Ein wichtiger Spezialfall das Zentralen Grenzwertsatzes besteht darin, dass die auftretenden Zufallsgrößen Bernoulli-verteilt sind.

## Korollar 116 (Grenzwertsatz von de Moivre)

*$X_1, \ldots, X_n$ seien unabhängige Bernoulli-verteilte Zufallsvariablen mit gleicher Erfolgswahrscheinlichkeit $p$. Dann gilt für die Zufallsvariable $H_n$ mit*

$$H_n := X_1 + \ldots + X_n$$

*für $n \geq 1$, dass die Verteilung der Zufallsvariablen*

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

*für $n \to \infty$ gegen die Standardnormalverteilung konvergiert.*

Beweis:

Die Behauptung folgt unmittelbar aus dem Zentralen Grenzwertsatz, da
$\mu = \frac{1}{n}\mathbb{E}[H_n] = p$ und $\sigma^2 = \frac{1}{n}\mathrm{Var}[H_n] = p(1-p)$. $\square$

**Bemerkung**

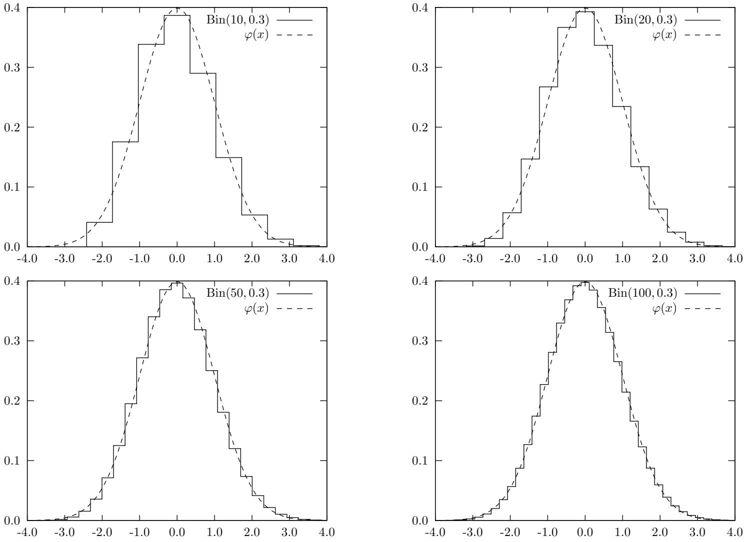
Wenn man $X_1, \ldots, X_n$ als Indikatorvariablen für das Eintreten eines Ereignisses $A$ bei $n$ unabhängigen Wiederholungen eines Experimentes interpretiert, dann gibt $H_n$ die absolute Häufigkeit von $A$ an.

## 4.1 Normalverteilung als Grenzwert der Binomialverteilung

Korollar 116 ermöglicht, die Normalverteilung als Grenzwert der Binomialverteilung aufzufassen. Die folgende Aussage ist eine Konsequenz von Korollar 116:

### Korollar 117

*Sei* $H_n \sim \mathrm{Bin}(n, p)$ *eine binomialverteilte Zufallsvariable. Die Verteilung von* $H_n/n$ *konvergiert gegen* $\mathcal{N}(p, p(1-p)/n)$ *für* $n \to \infty$.




# Vergleich von Binomial- und Normalverteilung

$\mathrm{Bin}(n, 0.3)$ bei $0.3n$ zentriert, mit $\sqrt{0.3 \cdot 0.7n}$ horizontal gestaucht und vertikal gestreckt

Historisch gesehen entstand Korollar 116 vor Satz 115.

Für den Fall $p = 1/2$ wurde Korollar 116 bereits von Abraham de Moivre (1667–1754) bewiesen. De Moivre war gebürtiger Franzose, musste jedoch aufgrund seines protestantischen Glaubens nach England fliehen. Dort wurde er unter anderem Mitglied der Royal Society, erhielt jedoch niemals eine eigene Professur.

Die allgemeine Formulierung von Korollar 116 geht auf Pierre Simon Laplace (1749–1827) zurück. Allerdings vermutet man, dass die Lösung des allgemeinen Falls $p \neq 1/2$ bereits de Moivre bekannt war.

## 4.2 Elementarer Beweis des Grenzwertsatzes von de Moivre für $p = 1/2$

Wir betrachten die Wahrscheinlichkeit $\Pr[a \le H_{2n}^* \le b]$ für $p = 1/2$ und $a, b \in \mathbb{R}$ mit $a \le b$. Wenn die Verteilung von $H_{2n}^*$, wie in Korollar 116 angegeben, gegen $\mathcal{N}(0,1)$ konvergiert, so sollte $\Pr[a \le H_{2n}^* \le b] \approx \int_a^b \varphi(t)\,\mathrm{d}\,t$ für genügend große $n$ gelten.
Wir schreiben $f(n) \sim_\infty g(n)$ für $\lim_{n\to\infty} f(n)/g(n) = 1$, wollen also zeigen:

$$\Pr[a \le H_{2n}^* \le b] \sim_\infty \int_a^b \varphi(t)\,\mathrm{d}\,t.$$

Da für $H_{2n} \sim \mathrm{Bin}(2n, 1/2)$ gilt, dass $\mathbb{E}[H_{2n}] = n$ und $\mathrm{Var}[H_{2n}] = n/2$ ist, erhalten wir

$$H_{2n}^* = \frac{H_{2n} - n}{\sqrt{n/2}},$$

und es folgt

$$\begin{aligned}\Pr[a \leq H_{2n}^* \leq b] &= \Pr[n + a\sqrt{n/2} \leq H_{2n} \leq n + b\sqrt{n/2}] \\ &= \sum_{i \in I_n} \Pr[H_{2n} = n + i]\end{aligned}$$

für $I_n := \{z \in \mathbb{Z} \mid a\sqrt{n/2} \leq z \leq b\sqrt{n/2}\}$. Damit ist

$$\Pr[a \leq H_{2n}^* \leq b] = \sum_{i \in I_n} \underbrace{\binom{2n}{n+i} \cdot \left(\frac{1}{2}\right)^{2n}}_{=:p_{n,i}}.$$

Es gilt

$$\max_i p_{n,i} \leq p_n^* := \binom{2n}{n} \cdot \left(\frac{1}{2}\right)^{2n} = \frac{(2n)!}{(n!)^2} \cdot \left(\frac{1}{2}\right)^{2n},$$

und mit der Stirling'schen Approximation für $n!$

$$p_n^* \sim_\infty \frac{(2n)^{2n} \cdot e^{-2n} \cdot \sqrt{2\pi \cdot 2n}}{(n^n \cdot e^{-n} \cdot \sqrt{2\pi n})^2} \cdot \left(\frac{1}{2}\right)^{2n} = \frac{1}{\sqrt{\pi n}}.$$

Ersetzen wir nun die $p_{n,i}$ durch $p_n^*$, so entsteht dabei ein Fehler, den wir mit $q_{n,i} := \frac{p_{n,i}}{p_n^*}$ bezeichnen.

Für $i > 0$ gilt

$$\begin{aligned} q_{n,i} = \frac{\binom{2n}{n+i} \cdot \left(\frac{1}{2}\right)^{2n}}{\binom{2n}{n} \cdot \left(\frac{1}{2}\right)^{2n}} &= \frac{(2n)! \cdot n! \cdot n!}{(n+i)! \cdot (n-i)! \cdot (2n)!} \\ = \frac{\prod_{j=0}^{i-1}(n-j)}{\prod_{j=1}^{i}(n+j)} &= \prod_{j=1}^{i} \frac{n-j+1}{n+j} = \prod_{j=1}^{i} \left(1 - \frac{2j-1}{n+j}\right). \end{aligned}$$

Wegen der Symmetrie der Binomialkoeffizienten gilt $q_{n,-i} = q_{n,i}$, womit auch der Fall $i < 0$ abgehandelt ist.

Man macht sich leicht klar, dass $1 - 1/x \leq \ln x \leq x - 1$ für $x > 0$ gilt. Damit schließen wir, dass

$$
\begin{aligned}
\ln\left(\prod_{j=1}^{i}\left(1 - \frac{2j-1}{n+j}\right)\right) &= \sum_{j=1}^{i} \ln\left(1 - \frac{2j-1}{n+j}\right) \\
&\leq -\sum_{j=1}^{i} \frac{2j-1}{n+j} \leq -\sum_{j=1}^{i} \frac{2j-1}{n+i} \\
&= -\frac{i(i+1)-i}{n+i} = -\frac{i^2}{n} + \frac{i^3}{n(n+i)} \\
&= -\frac{i^2}{n} + \mathcal{O}\left(\frac{1}{\sqrt{n}}\right),
\end{aligned}
$$

da $i = \mathcal{O}(\sqrt{n})$ für $i \in I_n$.

Ebenso erhalten wir

$$\begin{aligned}\ln\left(\prod_{j=1}^{i}\left(1-\frac{2j-1}{n+j}\right)\right) &\geq \sum_{j=1}^{i}\left(1-\left(1-\frac{2j-1}{n+j}\right)^{-1}\right)\\ &= \sum_{j=1}^{i}\frac{-2j+1}{n-j+1} \geq -\sum_{j=1}^{i}\frac{2j-1}{n-i}\\ &= -\frac{i^2}{n-i} = -\frac{i^2}{n}-\mathcal{O}\left(\frac{1}{\sqrt{n}}\right).\end{aligned}$$

Zusammen haben wir

$$e^{-\frac{i^2}{n-i}} = e^{-\frac{i^2}{n}-\mathcal{O}\left(\frac{1}{\sqrt{n}}\right)} \leq q_{n,i} \leq e^{-\frac{i^2}{n}+\mathcal{O}\left(\frac{1}{\sqrt{n}}\right)}$$

Wegen $e^{\pm\mathcal{O}(1/\sqrt{n})} = 1 \pm o(1)$ folgt daraus $q_{n,i} \sim_\infty e^{-i^2/n}$.

Damit schätzen wir nun $\Pr[a \le H_{2n}^* \le b]$ weiter ab:

$$\Pr[a \le H_{2n}^* \le b] = \sum_{i \in I_n} p_n^* \cdot q_{n,i} \;\sim_\infty\; \underbrace{\frac{1}{\sqrt{\pi n}} \cdot \sum_{i \in I_n} e^{-i^2/n}}_{=:S_n}.$$

Mit $\delta := \sqrt{2/n}$ können wir die Summe $S_n$ umschreiben zu

$$S_n = \frac{1}{\sqrt{2\pi}} \cdot \sum_{i \in I_n} \delta e^{-(i\delta)^2 \cdot \frac{1}{2}}.$$

Diese Summe entspricht einer Näherung für $\int_a^b \varphi(t)\,\mathrm{d}\,t = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-t^2/2}\,\mathrm{d}\,t$ durch Aufteilung der integrierten Fläche in Balken der Breite $\delta$. Für $n \to \infty$ konvergiert die Fläche der Balken gegen das Integral, d. h. $S_n \sim_\infty \int_a^b \varphi(t)\,\mathrm{d}\,t$.

*q. e. d.*

## 4.3 Verschiedene Approximationen der Binomialverteilung

Sei $H_n \sim \mathrm{Bin}(n,p)$ eine binomialverteilte Zufallsvariable mit der Verteilungsfunktion $F_n$. Für $n \to \infty$ gilt

$$\begin{aligned} F_n(t) &= \Pr[H_n/n \le t/n] \\ &\to \Phi\left(\frac{t/n - p}{\sqrt{p(1-p)/n}}\right) = \Phi\left(\frac{t-np}{\sqrt{p(1-p)n}}\right). \end{aligned}$$

Wir können $F_n$ somit für große $n$ durch $\Phi$ approximieren. Diese Approximation ist in der Praxis deshalb von Bedeutung, da die Auswertung der Verteilungsfunktion der Binomialverteilung für große $n$ sehr aufwändig ist, während für die Berechnung der Normalverteilung effiziente numerische Methoden vorliegen.

## Beispiel 118

Wenn man die Wahrscheinlichkeit berechnen möchte, mit der bei $10^6$ Würfen mit einem idealen Würfel mehr als 500500-mal eine gerade Augenzahl fällt, so muss man eigentlich folgenden Term auswerten:

$$T := \sum_{i=5{,}005\cdot 10^5}^{10^6} \binom{10^6}{i} \left(\frac{1}{2}\right)^{10^6}.$$

Dies ist numerisch kaum effizient möglich.

Die numerische Integration der Dichte $\varphi$ der Normalverteilung ist hingegen relativ einfach. Auch andere Approximationen der Verteilung $\Phi$, beispielsweise durch Polynome, sind bekannt. Entsprechende Funktionen werden in zahlreichen Softwarebibliotheken als „black box" angeboten.

## Beispiel

*Mit der Approximation durch die Normalverteilung erhalten wir*

$$\begin{aligned} T &\approx 1 - \Phi\left(\frac{5{,}005 \cdot 10^5 - 5 \cdot 10^5}{\sqrt{2{,}5 \cdot 10^5}}\right) \\ &= 1 - \Phi\left(\frac{5 \cdot 10^2}{5 \cdot 10^2}\right) \\ &= 1 - \Phi(1) \approx 0{,}1573 \,. \end{aligned}$$

Bei der Approximation der Binomialverteilung mit Hilfe von Korollar 116 führt man oft noch eine so genannte Stetigkeitskorrektur durch. Zur Berechnung von $\Pr[X \leq x]$ für $X \sim \mathrm{Bin}(n,p)$ setzt man

$$\Pr[X \leq x] \approx \Phi\left(\frac{x + 0{,}5 - np}{\sqrt{np(1-p)}}\right)$$

statt

$$\Pr[X \leq x] \approx \Phi\left(\frac{x - np}{\sqrt{np(1-p)}}\right)$$

an.

Der Korrekturterm läßt sich in der Histogramm-Darstellung der Binomialverteilung veranschaulichen. Die Binomialverteilung wird dort durch Balken angegeben, deren Fläche in etwa der Fläche unterhalb der Dichte $\varphi$ von $\mathcal{N}(0,1)$ entspricht. Wenn man die Fläche der Balken mit „$X \le x$“ durch das Integral von $\varphi$ approximieren möchte, so sollte man bis zum Ende des Balkens für „$X = x$“ integrieren und nicht nur bis zur Mitte. Dafür sorgt der Korrekturterm $0{,}5$.

## Approximationen für die Binomialverteilung

- Approximation durch die Poisson-Verteilung: $\mathrm{Bin}(n,p)$ wird approximiert durch $\mathrm{Po}(np)$. Diese Approximation funktioniert sehr gut für seltene Ereignisse, d. h. wenn $np$ sehr klein gegenüber $n$ ist. Als Faustregel fordert man $n \geq 30$ und $p \leq 0{,}05$.
- Approximation durch die Chernoff-Schranken: Bei der Berechnung der tails der Binomialverteilung liefern diese Ungleichungen meist sehr gute Ergebnisse. Ihre Stärke liegt darin, dass es sich bei den Schranken nicht um Approximationen, sondern um echte Abschätzungen handelt. Dies ist vor allem dann wichtig, wenn man nicht nur numerische Näherungen erhalten möchte, sondern allgemeine Aussagen über die Wahrscheinlichkeit von Ereignissen beweisen möchte.

- Approximation durch die Normalverteilung: Als Faustregel sagt man, dass die Verteilungsfunktion $F_n(t)$ von $\mathrm{Bin}(n, p)$ durch

$$F_n(t) \approx \Phi((t - np)/\sqrt{p(1-p)n})$$

approximiert werden kann, wenn $np \geq 5$ und $n(1-p) \geq 5$ gilt.

# Kapitel III Induktive Statistik

## 1. Einführung

Das Ziel der induktiven Statistik besteht darin, aus gemessenen Zufallsgrößen auf die zugrunde liegenden Gesetzmäßigkeiten zu schließen. Im Gegensatz dazu spricht man von deskriptiver Statistik, wenn man sich damit beschäftigt, große Datenmengen verständlich aufzubereiten, beispielsweise durch Berechnung des Mittelwertes oder anderer abgeleiteter Größen.

# 2. Schätzvariablen

Wir betrachten die Anzahl $X$ von Lesezugriffen auf eine Festplatte bis zum ersten Lesefehler und nehmen an, dass $\Pr[X = i] = (1-p)^{i-1}p$, setzen also für $X$ eine geometrische Verteilung an. Dahinter verbirgt sich die Annahme, dass bei jedem Zugriff unabhängig und mit jeweils derselben Wahrscheinlichkeit $p$ ein Lesefehler auftreten kann.

Unter diesen Annahmen ist die Verteilung der Zufallsvariablen $X$ eindeutig festgelegt. Allerdings entzieht sich der numerische Wert des Parameters $p$ noch unserer Kenntnis. Dieser soll daher nun empirisch geschätzt werden. Statt $p$ können wir ebensogut $\mathbb{E}[X]$ bestimmen, da wir daraus nach den Eigenschaften der geometrischen Verteilung $p$ mittels $p = \frac{1}{\mathbb{E}[X]}$ berechnen können.

Dazu betrachten wir $n$ baugleiche Platten und die zugehörigen Zufallsvariablen $X_i$ (für $1 \leq i \leq n$), d. h. wir zählen für jede Platte die Anzahl von Zugriffen bis zum ersten Lesefehler. Die Zufallsvariablen $X_i$ sind dann unabhängig und besitzen jeweils dieselbe Verteilung wie $X$. Wir führen also viele Kopien eines bestimmten Zufallsexperiments aus, um Schlüsse auf die Gesetzmäßigkeiten des einzelnen Experiments ziehen zu können. Dies ist das Grundprinzip der induktiven Statistik. Die $n$ Messungen heißen Stichproben, und die Variablen $X_i$ nennt man Stichprobenvariablen.

**Grundprinzip statistischer Verfahren**

Wir erinnern an das Gesetz der großen Zahlen (Satz 63) bzw. den Zentralen Grenzwertsatz (Satz 115). Wenn man ein Experiment genügend oft wiederholt, so nähert sich der Durchschnitt der Versuchsergebnisse immer mehr dem Verhalten an, das man „im Mittel“ erwarten würde. Je mehr Experimente wir also durchführen, umso genauere und zuverlässigere Aussagen können wir über den zugrunde liegenden Wahrscheinlichkeitsraum ableiten. Auf diesem Grundprinzip beruhen alle statistischen Verfahren.

Um $\mathbb{E}[X]$ empirisch zu ermitteln, bietet es sich an, aus den Zufallsvariablen $X_i$ das arithmetische Mittel $\overline{X}$ zu bilden, das definiert ist durch

$$\overline{X} := \frac{1}{n} \sum_{i=1}^{n} X_i.$$

Es gilt

$$\mathbb{E}[\overline{X}] = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}[X_i] = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}[X] = \mathbb{E}[X].$$

$\overline{X}$ liefert uns also im Mittel den gesuchten Wert $\mathbb{E}[X]$. Da wir $\overline{X}$ zur Bestimmung von $\mathbb{E}[X]$ verwenden, nennen wir $\overline{X}$ einen Schätzer für den Erwartungswert $\mathbb{E}[X]$. Wegen der obigen Eigenschaft ist $\overline{X}$ sogar ein so genannter erwartungstreuer Schätzer.

## Definition 119

Gegeben sei eine Zufallsvariable $X$ mit der Dichte $f(x;\theta)$. Eine Schätzvariable oder kurz Schätzer für den Parameter $\theta$ der Dichte von $X$ ist eine Zufallsvariable, die aus mehreren (meist unabhängigen und identisch verteilten) Stichprobenvariablen zusammengesetzt ist. Ein Schätzer $U$ heißt erwartungstreu, wenn gilt

$$\mathbb{E}[U] = \theta.$$

**Bemerkung:**

Die Größe $\mathbb{E}[U - \theta]$ nennt man Bias der Schätzvariablen $U$. Bei erwartungstreuen Schätzvariablen ist der Bias gleich Null.

Der Schätzer $\overline{X}$ ist also ein erwartungstreuer Schätzer für den Erwartungswert von $X$.
Ein wichtiges Maß für die Güte eines Schätzers ist die mittlere quadratische Abweichung, kurz MSE für mean squared error genannt. Diese berechnet sich durch $MSE := \mathbb{E}[(U - \theta)^2]$. Wenn $U$ erwartungstreu ist, so folgt $MSE = \mathbb{E}[(U - \mathbb{E}[U])^2] = \mathrm{Var}[U]$.

## Definition 120

Wenn die Schätzvariable $A$ eine kleinere mittlere quadratische Abweichung besitzt als die Schätzvariable $B$, so sagt man, dass $A$ effizienter ist als $B$.
Eine Schätzvariable heißt konsistent im quadratischen Mittel, wenn $\mathrm{MSE} \to 0$ für $n \to \infty$ gilt. Hierbei bezeichne $n$ den Umfang der Stichprobe.

Für $\overline{X}$ erhalten wir wegen der Unabhängigkeit von $X_1, \ldots, X_n$

$$\begin{aligned} MSE = \mathrm{Var}[\overline{X}] &= \mathrm{Var}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] \\ &= \frac{1}{n^2}\sum_{i=1}^{n} \mathrm{Var}[X_i] = \frac{1}{n}\mathrm{Var}[X]. \end{aligned}$$

Bei jeder Verteilung mit endlicher Varianz folgt $MSE = \mathcal{O}(1/n)$ und somit $MSE \to 0$ für $n \to \infty$. Der Schätzer $\overline{X}$ ist also konsistent.

Aus der Konsistenz von $\overline{X}$ im quadratischen Mittel können wir mit Hilfe des Satzes von Chebyshev (siehe Satz 61) folgende Konsequenz ableiten. Sei $\varepsilon > 0$ beliebig, aber fest. Dann gilt

$$\Pr[|\overline{X} - \theta| \geq \varepsilon] = \Pr[|\overline{X} - \mathbb{E}[X]| \geq \varepsilon] \leq \frac{\mathrm{Var}[\overline{X}]}{\varepsilon^2} \to 0$$

für $n \to \infty$. Für genügend große $n$ liegen also die Werte von $\overline{X}$ beliebig nahe am gesuchten Wert $\theta = \mathbb{E}[X]$. Diese Eigenschaft nennt man auch schwache Konsistenz, da sie aus der Konsistenz im quadratischen Mittel folgt.