k} = \lim_{n\to\infty} \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1-\frac{\lambda}{n}\right)^{n-k} = \frac{\lambda^k}{k!}\, e^{-\lambda} \ .$$

Wir betrachten hier nun einen anderen Grenzübergang: den Grenzfall $n \to \infty$, wobei p konstant gehalten wird.

Für binomial-verteiltes S_n (= Anzahl der Erfolge bei bei n Wiederholungen eines Bernoulli-Experiments mit Erfolgsw'keit p) betrachte die relative Erfolgshäufigkeit $\bar{X}_n = S_n/n$ und

$$Y_n = \frac{S_n - n\,p}{\sqrt{n\,p\,q}} = \sqrt{\frac{n}{p\,q}}\, \left(\bar{X}_n - p\right) .$$

Es gilt[1]

$$\mathrm{E}\,(Y_n) = \frac{1}{\sqrt{n\,p\,q}}\ \mathrm{E}\left(S_n - n\,p\right) \overset{(*)}{=} 0 \ ,$$

$$\mathrm{Var}\,(Y_n) \overset{(**)}{=} \frac{1}{n\,p\,q}\ \mathrm{Var}\left(S_n - n\,p\right) \overset{(**)}{=} \frac{1}{n\,p\,q}\ \mathrm{Var}\,(S_n) \overset{(*)}{=} 1 \ .$$

Man nennt dieses Y_n auch „zentrierte" (Erwartungswert 0) und „normierte" (Streuung 1) relative Erfolgshäufigkeit. Diese ist für großes n normalverteilt:

[1] Zu $(*)$ beachte man Beispiel 2.38. Zu $(**)$ beachte man:

$$\mathrm{Var}\,(\alpha\,X+\beta) = \mathrm{E}\,((\alpha\,X+\beta)^2) - (\mathrm{E}\,(\alpha\,X+\beta))^2$$
$$= \alpha^2\mathrm{E}\,(X^2) + 2\,\alpha\beta\mathrm{E}\,(X) + \beta^2 - (\alpha\mathrm{E}\,(X)+\beta)^2 = \alpha^2\mathrm{E}\,(X^2) - \alpha^2\mathrm{E}\,(X)^2 = \alpha^2\mathrm{Var}\,(X) \ .$$

Satz 4.7: (Moivre-Laplace)

Sei $S_n : \Omega \to \{0, 1, \ldots, n\}$ binomial-verteilt:

$$P(S_n = k) \; = \; \binom{n}{k} \, p^k \, q^{n-k} \; , \; q = 1 - p.$$

Für $p \cdot q \neq 0$ gilt:

$$\lim_{n \to \infty} P\Big(a \leq \frac{S_n - n\,p}{\sqrt{n\,p\,q}} \leq b \Big) \; = \; \frac{1}{\sqrt{2\,\pi}} \int_a^b e^{-x^2/2} \, dx \; .$$

In Worten: für großes n genügt die zentrierte und normierte Erfolgshäufigkeit einer Normalverteilung (mit Erwartungswert 0 und Streuung 1).

Der Approximationsfehler für großes n ist $O(1/\sqrt{n}\,)$:

$$P\Big(a \leq \frac{S_n - n\,p}{\sqrt{n\,p\,q}} \leq b \Big) \; = \; \frac{1}{\sqrt{2\,\pi}} \int_a^b e^{-x^2/2} \, dx \; + \; O\Big(\frac{1}{\sqrt{n}} \Big) \; .$$

Beweisskizze: Elementares Abschätzen der Binomialkoeffizienten $\binom{n}{k}$ über die Stirling-Formel lässt die Exponentialfunktion auftauchen. Die exakte W'keit $P\Big(a \leq \frac{S_n - n\,p}{\sqrt{n\,p\,q}} \leq b\Big)$ ist eine Summe über Binomial–Terme, die nach der Approximation als Riemann–Summe für das Integral interpretiert werden kann. Die Riemann–Summe konvergiert für $n \to \infty$ gegen das Integral.

Hier die technischen Details:

Es gilt

$$P\Big(a \leq \frac{S_n - n\,p}{\sqrt{n\,p\,q}} \leq b \Big) = P\Big(n\,p + a\,\sqrt{n\,p\,q} \leq S_n \leq n\,p + b\,\sqrt{n\,p\,q} \Big)$$

$$= \sum_{k=k_a}^{k_b} \binom{n}{k} \, p^k \, q^{n-k} \;, \quad k_a = \lceil n\,p + a\,\sqrt{n\,p\,q}\,\rceil \;, \quad k_b = \lfloor n\,p + b\,\sqrt{n\,p\,q}\,\rfloor$$

abzuschätzen. Die Terme sind damit für Werte von k zu approximieren, für die $|k - n\,p| \leq \text{const}\,\sqrt{n}$ gilt, also $1/k = O(1/n)$ und $1/(n-k) = O(1/n)$. Mit der Stirling-Formel für Fakultäten (siehe Abschnitt 2.4.2) folgt

$$\binom{n}{k} \, p^k \, q^{n-k} = \frac{n! \, p^k \, q^{n-k}}{k! \, (n-k)!}$$

$$= \frac{p^k \, q^{n-k}}{\sqrt{2\,\pi}} \, \sqrt{\frac{n}{k\,(n-k)}} \Big(\frac{n}{e}\Big)^n \Big(\frac{e}{k}\Big)^k \Big(\frac{e}{n-k}\Big)^{n-k} \frac{1 + O(\frac{1}{n})}{(1 + O(\frac{1}{k}))\,(1 + O(\frac{1}{n-k}))}$$

$$= \frac{1}{\sqrt{2\,\pi}}\,\sqrt{\frac{n}{k\,(n-k)}}\Big(\frac{n\,p}{k}\Big)^k\Big(\frac{n\,q}{n-k}\Big)^{n-k}\,\Big(1+O\Big(\frac{1}{n}\Big)\Big)\ .$$

Ersetze k durch die Abweichungen vom Erwartungswert $y = y(k) = k - n\,p$. Mit den relevanten Werten von k gilt $|y| \le \text{const}\,\sqrt{n}$, also $y = O(\sqrt{n}\,)$. Mit $k = np + y$, $n - k = n - np - y = nq - y$ gilt:

$$\sqrt{\frac{n}{k\,(n-k)}} = \sqrt{\frac{n}{(np+y)\,(nq-y)}} = \sqrt{\frac{n}{n^2\,p\,q + n\,y\,(q-p) - y^2}}$$

$$= \frac{1}{\sqrt{n\,p\,q\,(1+\frac{y\,(q-p)}{n\,p\,q} - \frac{y^2}{n^2\,p\,q})}} = \frac{1}{\sqrt{n\,p\,q}}\,\Big(1{+}O\Big(\frac{y}{n}\Big)\Big) = \frac{1}{\sqrt{n\,p\,q}}\,\Big(1{+}O\Big(\frac{1}{\sqrt{n}}\Big)\Big)\ .$$

Für den nächsten Faktor gilt:

$$\Big(\frac{n\,p}{k}\Big)^k = \Big(\frac{n\,p}{n\,p+y}\Big)^{n\,p+y} = \Big(\frac{1}{1+\frac{y}{n\,p}}\Big)^{n\,p+y} = e^{-n\,p\,(1+\frac{y}{n\,p})\,\ln(1+\frac{y}{n\,p})}$$

$$= e^{-y-\frac{y^2}{2\,n\,p}+O(\frac{y^3}{n^2})} = e^{-y-\frac{y^2}{2\,n\,p}}\,\Big(1+O\Big(\frac{y^3}{n^2}\Big)\Big) = e^{-y-\frac{y^2}{2\,n\,p}}\,\Big(1+O\Big(\frac{1}{\sqrt{n}}\Big)\Big)\ .$$

Für den nächsten Faktor gilt analog:

$$\Big(\frac{n\,q}{n-k}\Big)^{n-k} = \Big(\frac{n\,q}{n\,q-y}\Big)^{n\,q-y} = \Big(\frac{1}{1-\frac{y}{n\,q}}\Big)^{n\,q-y} = e^{-n\,q\,(1-\frac{y}{n\,q})\,\ln(1-\frac{y}{n\,q})}$$

$$= e^{y-\frac{y^2}{2\,n\,q}+O(\frac{y^3}{n^2})} = e^{y-\frac{y^2}{2\,n\,q}}\,\Big(1+O\Big(\frac{y^3}{n^2}\Big)\Big) = e^{y-\frac{y^2}{2\,n\,q}}\,\Big(1+O\Big(\frac{1}{\sqrt{n}}\Big)\Big)\ .$$

Insgesamt folgt (beachte $\frac{1}{p}+\frac{1}{q} = \frac{p+q}{p\,q} = \frac{1}{p\,q}$):

$$\binom{n}{k}\,p^k\,q^{n-k} = \frac{1}{\sqrt{2\,\pi}}\,\frac{1}{\sqrt{n\,p\,q}}\,e^{-y-\frac{y^2}{2\,n\,p}}\,e^{y-\frac{y^2}{2\,n\,q}}\,\Big(1+O\Big(\frac{1}{\sqrt{n}}\Big)\Big)$$

$$= \frac{1}{\sqrt{2\,\pi}}\,\frac{1}{\sqrt{n\,p\,q}}\,e^{-\frac{y^2}{2\,n\,p\,q}}\,\Big(1+O\Big(\frac{1}{\sqrt{n}}\Big)\Big) = \frac{1}{\sqrt{2\,\pi}}\,\frac{1}{\sqrt{n\,p\,q}}\,e^{-\frac{y^2}{2\,n\,p\,q}}\ +\ O\Big(\frac{1}{n}\Big)\ .$$

Die gesuchte W'keit lässt sich nach diesen Vorbereitungen abschätzen. Es wird y wieder durch $y = k - n\,p$ ersetzt:

$$P := P\Big(a \le \frac{S_n - n\,p}{\sqrt{n\,p\,q}} \le b\Big) = \sum_{k=k_a}^{k_b}\binom{n}{k}\,p^k\,q^{n-k}$$

$$= \underbrace{\sum_{k=k_a}^{k_b}\frac{1}{\sqrt{2\,\pi}}\,\frac{1}{\sqrt{n\,p\,q}}\,e^{-\frac{(k-n\,p)^2}{2\,n\,p\,q}}}_{I}\ +\ \underbrace{\sum_{k=k_a}^{k_b} O\Big(\frac{1}{n}\Big)}_{F}\ .$$

Setzt man $x_k = (k - n\,p)/\sqrt{n\,p\,q}$, so lässt sich der erste Term I als Riemann-Summe mit einer äquidistanten Zerlegung des Intervalls $[x_{k_a}, x_{k_b+1}]$ mit konstanter Schrittweite $x_{k+1} - x_k = 1/\sqrt{n\,p\,q}$ interpretieren. Hierbei ist

$$x_{k_a} = a + O\Big(\frac{1}{\sqrt{n}}\Big)\ , \quad x_{k_b+1} = b + O\Big(\frac{1}{\sqrt{n}}\Big)$$

und es folgt

$$I = \sum_{k=k_a}^{k_b} \frac{1}{\sqrt{2\,\pi}}\ e^{-x_k^2/2}\ (x_{k+1} - x_k) = \frac{1}{\sqrt{2\,\pi}}\ \int_a^b e^{-x^2/2}\ dx\ +\ O\Big(\frac{1}{\sqrt{n}}\Big)\ .$$

(Die Approximation eines Integrals durch eine äquidistante Riemann-Summe hat den Fehler $O(1/\text{Schrittweite})$.) Im Fehlerterm F erstreckt sich die Summe nur über $k_b - k_a + 1 \approx 1 + (b - a)\ \sqrt{n\,p\,q} = O(\sqrt{n}\,)$ Terme, also gilt $|F| \leq O(\sqrt{n}\,)\,O(1/n) = O(1/\sqrt{n}\,)$. Insgesamt ergibt sich damit die Behauptung

$$P = \frac{1}{\sqrt{2\,\pi}}\ \int_a^b e^{-x^2/2}\ dx\ +\ O\Big(\frac{1}{\sqrt{n}}\Big)\ .$$

Für die in Bemerkung 4.8 behauptete gleichmäßige Konvergenz bezüglich a und b müsste man die Abschätzungen verfeinern. Dies ersparen wir uns hier.

Q.E.D.

Bemerkung 4.8: *Man kann zeigen, dass diese Approximation durch die Normalverteilung* gleichmäßig *in den Grenzen a und b ist, d.h., der Fehlerterm $O(1/\sqrt{n}\,)$ kann unabhängig von a und b nach oben abgeschätzt werden durch einen nur von n abhängenden Fehlerterm, der für $n \to \infty$ verschwindet.*

Visualisierung 4.9:

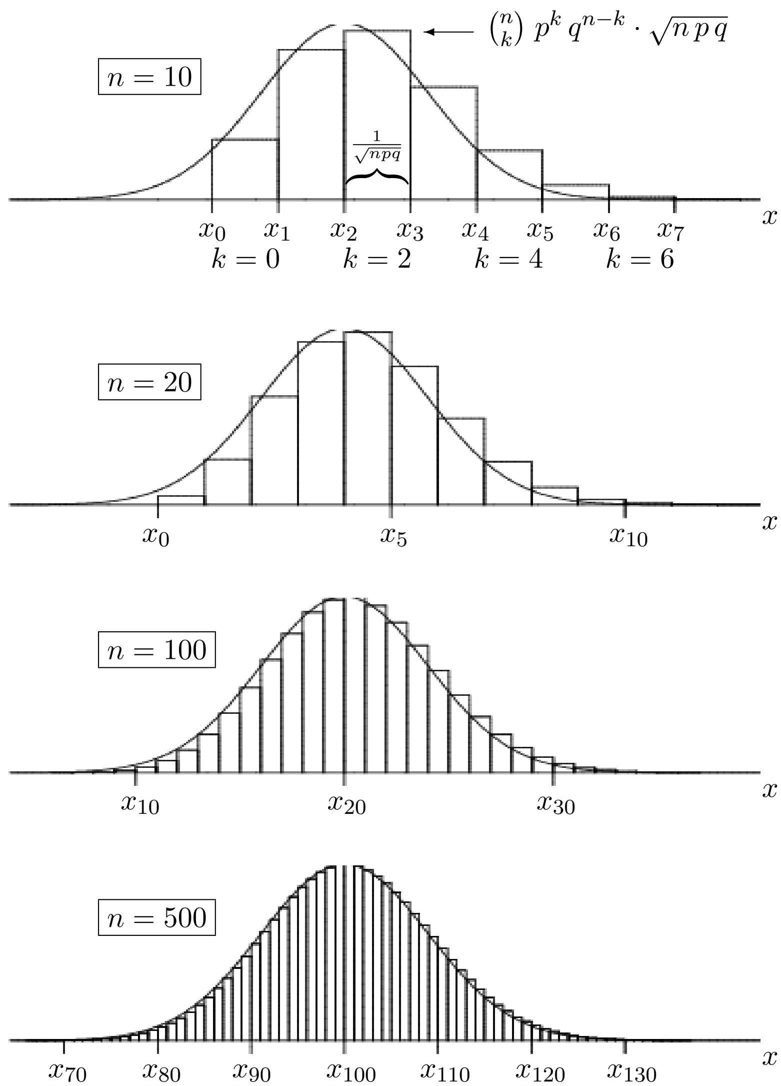
Die folgenden Bilder visualisieren den Grenzübergang. Über dem Intervall $[a, b] = [-4, 5]$ wird die Dichte $\rho(x) = e^{-x^2/2}/\sqrt{2\,\pi}$ der $N(0,1)$-Verteilung gezeichnet. Weiterhin wird für $p = 0.2$ und verschiedene Werte von n das Analogon einer "Dichte" für die diskrete Variable $Y_n = (S_n - n\,p)/\sqrt{n\,p\,q}$ als Balkendiagramm folgendermaßen dargestellt:

Die $Bi(n, p)$-verteilte Variable S_n nimmt die Werte $k = 0, \ldots, n$ an. Diesen k-Werten werden auf der x-Achse die Intervalle $[x_k, x_{k+1})$ der Länge $dx = 1/\sqrt{n\,p\,q}$ von jeweils

$$x_k = \frac{k - n\,p}{\sqrt{n\,p\,q}}$$

bis $x_{k+1} = x_k + dx$ zugeordnet. Über diesen Intervallen sind jeweils die konstanten W'keitswerte $y_k = \binom{n}{k} p^k q^{n-k}/dx$ der Binomial-Verteilung dargestellt. Der Skalierungsfaktor $1/dx$ sorgt dabei dafür, dass die Fläche $y_k \cdot dx$ eines Balkens (mit Breite dx) genau die W'keit $P(S_n = k) = \binom{n}{k} p^k q^{n-k}$ ist.

Mit wachsendem n konvergiert die diskrete $Bi(n, p)$-Verteilung offensichtlich gegen die kontinuierliche $N(0, 1)$-Verteilung:



Folgerungen 4.10:

Die spezielle Form der Moivre-Laplace-Näherung in Satz 4.7 wird in äquivalente Aussagen umgeschrieben. Mit

$$P\Big(a \le \frac{S_n - n\,p}{\sqrt{n\,p\,q}} \le b\Big) \;=\; P\Big(\underbrace{n\,p + a\,\sqrt{n\,p\,q}}_{a'} \le S_n \le \underbrace{n\,p + b\,\sqrt{n\,p\,q}}_{b'}\Big)$$

oder auch (mit der mittleren Häufigkeit $\bar{X}_n = S_n/n$)

$$P\Big(a \le \frac{S_n - n\,p}{\sqrt{n\,p\,q}} \le b\Big) \;=\; P\Big(a \le \frac{\sqrt{n}\,(\bar{X}_n - p)}{\sqrt{p\,q}} \le b\Big) \;=$$

$$P\Big(\underbrace{a\,\sqrt{\frac{p\,q}{n}}}_{A} \le \bar{X}_n - p \le \underbrace{b\,\sqrt{\frac{p\,q}{n}}}_{B}\Big) = P\Big(\underbrace{p + a\,\sqrt{\frac{p\,q}{n}}}_{A'} \le \bar{X}_n \le \underbrace{p + b\,\sqrt{\frac{p\,q}{n}}}_{B'}\Big)$$

folgen unmittelbar allgemeine Approximationen für binomial-verteilte Variablen S_n und die verwandten mittleren Häufigkeiten $\bar{X}_n = S_n/n$. Aus der Moivre-Laplace-Näherung

$$\boxed{P\Big(a \le \frac{S_n - n\,p}{\sqrt{n\,p\,q}} \le b\Big) \;=\; \frac{1}{\sqrt{2\,\pi}}\,\int_a^b e^{-x^2/2}\,dx \;+\; O\Big(\frac{1}{\sqrt{n}}\Big)}$$

folgt mit $a' = n\,p + a\,\sqrt{n\,p\,q}$, $b' = n\,p + b\,\sqrt{n\,p\,q}$, also $a = \frac{a' - n\,p}{\sqrt{n\,p\,q}}$, $b = \frac{b' - n\,p}{\sqrt{n\,p\,q}}$:

$$\boxed{P(a' \le S_n \le b') \;=\; \frac{1}{\sqrt{2\,\pi}}\,\int_{\frac{a'-n\,p}{\sqrt{n\,p\,q}}}^{\frac{b'-n\,p}{\sqrt{n\,p\,q}}} e^{-x^2/2}\,dx \;+\; O\Big(\frac{1}{\sqrt{n}}\Big)\;.}$$

Mit $A = a\,\sqrt{\frac{p\,q}{n}}$, $B = b\,\sqrt{\frac{p\,q}{n}}$, also $a = A\,\sqrt{\frac{n}{p\,q}}$, $b = B\,\sqrt{\frac{n}{p\,q}}$:

$$\boxed{P(A \le \bar{X}_n - p \le B) \;=\; \frac{1}{\sqrt{2\,\pi}}\,\int_{A\,\sqrt{\frac{n}{p\,q}}}^{B\,\sqrt{\frac{n}{p\,q}}} e^{-x^2/2}\,dx \;+\; O\Big(\frac{1}{\sqrt{n}}\Big)\;.}$$

Mit $A = A' - p$, $B = B' - p$:

$$\boxed{P(A' \le \bar{X}_n \le B') \;=\; \frac{1}{\sqrt{2\,\pi}}\,\int_{(A'-p)\,\sqrt{\frac{n}{p\,q}}}^{(B'-p)\,\sqrt{\frac{n}{p\,q}}} e^{-x^2/2}\,dx \;+\; O\Big(\frac{1}{\sqrt{n}}\Big)\;.}$$

Die Integrale können z.B. mit der Wertetabelle der Funktion Φ auf Seite 81 numerisch ermittelt werden:

$$\int_a^b e^{-x^2/2}\,dx = \int_0^b e^{-x^2/2}\,dx \;-\; \int_0^a e^{-x^2/2}\,dx$$

$$= \frac{1}{2}\Big(\underbrace{\int_{-b}^{b} e^{-x^2/2}\,dx}_{\Phi(b)\text{ für } b\geq 0} - \underbrace{\int_{-a}^{a} e^{-x^2/2}\,dx}_{\Phi(a)\text{ für } a\geq 0}\Big)$$

$$\Rightarrow \quad \boxed{\frac{1}{\sqrt{2\,\pi}}\int_a^b e^{-x^2/2}\,dx = \frac{1}{2}\Big(\mathrm{sign}(b)\,\Phi(|b|) - \mathrm{sign}(a)\,\Phi(|a|)\Big).}$$

Bemerkung und Warnung 4.11: $\boxed{\downarrow 2.7.07}$

Für großes n liefert Moivre-Laplace gute Näherungen, falls die zu berechnenden W'keiten nicht allzu klein sind. Man sollte aber nicht erwarten, dass die Näherungen (außer bei riesigem n) auf mehr als einige wenige Dezimalstellen genau sind. Sind die zu berechnenden W'keiten klein, so macht sich der additive Fehler $O(1/\sqrt{n}\,)$ stark bemerkbar.

Beispiel: Betrachte eine unfaire Münze mit $p = P(„Kopf“) = 0.6$. Die Variable S_n = „Anzahl der Köpfe bei n Würfen“ ist $Bi(n,p)$-verteilt. Wir vergleichen die exakte Formel

$$P(a' \leq S_n \leq b') = \sum_{k=a'}^{b'} \binom{n}{k} p^k\, q^{n-k}$$

mit der Moivre-Laplace-Näherung

$$P(a' \leq S_n \leq b') \;=\; \frac{1}{\sqrt{2\,\pi}} \int_{\frac{a'-n\,p}{\sqrt{n\,p\,q}}}^{\frac{b'-n\,p}{\sqrt{n\,p\,q}}} e^{-x^2/2}\,dx \;+\; O\Big(\frac{1}{\sqrt{n}}\Big)\;.$$

Für $n = 100$:

$$P(55 \leq S_n \leq 70) \;=\; \underbrace{0.8541...}_{\textit{exakt}} \;\approx\; \underbrace{0.8256...}_{\textit{Moivre}}\;,\; \textit{Fehler} \;\approx\; 0.0285\;.$$

Für $n = 10\,000$:

$$P(5950 \leq S_n \leq 6100) \;=\; \underbrace{0.8286...}_{\textit{exakt}} \;\approx\; \underbrace{0.8256...}_{\textit{Moivre}}\;,\; \textit{Fehler} \;\approx\; 0.0030\;.$$

Die Fehler verhalten sich in der Tat wie $O(1/\sqrt{n}\,)$: sie fallen etwa um den Faktor $1/10$, wenn n um den Faktor 100 wächst.

Achtung: werden kleine W'keiten $\ll 1$ approximiert, macht sich der additive Fehler der Moivre-Laplace-Näherung bemerkbar. Z.B., für $n = 10\,000$:

$$P(5500 \leq S_n \leq 5700) \;=\; \underbrace{5.62\cdot 10^{-10}}_{\textit{exakt}} \;\approx\; \underbrace{4.57\cdot 10^{-10}}_{\textit{Moivre}}\;.$$

Bemerkung 4.12: *Wählt man in den Folgerungen 4.10 speziell $B = -A = \epsilon > 0$, so erhält man*

$$P(-\epsilon \leq \bar{X}_n - p \leq \epsilon) \;=\; \frac{1}{\sqrt{2\,\pi}} \int_{-\epsilon\,\sqrt{\frac{n}{p\,q}}}^{\epsilon\,\sqrt{\frac{n}{p\,q}}} e^{-x^2/2}\,dx \;+\; O\Big(\frac{1}{\sqrt{n}}\Big)$$

$$= \; \Phi\Big(\epsilon\,\sqrt{\frac{n}{p\,q}}\,\Big) \;+\; O\Big(\frac{1}{\sqrt{n}}\Big)\;.$$

Damit folgt

$$\lim_{n\to\infty} P(|\bar{X}_n - p| \leq \epsilon) = 1\;,$$

wie bereits durch das Gesetz der großen Zahl vorausgesagt wurde: der Erwartungswert $\mathrm{E}\,(X) = p$ der Bernoulli-Variable $X : \Omega \mapsto \{0,1\}$ kann als empirischer Mittelwert $\bar{X}_n$ häufiger Wiederholungen X_i mit großer W'keit („Sicherheit“) S bestimmt werden. Das Gesetz der großen Zahl 4.1 liefert für das Bernoulli-Experiment X mit Erwartungswert $\mu = p$ und Varianz $\sigma^2 = p\,q$

$$S \;=\; P(|\bar{X}_n - p| < \epsilon) \;\geq\; 1 - \frac{p\,q}{\epsilon^2\,n}\;,$$

dies wird durch Moivre-Laplace verfeinert zu

$$S \;=\; P(|\bar{X}_n - p| \leq \epsilon) \;=\; \Phi\Big(\epsilon\,\sqrt{\frac{n}{p\,q}}\,\Big) \;+\; O\Big(\frac{1}{\sqrt{n}}\Big)\;.$$

Damit haben wir nun eine bessere Antwort auf die Frage

> *„Wie oft muss das Experiment X wiederholt werden, um den Erwartungswert $\mathrm{E}\,(X) = p$ durch den Mittelwert $\bar{X}_n = \frac{1}{n}\sum_i X_i$ unabhängiger Messungen X_i mit Sicherheit $S = P(|\bar{X}_n - p| < \epsilon)$ auf eine absolute Genauigkeit ϵ festlegen zu können?“*

Das Gesetz der großen Zahl liefert die Abschätzung (siehe auch die Beispiele 4.5 und 4.6)

$$n \;\geq\; \frac{1}{4\,\epsilon^2}\,\frac{1}{1-S} \;\geq\; \frac{p\,q}{\epsilon^2}\,\frac{1}{1-S}\;,$$

über Moivre-Laplace erhalten wir aus $S \approx \Phi(\epsilon\,\sqrt{\frac{n}{pq}}\,)$ mit der Umkehrfunktion Φ^{-1} nun die realistischeren kleineren Approximationen

$$n \;\approx\; \frac{p\,q}{\epsilon^2}\,(\Phi^{-1}(S))^2$$

bzw.

$$n \;\geq\; \frac{1}{4\,\epsilon^2}\,(\Phi^{-1}(S))^2 \;\geq\; \frac{p\,q}{\epsilon^2}\,(\Phi^{-1}(S))^2$$

(wir können immer $p\,q \leq 1/4$ benutzen, falls die Größenordnung von p, q nicht bekannt ist). Zum Vergleich für verschiedene vorgegebene Sicherheiten S:

S	$\frac{1}{1-S}$	$(\Phi^{-1}(S))^2$
0.5	2	0.455
0.683	3.2	1.002
0.9	10	2.706
0.954	21.7	4.019
0.99	100	6.635
0.999	1 000	10.828

Fazit: Moivre-Laplace erleichtert das Leben (liefert realistische Abschätzungen), wenn man mit großer Sicherheit den Parameter p empirisch bestimmen möchte.

Beispiel 4.13: Wir betrachten noch einmal Beispiel 4.6: wie oft muss ein Würfel geworfen werden, um $p = P(\text{„Sechs“})$ mit einer Sicherheit von 97.5% auf eine absolute Genauigkeit $\epsilon = 1/1000$ festlegen zu können? Mit der relativen Häufigkeit $\bar{X}_n$ der bei n Würfen beobachteten „Sechsen“ ist n so zu bestimmen, dass

$$P\Big(|\bar{X}_n - p| < \frac{1}{1000}\Big) \geq 0.975 =: S$$

gilt. Nach der vorigen Bemerkung 4.12 gilt

$$S \;\leq\; P\Big(|\bar{X}_n - p| < \epsilon\Big) \;\approx\; \Phi\Big(\epsilon\,\sqrt{\frac{n}{p\,q}}\,\Big) \quad\Rightarrow\quad n \;\geq\; \frac{p\,q}{\epsilon^2}\,(\Phi^{-1}(S))^2\,.$$

Die Wertetabelle für Φ auf Seite 81 kann zur Bestimmung von Werten der Umkehrfunktion Φ^{-1} benutzt werden. Man findet $\Phi(2.241) \approx 0.975 = S$, also muss in guter Näherung

$$n \;\geq\; \frac{p\,q}{\epsilon^2}\,(2.241)^2$$

gelten. Für $\epsilon = 1/1000$ folgt <u>mit der zusätzlichen Annahme</u> eines annähernd fairen Würfels $p \approx 1/6$, $q \approx 5/6$, dass

$$n \;\geq\; 697\,512\,.\,25$$

Würfe durchzuführen sind (vergleiche mit den in Beispiel 4.6 gefundenen Werten). <u>Ohne die Zusatzannahme</u> $p \approx 1/6$, $q \approx 5/6$ kann man wegen $p\,q = p\,(1-p) \leq 1/4$ die Genauigkeit ϵ mit der Sicherheit $S = 0.975$ <u>in jedem Fall</u> für

$$n \;\geq\; \frac{1}{4\,\epsilon^2}\,(2.241)^2 \approx 1\,255\,520\,.\,25$$

Würfe garantieren.

4.3 Der Zentrale Grenzwertsatz

↓4.7.07

Der letzte Abschnitt zeigte, dass eine $\mathrm{Bi}(n,p)$-verteilte Variable S_n sich für großes n durch eine normalverteilte Variable annähern lässt. Die Binomial-Verteilung ist die Verteilung einer Summenvariable: $S_n = \sum_{n=1}^n X_i$ ist die Anzahl der Erfolge bei unabhängigen Wiederholungen X_i eines Bernoulli-Experiments $X : \Omega \mapsto \{0,1\}$. Es stellt sich nun heraus, dass dies nur ein Spezialfall eines viel allgemeineren Gesetzes ist: jede Summe von unabhängigen Wiederholungen X_i eines *beliebigen* Experiments X nähert sich für großes n einer Normalverteilung an, egal ob das Experiment ein Bernoulli-Experiment ist oder einer (eventuell beliebig komplizierten) Verteilung gehorcht!
Diese bemerkenswerte Eigenschaft erlaubt in vielen Fällen, praktische Anwendungen auch dann quantitativ zu erfassen, wenn nichts über den unterliegenden Zufallsmechanismus bekannt ist. Nach häufigen Wiederholungen weiß man, dass Summen und Mittelwerte sich approximativ normalverteilt verhalten (wenn man nur genügend oft wiederholt hat).

Satz 4.14: (Zentraler Grenzwertsatz)

Sei $X_1, X_2, \ldots$ ein Familie unabhängiger Zufallsvariablen über einem gemeinsamen Stichprobenraum $(\Omega, \mathcal{E}, P)$. Alle Variablen haben die selbe Verteilung, der gemeinsame Erwartungswert sei $\mu = \mathrm{E}\,(X_1) = \mathrm{E}\,(X_2) = \ldots$, die gemeinsame Streuung sei $\sigma = \sigma(X_1) = \sigma(X_2) = \ldots$.

Für $S_n = \sum_{k=1}^n X_k$ gilt:

$$\lim_{n\to\infty} P\Big(a \le \frac{S_n - n\,\mu}{\sqrt{n}\,\sigma} \le b\Big) \;=\; \frac{1}{\sqrt{2\,\pi}} \int_a^b e^{-x^2/2}\,dx\ .$$

Die Konvergenz ist gleichmäßig in a und b. Der Approximationsfehler für großes n ist $O(1/\sqrt{n}\,)$:

$$P\Big(a \le \frac{S_n - n\,\mu}{\sqrt{n}\,\sigma} \le b\Big) \;=\; \frac{1}{\sqrt{2\,\pi}} \int_a^b e^{-x^2/2}\,dx \;+\; O\Big(\frac{1}{\sqrt{n}}\Big)\ .$$

Beweisskizze: (nur eine sehr grobe Andeutung)

Zu einer Zufallsvariable $X : \Omega \mapsto \mathbb{R}$ definiert man die sogenannte „Fourier-Transformierte“ $\hat{X}(\alpha) = \mathrm{E}\,(e^{i\,\alpha\,X})$ (mit $i = \sqrt{-1}$, $\alpha \in \mathbb{R}$). Die Verteilungsfunktion F_X von X ist durch die Funktion $\hat{X}(\alpha)$ eindeutig bestimmt. Beispielsweise lässt sich für eine kontinuierliche Variable die Dichte $\rho(x) = F_X'(x)$ durch sogenannte „Fourier-Rücktransformation“

$$\rho(x) = \frac{1}{2\,\pi} \int_{-\infty}^{\infty} e^{-i\,\alpha\,x}\ \hat{X}(\alpha)\ d\alpha$$

aus $\hat{X}(\alpha)$ rekonstruieren. Die Fourier-Transformation der $N(0,1)$-Verteilung berechnet sich zu

$$\hat{X}(\alpha) = \int_{-\infty}^{\infty} e^{i\,\alpha\,x} \frac{1}{\sqrt{2\,\pi}}\; e^{-x^2/2}\; dx \;=\; e^{-\alpha^2/2}\;.$$

Es wird gezeigt, dass sich die Fourier-Transformation der Variable

$$Y_n = \frac{S_n - n\,\mu}{\sqrt{n}\,\sigma}$$

für großes n der Fourier-Transformation $e^{-\alpha^2/2}$ der $N(0,1)$-Verteilung annähert. Hieraus folgt dann, dass sich auch die Verteilung von Y_n der $N(0,1)$-Verteilung annähert.
Es gilt

$$\mathrm{E}\,(e^{i\,\alpha\,Y_n}) = \mathrm{E}\left(e^{i\,\alpha\,\frac{1}{\sqrt{n}\,\sigma}\sum_k (X_k-\mu)}\right) = \mathrm{E}\left(\prod_{k=1}^{n} e^{i\,\alpha\,\frac{X_k-\mu}{\sqrt{n}\,\sigma}}\right)\;.$$

Da mit den X_k auch die Variablen $e^{i\,\alpha\,(X_k-\mu)/(\sqrt{n}\,\sigma)}$ unabhängig sind, gilt nach Satz 2.53 die Produktzerlegung des Erwartungswerts:

$$\mathrm{E}\,(e^{i\,\alpha\,Y_n}) = \prod_{k=1}^{n} \mathrm{E}\left(e^{i\,\alpha\,\frac{X_k-\mu}{\sqrt{n}\,\sigma}}\right) = \left(\mathrm{E}\left(e^{i\,\alpha\,\frac{X_1-\mu}{\sqrt{n}\,\sigma}}\right)\right)^n\;.$$

Hierbei wurde verwendet, dass alle Variablen die selbe Verteilung haben. Durch Reihenentwicklung von exp und ln folgt:

$$\ln\left(\mathrm{E}\,(e^{i\,\alpha\,Y_n})\right) \;=\; n\;\ln\left(\mathrm{E}\left(e^{i\,\alpha\,\frac{X_1-\mu}{\sqrt{n}\,\sigma}}\right)\right)$$

$$=\; n\;\ln\left(\mathrm{E}\left(1 \;+\; \frac{1}{1!}\frac{i\,\alpha}{\sqrt{n}\,\sigma}(X_1-\mu) \;+\; \frac{1}{2!}\left(\frac{i\,\alpha}{\sqrt{n}\,\sigma}(X_1-\mu)\right)^2 \;+\; O\Big(\frac{1}{\sqrt{n}^{\,3}}\Big)\right)\right)$$

$$=\; n\;\ln\left(\mathrm{E}\,(1) \;+\; \frac{1}{1!}\frac{i\,\alpha}{\sqrt{n}\,\sigma}\underbrace{\mathrm{E}\,(X_1-\mu)}_{0} \;-\; \frac{1}{2!}\frac{\alpha^2}{n\,\sigma^2}\underbrace{\mathrm{E}\,((X_1-\mu)^2)}_{\sigma^2} \;+\; O\Big(\frac{1}{\sqrt{n}^{\,3}}\Big)\right)$$

$$=\; n\;\ln\left(1 - \frac{\alpha^2}{2\,n} + O\Big(\frac{1}{\sqrt{n}^{\,3}}\Big)\right) = n\left(-\frac{\alpha^2}{2\,n} + O\Big(\frac{1}{\sqrt{n}^{\,3}}\Big)\right) \;=\; -\frac{\alpha^2}{2} + O\Big(\frac{1}{\sqrt{n}}\Big)\;.$$

Es folgt, dass die Fourier-Transformierte von Y_n gegen die Fourier-Transformierte $e^{-\alpha^2/2}$ der $N(0,1)$-Verteilung konvergiert:

$$\mathrm{E}\,(e^{i\,\alpha\,Y_n}) \;=\; e^{-\alpha^2/2}\;\left(1 + O\Big(\frac{1}{\sqrt{n}}\Big)\right) \;\stackrel{n\to\infty}{\longrightarrow}\; e^{-\alpha^2/2}\;.$$

Q.E.D.

In Analogie zu den Folgerungen 4.10 ergeben sich folgende Umformulierungen. Hierbei braucht in 4.10 lediglich der Erwartungswert p und die Varianz $p\,q$ des Bernoulli-Experiments durch einen allgemeinen Erwartungswert μ bzw. eine allgemeine Varianz σ^2 ersetzt zu werden:

Folgerungen 4.15:

Mit

$$P\Big(a \le \frac{S_n - n\,\mu}{\sqrt{n}\,\sigma} \le b\Big) \;=\; P\Big(\underbrace{n\,\mu + a\,\sqrt{n}\,\sigma}_{a'} \le S_n \le \underbrace{n\,\mu + b\,\sqrt{n}\,\sigma}_{b'}\Big)$$

oder auch (mit den Mittelwerten $\bar{X}_n = S_n/n$)

$$P\Big(a \le \frac{S_n - n\,\mu}{\sqrt{n}\,\sigma} \le b\Big) \;=\; P\Big(a \le \frac{\sqrt{n}\,(\bar{X}_n - \mu)}{\sigma} \le b\Big) \;=$$

$$P\Big(\underbrace{a\,\frac{\sigma}{\sqrt{n}}}_{A} \le \bar{X}_n - \mu \le \underbrace{b\,\frac{\sigma}{\sqrt{n}}}_{B}\Big) = P\Big(\underbrace{\mu + a\,\frac{\sigma}{\sqrt{n}}}_{A'} \le \bar{X}_n \le \underbrace{\mu + b\,\frac{\sigma}{\sqrt{n}}}_{B'}\Big)$$

folgen unmittelbar allgemeine Approximationen für Summenvariablen S_n und Mittelwerte $\bar{X}_n = S_n/n$ für großes n. Aus dem Zentralen Grenzwertsatz

$$\boxed{P\Big(a \le \frac{S_n - n\,\mu}{\sqrt{n}\,\sigma} \le b\Big) \;\approx\; \frac{1}{\sqrt{2\,\pi}} \int_a^b e^{-x^2/2}\,dx}$$

folgt mit $a' = n\,\mu + a\,\sqrt{n}\,\sigma$, $b' = n\,\mu + b\,\sqrt{n}\,\sigma$, also $a = \frac{a' - n\,\mu}{\sqrt{n}\,\sigma}$, $b = \frac{b' - n\,\mu}{\sqrt{n}\,\sigma}$:

$$\boxed{P(a' \le S_n \le b') \;\approx\; \frac{1}{\sqrt{2\,\pi}} \int_{\frac{a' - n\,\mu}{\sqrt{n}\,\sigma}}^{\frac{b' - n\,\mu}{\sqrt{n}\,\sigma}} e^{-x^2/2}\,dx\;.}$$

Mit $A = a\,\frac{\sigma}{\sqrt{n}}$, $B = b\,\frac{\sigma}{\sqrt{n}}$, also $a = A\,\frac{\sqrt{n}}{\sigma}$, $b = B\,\frac{\sqrt{n}}{\sigma}$:

$$\boxed{P(A \le \bar{X}_n - \mu \le B) \;\approx\; \frac{1}{\sqrt{2\,\pi}} \int_{A\,\frac{\sqrt{n}}{\sigma}}^{B\,\frac{\sqrt{n}}{\sigma}} e^{-x^2/2}\,dx\;.}$$

Mit $A = A' - \mu$, $B = B' - \mu$:

$$\boxed{P(A' \le \bar{X}_n \le B') \;\approx\; \frac{1}{\sqrt{2\,\pi}} \int_{(A' - \mu)\,\frac{\sqrt{n}}{\sigma}}^{(B' - \mu)\,\frac{\sqrt{n}}{\sigma}} e^{-x^2/2}\,dx\;.}$$

Die Integrale können z.B. mit der Wertetabelle der Funktion Φ auf Seite 81 numerisch ermittelt werden:

$$\boxed{\frac{1}{\sqrt{2\,\pi}} \int_a^b e^{-x^2/2}\,dx = \frac{1}{2}\Big(\mathrm{sign}(b)\,\Phi(|b|) - \mathrm{sign}(a)\,\Phi(|a|)\Big).}$$

In Analogie zur Bemerkung 4.12 folgt

Bemerkung 4.16: *Wählt man in den Folgerungen 4.15* $B = -A = \epsilon > 0$, *so erhält man für großes* n

$$P(-\epsilon \le \bar{X}_n - \mu \le \epsilon) \;\approx\; \frac{1}{\sqrt{2\,\pi}} \int_{-\epsilon\,\frac{\sqrt{n}}{\sigma}}^{\epsilon\,\frac{\sqrt{n}}{\sigma}} e^{-x^2/2}\,dx \;=\; \Phi\Big(\epsilon\,\frac{\sqrt{n}}{\sigma}\Big)$$

und damit

$$\lim_{n\to\infty} P(|\bar{X}_n - \mu| \le \epsilon) = 1\;,$$

wie bereits durch das Gesetz der großen Zahl vorausgesagt wurde: der Erwartungswert $\mathrm{E}\,(X) = \mu$ *einer* beliebigen *Variable* X *kann als empirischer Mittelwert* $\bar{X}_n$ *häufiger Wiederholungen* X_i *mit großer W'keit („Sicherheit“)* S *bestimmt werden. Das Gesetz der großen Zahl 4.1 liefert für das Experiment* X *mit Erwartungswert* μ *und Varianz* σ^2

$$P(|\bar{X}_n - \mu| < \epsilon) \;\ge\; 1 - \frac{\sigma^2}{\epsilon^2\,n}\;,$$

dies wird durch den Zentralen Grenzwertsatz verfeinert zu

$$P(|\bar{X}_n - \mu| \le \epsilon) \;\approx\; \Phi\Big(\epsilon\,\frac{\sqrt{n}}{\sigma}\Big)\;.$$

Bemerkung 4.17: *Führt man in den Folgerungen 4.15*

$$P(A' \le \bar{X}_n \le B') \;\approx\; \frac{1}{\sqrt{2\,\pi}} \int_{(A'-\mu)\,\frac{\sqrt{n}}{\sigma}}^{(B'-\mu)\,\frac{\sqrt{n}}{\sigma}} e^{-x^2/2}\,dx$$

die Substitution $x = (r - \mu)\,\frac{\sqrt{n}}{\sigma}$ *durch, so ergibt sich*

$$P(A' \le \bar{X}_n \le B') \;\approx\; \frac{1}{\sqrt{2\,\pi}\;(\sigma/\sqrt{n}\,)} \int_{A'}^{B'} e^{-\frac{(r-\mu)^2}{2\,\sigma^2/n}}\,dr\;.$$

Dieses Integral entspricht den W'keiten einer $N(\mu, \sigma^2/n)$*-verteilten Variable (siehe Abschnitt 2.4.7), d.h., es gilt die Aussage:*

Der Mittelwert von n unabhängigen Messungen einer beliebigen Zufallsvariable mit Erwartungswert μ und Streuung σ genügt für großes n approximativ einer Normalverteilung mit Erwartungswert μ und Streuung $\sigma/\sqrt{n}$.

Bemerkung 4.18: *In der Praxis ergibt sich allerdings das Problem: wie groß muss n denn nun wirklich sein, damit die Annahme einer Normalverteilung realistisch ist? Es gilt folgende Abschätzung nach Berry und Esseen, falls man neben dem Erwartungswert μ und der Varianz $\sigma^2 = \mathrm{E}\left((X-\mu)^2\right)$ der Variablen auch noch die Größe $\kappa = \mathrm{E}\left(|X-\mu|^3\right)$ kennt. Für beliebiges $b \in \mathbb{R}$ gilt:*

$$\left| \; P\Big(\frac{S_n - n\,\mu}{\sqrt{n}\,\sigma} \leq b\Big) \; - \; \frac{1}{\sqrt{2\,\pi}} \int_{-\infty}^{b} e^{-x^2/2}\,dx \; \right| \; \leq \; \frac{0.8\;\kappa}{\sigma^3\,\sqrt{n}} \; .$$

Mit $\int_a^b = \int_{-\infty}^b - \int_{-\infty}^a$ folgt durch die Dreiecksungleichung

$$\left| \; P\Big(a < \frac{S_n - n\,\mu}{\sqrt{n}\,\sigma} \leq b\Big) \; - \; \frac{1}{\sqrt{2\,\pi}} \int_{a}^{b} e^{-x^2/2}\,dx \; \right| \; \leq \; \frac{1.6\;\kappa}{\sigma^3\,\sqrt{n}} \; .$$

Hieran sieht man deutlich den Approximationsfehler $O(1/\sqrt{n}\,)$ des Zentralen Grenzwertsatzes.

Bemerkung 4.19: *In der Praxis kennt man in der Regel weder den Erwartungswert μ, noch die Streuung σ, noch die Größe κ. Hier setzt nun die Statistik ein: es seien n unabhängige Messungen $x_1, \ldots, x_n$ einer Zufallsvariablen X gegeben. Man schätzt dann μ, σ, κ ab durch*

$$\bar{x}_n = \frac{1}{n} \sum_{i=1}^{n} x_i \; , \quad \bar{v}_n = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x}_n)^2 \; ,$$

$$\bar{k}_n = \frac{n}{(n-1)\,(n-2)} \sum_{i=1}^{n} |x_i - \bar{x}_n|^3 \; .$$

Für großes n liegt $\bar{x}_n$ mit großer W'keit in der Nähe von μ (das haben wir in diesem Kapitel gezeigt). Man nennt $\bar{x}_n$ einen „Schätzwert“ für μ. Analog kann man zeigen, dass $\bar{v}_n$ eine Schätzung für σ^2 liefert (siehe auch Übungsaufgabe 48). Heuristisch liefert $\bar{k}_n$ eine Näherung für κ.

Beispiel 4.20: (Eine typische „Fehlerrechnung“ der Physik)
Eine physikalische Größe X wird mehrfach gemessen, die Größe (oder der Messprozess) unterliegt einer stochastischen Störung. Es ergeben sich die $n = 10$ Werte

$$99, \; 100, \; 103, \; 99, \; 102, \; 100, \; 98, \; 101, \; 102, \; 101.$$

Der „wahre“ Wert (der Erwartungswert μ von X) soll mit einer Sicherheit von $S = 0.954$ abgeschätzt werden.

Als Mittelwert und Schätzung der Varianz berechnet man

$$\bar{x}_n \;=\; \frac{1}{n} \sum_{i=1}^{n} x_i \;=\; 100.5 \; , \quad \bar{v}_n \;=\; \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x}_n)^2 \;=\; 2.5 \; .$$

Wir schätzen die unbekannte Streuung σ von X durch die empirische Streuung $\bar{v}_n$ ab: mit $\sigma^2 \approx \bar{v}_n$ ergibt sich $\sigma \approx \sqrt{\bar{v}_n} \approx 1.6$. Setzt man der Messung $\bar{x}_n$ entsprechende Mittelwertvariable als normalverteilt voraus, so liefert Bemerkung 4.16 die Sicherheit

$$S \;=\; P(|\bar{x}_n - \mu| \le \epsilon) \;=\; \Phi\Big(\epsilon\,\frac{\sqrt{n}}{\sigma}\Big)\;.$$

Es ist die absolute Genauigkeit ϵ gefragt, mit der $\bar{x}_n$ den gesuchten Erwartungswert μ approximiert:

$$\epsilon \;=\; \frac{\sigma}{\sqrt{n}}\,\Phi^{-1}(S) \;\approx\; \frac{1.6}{\sqrt{10}}\,\Phi^{-1}(S) \;\approx\; 0.51 \cdot \Phi^{-1}(S)\;.$$

Für $S = 0.954$ gilt $\Phi^{-1}(S) \approx 2.00$; es folgt $\epsilon \approx 1.0$. Also gilt mit 95.4%-iger Sicherheit (in Physikernotation):

$$\mu \;=\; 100.5 \;\pm\; 1.0\;.$$

Für $S = 0.683$ gilt $\Phi^{-1}(S) \approx 1.00$; es folgt $\epsilon \approx 0.51$. Also gilt mit 68.3%-iger Sicherheit:

$$\mu \;=\; 100.5 \;\pm\; 0.51\;.$$

Diese Aussagen sind natürlich dadurch etwas verfälscht, dass man σ nicht wirklich kennt, sondern empirisch über $\sigma^2 \approx \bar{v}_n$ approximiert hat.

Man findet in der Tat bei Naturwissenschaftlern und Ingenieuren folgendes Rezept für „Fehlerrechnungen“:

> Bei n Messungen $x_1, \ldots, x_n$ einer Größe X gilt $X \;=\; \bar{x}_n \;\pm\; \dfrac{2\,\sigma_n}{\sqrt{n}}$ mit
>
> $$\bar{x}_n \;=\; \frac{1}{n}\sum_{i=1}^{n} x_i\;, \qquad \sigma_n^2 \;=\; \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x}_n)^2\;.$$

Unter der Approximation des Zentralen Grenzwertsatzes gilt diese Aussage mit 95.4%-iger Sicherheit (der Faktor 2 in $2\,\sigma_n/\sqrt{n}$ entspricht $\Phi^{-1}(0.954) \;\approx\; 2.00$).

Zusatzfrage: wieviele Messungen muss man machen, um μ mit 95.4%iger Sicherheit auf die Genauigkeit $\epsilon = 0.1$ festzulegen? Nun ist

$$S \;=\; P(|\bar{x}_n - \mu| \le \epsilon) \;=\; \Phi\Big(\epsilon\,\frac{\sqrt{n}}{\sigma}\Big)$$

nach n aufzulösen:

$$n \;=\; \frac{\sigma^2}{\epsilon^2}\,(\Phi^{-1}(S))^2 \;\approx\; \frac{\sigma_n^2}{\epsilon^2}\,(\Phi^{-1}(S))^2 \;\approx\; \frac{2.5}{0.1^2}\,2.0^2 \;=\; 1\,000\;.$$