

Folien zur Übung Induktive Statistik

Michael Westermann

Zufallsexperiment und Ereignisraum

Zufallsexperiment (ZE), besser: Zufallsvorgang

Ein *Zufallsvorgang* zeichnet sich durch folgende Eigenschaften aus:

- Die Versuchsausgänge sind wohldefiniert;
- Er ist beliebig oft (unter gleichen Bedingungen) wiederholbar;
- Die einzelnen Wiederholungen sind unabhängig;
- Der Ausgang ist weder zu beeinflussen noch vorhersagbar;
- Bei häufigen Durchführungen ist eine gewisse Regelmäßigkeiten zu beobachten.

Beispiele: Werfen einer Münze, Geburt eines Kindes, Ausgang einer Fußballbegegnung, Wartezeit an einer Kasse, Besucher bei einer Theaterveranstaltung.

Stichprobenraum/Ereignisraum

Unter dem *Stichprobenraum* versteht man alle, sich gegenseitig ausschließenden Ereignisse/Ausgänge eines Zufallsexperiments/-vorgangs:

$$\Omega = \{\omega_1, \omega_2, \omega_3, \dots, \omega_n\}.$$

Der Umfang von Ω kann endlich oder (überabzählbar) unendlich groß sein.

Beispiele:

Einmaliges Werfen einer Münze	$\Omega = \{K, W\}$	endlicher Stichprobenraum
Anzahl der Spiele bis zum ersten Sieg	$\Omega = \{1, 2, 3, \dots\}$	abzählbar unendlicher Stichprobenraum
Wartezeit an einer Kasse	$\Omega = \{0, \infty\}$ jede positive reelle Zahl	überabzählbar unendlicher Stichprobenraum

Grundaufgaben der Kombinatorik

Die Kombinatorik ist u.a. Grundlage für die Berechnung von Wahrscheinlichkeiten (*Kapitel 3*) nach dem klassischen Wahrscheinlichkeitsbegriff. Außerdem wird sie auch im Rahmen der Stichprobentheorie (*Kapitel 6*) zur Ermittlung der Anzahl aller möglichen Zufallsstichproben aus einer Grundgesamtheit verwendet.

Gesucht ist die Anzahl aller möglichen Ausgänge eines Zufallsexperiments, also der Umfang des Stichprobenraums Ω . Folgende Informationen über die Konstruktion des Zufallsexperiments werden dabei benötigt:

- Handelt es sich um eine Anordnung aller n Elemente oder um eine Auswahl von $i \leq n$ Elementen?
- Ist die Reihenfolge der Auswahl von Bedeutung oder nicht, d.h. ist

$$\{a, b\} \neq \{b, a\} \quad \text{bzw.} \quad \{a, b\} = \{b, a\}?$$

- Handelt es sich um eine Experiment mit Wiederholungen oder ohne Wiederholungen?

Grundaufgaben der Kombinatorik						
Anordnung			Auswahl			
Permutation			mit Reihenfolge	ohne Reihenfolge		
			Variation	Kombination		
o.W.	$P = n! = n \cdot (n-1) \cdot (n-2) \dots 2 \cdot 1$		o.W.	$V = \frac{n!}{(n-i)!}$	o.W.	$K = \binom{n}{i} = \frac{n!}{i! \cdot (n-i)!}$
m.W.	$P_W = \frac{n!}{n_1! \cdot n_2! \cdot \dots \cdot n_k!}$		m.W.	$V_W = n^i$	m.W.	$K_W = \binom{n+i-1}{i}$

Einleitung

(Zufalls-)Ereignis

Jede Teilmenge des Stichprobenraums heißt *Ereignis*. Es gibt folgende Arten von Ereignissen:

- sicheres Ereignis Ω
- unmögliches Ereignis \emptyset
- Elementarereignis ω_i
- zusammengesetzte Ereignisse

Beispiel: Einmaliges Werfen eines Würfels

sicheres Ereignis	$\Omega = \{1, 2, 3, 4, 5, 6\}$	
unmögliches Ereignis	\emptyset	
Elementarereignisse	$\{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}$	
zusammengesetzte Ereignisse	$A =$ alle Zahlen größer als 2	$A = \{3, 4, 5, 6\}$
	$B =$ alle ungeraden Zahlen	$B = \{1, 3, 5\}$

Darstellung von Ereignissen

Venn-Diagramme

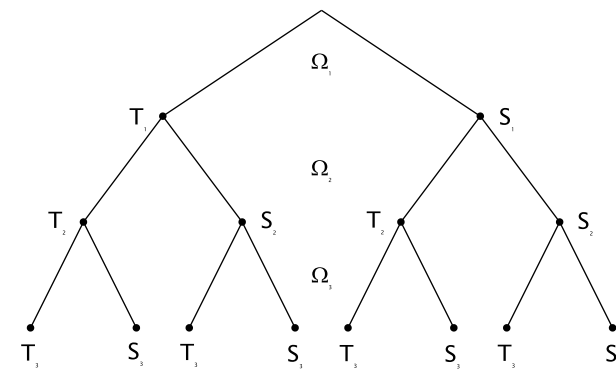
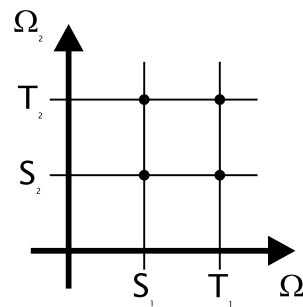
Venn-Diagramme werden zur Darstellung von Ereignissen, Ereignisoperationen und -relationen verwendet.

Produktmengen

Produktmengen werden zur Darstellung wiederholter Zufallsexperimente verwendet. Bei mehr als 2 Durchführungen verwendet man zur graphischen Darstellung der Produktmengen ein Baumdiagramm.

Beispiel: Das Zufallsexperiment lautet *Geburt eines Kindes*

1. Kind	$\Omega_1 = \{S_1, T_1\}$
2. Kind	$\Omega_2 = \{S_2, T_2\}$
3. Kind	$\Omega_3 = \{S_3, T_3\}$



$$\Omega_1 \times \Omega_2 = \{(S_1S_2), (S_1T_2), (T_1S_2), (T_1T_2)\}$$

$$\Omega_1 \times \Omega_2 \times \Omega_3 = \{(S_1S_2S_3), (S_1S_2T_3), (S_1T_2S_3), (S_1T_2T_3), (T_1S_2S_3), (T_1S_2T_3), (T_1T_2S_3), (T_1T_2T_3)\}$$

Potenzmenge

Die *Potenzmenge* $P(\Omega)$ ist die Menge aller möglichen Teilmengen des Stichprobenraums, d.h. die Menge aller möglichen Ereignisse.

Sigma-Algebra

Da die Potenzmenge u.U. überabzählbar unendlich groß werden kann, wird eine geeignete Teilmenge hiervon gebildet, welche den Definitionsbereich für ein Wahrscheinlichkeitsmaß darstellt. Diese Teilmenge heißt σ -Algebra.

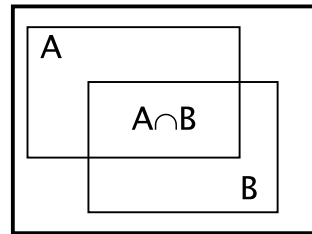
Wahrscheinlichkeitsbegriff

Wahrscheinlichkeitsbegriff			
axiomatisch	interpretierend		
Grundlage der Wahrscheinlichkeitsrechnung	wie gelangt man zu einer konkreten Wahrscheinlichkeit		
	subjektiv	objektiv	
Axiome von Kolmogoroff	Hypothesenwahrscheinlichkeiten	a priori	<ul style="list-style-type: none"> • klassisch • geometrisch
	Maß der Überzeugtheit von einem Ereigniss	a posteriori	<ul style="list-style-type: none"> • statistischer Wkt.-begriff • Grenzwert der relativen Häufigkeiten

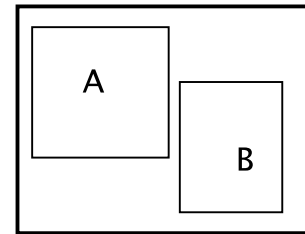
Rechnen mit Wahrscheinlichkeiten

Additionssatz $P(A \cup B)$ [mindestens ein Ereignis]

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

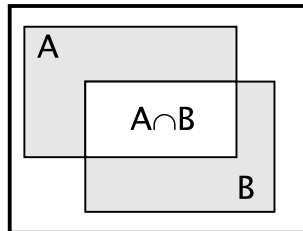


$$P(A \cup B) = P(A) + P(B) - \underbrace{P(A \cap B)}_{=0}$$



Additionssatz $P(A \cup B)$ [genau ein Ereignis]

$$P(A \cup B)_{\text{excl}} = P(A) + P(B) - 2 \cdot P(A \cap B) = P(A \cup B) - P(A \cap B)$$



Additionssatz

450 Studenten müssen im Rahmen ihres Grundstudiums 2 Klausuren in Statistik schreiben. Von den 450 Personen haben 400 die erste Klausur bestanden (*Ereignis A*) und 430 haben die zweite Klausur bestanden (*Ereignis B*). 390 Studenten haben beide Klausuren bestanden (*Ereignis $A \cap B$*).

$$\Rightarrow P(A) = \frac{400}{450}; \quad P(B) = \frac{430}{450}; \quad P(A \cap B) = \frac{390}{450}$$

Wie wahrscheinlich ist es, dass ein zufällig ausgewählter Student

- die erste Klausur nicht bestanden hat?

$$P(\bar{A}) = 1 - P(A) = \frac{50}{450} = 0,1111.$$

- mindestens eine Klausur bestanden hat?

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = \frac{400}{450} + \frac{430}{450} - \frac{390}{450} = \frac{440}{450} = 0,9778.$$

- genau eine Klausur bestanden hat?

$$P(A \cup B)_{\text{excl}} = P(A \cup B) - P(A \cap B) = \frac{440}{450} - \frac{390}{450} = \frac{50}{450} = 0,1111.$$

Bedingte Wahrscheinlichkeiten

1000 befragte Personen lassen sich wie folgt nach CDU-Anhängern (*Ereignis A*) und BILD-Lesern (*Ereignis B*) einteilen:

	<i>B</i>	\bar{B}	
<i>A</i>	400	200	600
\bar{A}	100	300	400
	500	500	1000

$$P(A) = \frac{600}{1000} = 0,6; \quad P(B) = \frac{500}{1000} = 0,5; \quad P(A \cap B) = \frac{400}{1000} = 0,4$$

Wie wahrscheinlich ist es

- einen CDU-Anhänger anzutreffen, wenn die BILD-Leser betrachtet werden?

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{0,4}{0,5} = 0,8.$$

- einen BILD-Leser anzutreffen, wenn die CDU-Anhänger betrachtet werden?

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{0,4}{0,6} = 0,6667.$$

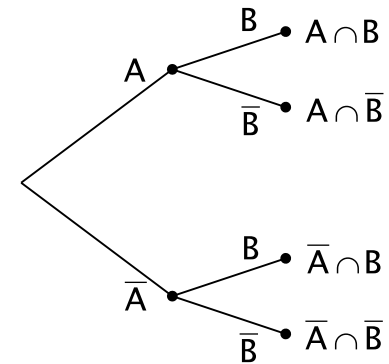
Multiplikationssatz

Aus einer Urne mit 5 weißen und 5 roten Kugeln wird zwei Mal ohne Zurücklegen gezogen. Wie groß ist die Wahrscheinlichkeit, dass beide Male eine rote Kugel gezogen wird?

$$\text{Ereignis A: 1. Kugel rot } P(A) = \frac{5}{10} = 0,5$$

$$\text{Ereignis B: 2. Kugel rot } P(B|A) = \frac{4}{9}, \text{ wenn 1. Kugel rot}$$

$$P(B|\bar{A}) = \frac{5}{9}, \text{ wenn 1. Kugel weiss}$$



gesucht ist:

$$P(A \cap B) = P(A) \cdot P(B|A) = \frac{5}{10} \cdot \frac{4}{9} = \frac{2}{9}.$$

Auf drei Züge ausgeweitet erhält man:

$$P(A \cap B \cap C) = P(A) \cdot P(B|A) \cdot P(C|A \cap B) = \frac{5}{10} \cdot \frac{4}{9} \cdot \frac{3}{8} = \frac{1}{12} = 0,0833.$$

Stochastische Unabhängigkeit

- Klausuren

$$P(A) = \frac{400}{450}; \quad P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{390/450}{430/450} = 0,9070$$

- CDU-Anhänger/BILD-Leser

$$P(A) = 0,6; \quad P(A|B) = 0,8$$

- Kugeln

$$P(B|A) = \frac{4}{9}; \quad P(B|\bar{A}) = \frac{5}{9}$$

Stochastische Unabhängigkeit liegt vor, wenn gilt:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = P(A) \quad \text{bzw.} \quad P(B|A) = \frac{P(A \cap B)}{P(A)} = P(B).$$

Daraus folgt der Multiplikationssatz für zwei Ereignisse bei Unabhängigkeit:

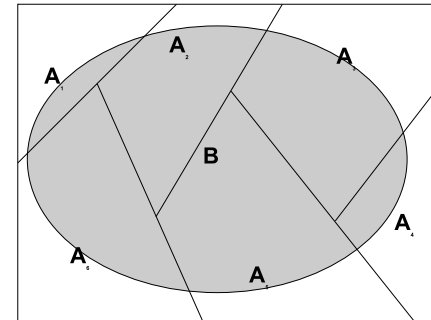
$$P(A \cap B) = P(A) \cdot P(B).$$

Satz der totalen Wahrscheinlichkeit, Theorem von Bayes

Satz der totalen Wahrscheinlichkeit

Gegeben sei eine vollständige Zerlegung von Ω in die Ereignisse A_i und ein Ereignis B .

- i) $\bigcup_{i=1}^m A_i = \Omega$,
- ii) $A_i \cap A_j = \emptyset$, für $i \neq j$,
- iii) $A_i \neq \emptyset$, für $i = 1, \dots, m$.



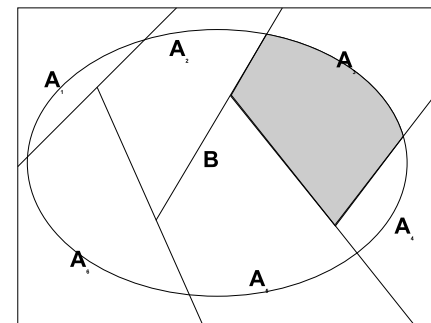
Ereignis B : $B = B \cap \Omega = B \cap (A_1 \cup A_2 \cup \dots \cup A_m) = (B \cap A_1) \cup (B \cap A_2) \cup \dots \cup (B \cap A_m)$.

Wahrscheinlichkeit für B : $P(B|A_1) \cdot P(A_1) + P(B|A_2) \cdot P(A_2) + \dots + P(B|A_m) \cdot P(A_m) = \sum_{i=1}^m P(B|A_i) \cdot P(A_i)$

Theorem von Bayes

Gesucht ist

$$\begin{aligned}
 P(A_i|B) &= \frac{P(A_i \cap B)}{P(B)} \quad \leftarrow \text{Multiplikationssatz} \\
 &= \frac{P(B|A_i) \cdot P(A_i)}{\sum_{i=1}^m P(B|A_i) \cdot P(A_i)} \quad \leftarrow \text{Satz der tot. Wkt.}
 \end{aligned}$$



Satz der totalen Wahrscheinlichkeit, Theorem von Bayes

Die Produktion von Armlehnen für Gartenstühle erfolgt auf drei Maschinen. Maschine A_1 erzeugt 10% der Gesamtproduktion, Maschine A_2 20% und Maschine A_3 70%. Der Ausschußanteil beträgt auf Maschine A_1 1%, auf A_2 3% und auf Maschine A_3 5%. Die Armlehnen werden gemischt gelagert, so daß eine dem Lager entnommene Armlehne nicht mehr einer Maschine zugeordnet werden kann.

$$\Rightarrow P(A_1) = 0,1; \quad P(A_2) = 0,2; \quad P(A_3) = 0,7; \quad P(D|A_1) = 0,01; \quad P(D|A_2) = 0,03; \quad P(D|A_3) = 0,05$$

i) Wie groß ist die Wahrscheinlichkeit, daß eine zufällig dem Lager entnommene Armlehne defekt ist (Ereignis D)?

$$P(D) = \sum_{i=1}^3 P(D|A_i) \cdot P(A_i) = 0,01 \cdot 0,1 + 0,03 \cdot 0,2 + 0,05 \cdot 0,7 = 0,042.$$

ii) Wie groß ist die Wahrscheinlichkeit, daß eine defekte Armlehne von Maschine A_3 produziert wurde?

$$P(A_3|D) = \frac{P(D|A_3) \cdot P(A_3)}{P(D)} = \frac{0,05 \cdot 0,7}{0,042} = 0,8333.$$

Drei Schützen S_1 , S_2 und S_3 schießen auf ein Ziel. Im gleichen Zeitraum gibt S_1 dreimal, S_2 doppelt so viele Schüsse ab, wie S_3 . Die Trefferquote von S_1 beträgt 0,3, die von S_2 0,6 und die von S_3 0,8. Es fällt ein Schuß, der das Ziel trifft. Wie wahrscheinlich ist es, daß dieser Schuß von S_2 abgefeuert wurde?

$$\Rightarrow P(S_1) = 3P(S_3); \quad P(S_2) = 2P(S_3); \quad P(T|S_1) = 0,3; \quad P(T|S_2) = 0,6; \quad P(T|S_3) = 0,8$$

Da nur drei Schützen am Start sind, muß gelten:

$$\begin{aligned} P(S_1) + P(S_2) + P(S_3) = 1 &\Rightarrow 3P(S_3) + 2P(S_3) + P(S_3) = 1 &\Rightarrow 6P(S_3) = 1 &\Rightarrow P(S_3) = 1/6 \\ & & &\Rightarrow P(S_1) = 3/6 = 1/2 \\ & & &\Rightarrow P(S_2) = 2/6 = 1/3 \end{aligned}$$

Wahrscheinlichkeit für Ereignis T : $P(T) = \sum_{i=1}^3 P(T|S_i) \cdot P(S_i) = 0,3 \cdot 0,5 + 0,6 \cdot 1/3 + 0,8 \cdot 1/6 = 0,4833$

$$P(S_2|T) = \frac{P(T|S_2) \cdot P(S_2)}{P(T)} = \frac{0,6 \cdot 1/3}{0,4833} = 0,4138.$$

Binomialverteilung $[X \sim B(n; \pi)]$

$$f(x) = \binom{n}{x} \pi^x (1 - \pi)^{n-x} \quad E(X) = n\pi \quad V(X) = n\pi(1 - \pi)$$

Tabelle 1: Binomialverteilung (Auszug)

n	x	0,1		0,2		0,3		0,4		0,5	
		$f(x)$	$F(x)$	$f(x)$	$F(x)$	$f(x)$	$F(x)$	$f(x)$	$F(x)$	$f(x)$	$F(x)$
1	0	0,9000	0,9000	0,8000	0,8000	0,7000	0,7000	0,6000	0,6000	0,5000	0,5000
	1	0,1000	1,0000	0,2000	1,0000	0,3000	1,0000	0,4000	1,0000	0,5000	1,0000
2	0	0,8100	0,8100	0,6400	0,6400	0,4900	0,4900	0,3600	0,3600	0,2500	0,2500
	1	0,1800	0,9900	0,3200	0,9600	0,4200	0,9100	0,4800	0,8400	0,5000	0,7500
	2	0,0100	1,0000	0,0400	1,0000	0,0900	1,0000	0,1600	1,0000	0,2500	1,0000
3	0	0,7290	0,7290	0,5120	0,5120	0,3430	0,3430	0,2160	0,2160	0,1250	0,1250
	1	0,2430	0,9720	0,3840	0,8960	0,4410	0,7840	0,4320	0,6480	0,3750	0,5000
	2	0,0270	0,9990	0,0960	0,9920	0,1890	0,9730	0,2880	0,9360	0,3750	0,8750
	3	0,0010	1,0000	0,0080	1,0000	0,0270	1,0000	0,0640	1,0000	0,1250	1,0000
4	0	0,6561	0,6561	0,4096	0,4096	0,2401	0,2401	0,1296	0,1296	0,0625	0,0625
	1	0,2916	0,9477	0,4096	0,8192	0,4116	0,6517	0,3456	0,4752	0,2500	0,3125
	2	0,0486	0,9963	0,1536	0,9728	0,2646	0,9163	0,3456	0,8208	0,3750	0,6875
	3	0,0036	0,9999	0,0256	0,9984	0,0756	0,9919	0,1536	0,9744	0,2500	0,9375
	4	0,0001	1,0000	0,0016	1,0000	0,0081	1,0000	0,0256	1,0000	0,0625	1,0000
5	0	0,5905	0,5905	0,3277	0,3277	0,1681	0,1681	0,0778	0,0778	0,0313	0,0313
	1	0,3281	0,9185	0,4096	0,7373	0,3602	0,5282	0,2592	0,3370	0,1563	0,1875
	2	0,0729	0,9914	0,2048	0,9421	0,3087	0,8369	0,3456	0,6826	0,3125	0,5000
	3	0,0081	0,9995	0,0512	0,9933	0,1323	0,9692	0,2304	0,9130	0,3125	0,8125
	4	0,0005	1,0000	0,0064	0,9997	0,0284	0,9976	0,0768	0,9898	0,1563	0,9688
	5	0,0000	1,0000	0,0003	1,0000	0,0024	1,0000	0,0102	1,0000	0,0313	1,0000

Quelle: v.d. Lippe (1999): Induktive Statistik – Formeln, Aufgaben, Klausurtraining, S. 117.

Binomialverteilung

Für ein Seminar sollen maximal 30 Teilnehmer zugelassen werden. Aus langjähriger Erfahrung weiß man, dass 10% der angemeldeten Studenten das zur Qualifikation erforderliche Grundstudium nicht schaffen, d.h. nicht teilnehmen können. Im letzten Seminar wurden daher 33 Anmeldungen zugelassen.

1. Wieviele Teilnehmer können erwartet werden?

Die Zufallsvariable X sei definiert als die Anzahl der Studenten, welche am Seminar teilnehmen. Diese Zufallsvariable ist somit binomialverteilt mit $n = 33$ (Teilnehmern) und $\pi = 0,9$ (Erfolgswahrscheinlichkeit für die Teilnahme). Es gilt somit:

$$X \sim B(33; 0,9) \quad \Rightarrow \quad E(X) = n\pi = 33 \cdot 0,9 = 29,7$$

Es können also 29,7 bzw. 30 Teilnehmer erwartet werden.

Alternativ dazu könnte man $Y = n - X$ auch definieren als Anzahl der Studenten, die das Vordiplom nicht schaffen und daher nicht am Seminar teilnehmen können. Die „Erfolgswahrscheinlichkeit“ hierfür beträgt dann $1 - \pi = 0,1$.

$$X \sim B(n; \pi) \quad \Rightarrow \quad Y = n - X \sim B(n; 1 - \pi)$$

2. Wie groß ist die Wahrscheinlichkeit, dass die maximale gewünschte Anzahl von 30 Studenten teilnimmt?

$$P(X = 30) = f(30) = \binom{33}{30} 0,9^{30} 0,1^3 = 0,2313.$$

3. Wie groß ist die Wahrscheinlichkeit, dass 3 Studenten an der maximalen Teilnehmerzahl fehlen?

$$P(X = 27) = f(27) = \binom{33}{27} 0,9^{27} 0,1^6 = 0,0644.$$

4. Wie wahrscheinlich ist es, dass das Seminar überfüllt ist?

$$\begin{aligned} P(X > 30) &= f(31) + f(32) + f(33) \\ &= 1 - P(X \leq 30) = 1 - F(30) \\ &= 1 - 0,6543 = 0,3457. \end{aligned}$$

Poissonverteilung $[X \sim P(\lambda)]$

$$f(x) = \frac{\lambda^x}{x!} e^{-\lambda} \quad E(X) = \lambda \quad V(X) = \lambda$$

Tabelle 2: Poissonverteilung (Auszug)

x	λ 1,5		2		3		4		5	
	f(x)	F(x)	f(x)	F(x)	f(x)	F(x)	f(x)	F(x)	f(x)	F(x)
0	0,2231	0,2231	0,1353	0,1353	0,0498	0,0498	0,0183	0,0183	0,0067	0,0067
1	0,3347	0,5578	0,2707	0,4060	0,1494	0,1991	0,0733	0,0916	0,0337	0,0404
2	0,2510	0,8088	0,2707	0,6767	0,2240	0,4232	0,1465	0,2381	0,0842	0,1247
3	0,1255	0,9344	0,1804	0,8571	0,2240	0,6472	0,1954	0,4335	0,1404	0,2650
4	0,0471	0,9814	0,0902	0,9473	0,1680	0,8153	0,1954	0,6288	0,1755	0,4405
5	0,0141	0,9955	0,0361	0,9834	0,1008	0,9161	0,1563	0,7851	0,1755	0,6160
6	0,0035	0,9991	0,0120	0,9955	0,0504	0,9665	0,1042	0,8893	0,1462	0,7622
7	0,0008	0,9998	0,0034	0,9989	0,0216	0,9881	0,0595	0,9489	0,1044	0,8666
8	0,0001	1,0000	0,0009	0,9998	0,0081	0,9962	0,0298	0,9786	0,0653	0,9319
9			0,0002	1,0000	0,0027	0,9989	0,0132	0,9919	0,0363	0,9682
10					0,0008	0,9997	0,0053	0,9972	0,0181	0,9863
11					0,0002	0,9999	0,0019	0,9991	0,0082	0,9945
12					0,0001	1,0000	0,0006	0,9997	0,0034	0,9980
13							0,0002	0,9999	0,0013	0,9993
14							0,0001	1,0000	0,0005	0,9998
15									0,0002	0,9999
16									0,0000	1,0000

Quelle: v.d. Lippe (1999): Induktive Statistk – Formeln, Aufgaben, Klausurtraining, S. 118.

Poissonverteilung

In einem Meldebüro finden sich täglich im Mittel 5 Personen ein, die von außerhalb in die Stadt ziehen (Zufallsvariable X). Andererseits wollen sich im Mittel 7 Personen pro Tag abmelden (Zufallsvariable Y).

1. Wie groß ist die Wahrscheinlichkeit, dass sich an einem Tag mehr als 12 Personen anmelden wollen?

Die Zufallsvariable X ist poissonverteilt mit $X \sim P(5)$. Es gilt also

$$P(X > 12) = 1 - P(X \leq 12) = 1 - F(12) = 1 - 0,9980 = 0,0020.$$

2. Wie groß ist die Wahrscheinlichkeit, dass sich an einem Tag höchstens 5 Personen abmelden wollen?

Die Zufallsvariable Y ist ebenfalls poissonverteilt mit $Y \sim P(7)$ und gesucht ist

$$P(Y \leq 5) = F(5) = 0,3007.$$

3. Wie groß ist die Wahrscheinlichkeit, dass sich mindestens 1 Person ummelden will?

Wenn zwei unabhängige Zufallsvariablen poissonverteilt sind mit λ_1 und λ_2 , so ist die Summe dieser ZV'n ebenfalls poissonverteilt mit $\lambda^* = \lambda_1 + \lambda_2$, also

$$X \sim P(\lambda_1), Y \sim P(\lambda_2) \quad \Rightarrow \quad Z = X + Y \sim P(\lambda_1 + \lambda_2)$$

Z (die Anzahl der sich ummeldenden Personen) ist somit poissonverteilt mit $Z \sim P(12)$. Gesucht wird die Wahrscheinlichkeit

$$P(Z \geq 1) = 1 - P(Z < 1) = 1 - P(Z = 0) \approx 1.$$

Oskar geht gerne angeln. Bei seinem Stammteich beträgt die Wahrscheinlichkeit $\pi = 0,2$, dass der Petri-Jünger erfolgreich ist.

1. Mit welcher Wahrscheinlichkeit ist er an fünf Tagen

- (a) genau 3 mal erfolgreich?
- (b) überhaupt nicht erfolgreich?
- (c) mehr als einmal erfolgreich?

Die Anzahl X der erfolgreichen Tage ist binomialverteilt mit $n = 5$ Versuchen und einer Erfolgswahrscheinlichkeit $\pi = 0,2$. Somit gilt $X \sim B(5; 0,2)$.

- (a) $P(X = 3) = f(3) = 0,0512$
- (b) $P(X = 0) = f(0) = 0,3277$
- (c) $P(X > 1) = 1 - P(X \leq 1) = 1 - F(1) = 1 - 0,7373 = 0,2627$

2. Wie groß ist die Wahrscheinlichkeit, dass er höchstens 4 mal vergeblich zum Teich marschiert und erst dann erfolgreich ist?

Es sei Y die Anzahl der Mißerfolge bis zum ersten Erfolg. Diese ist geometrisch verteilt mit $\pi = 0,2$, also $Y \sim G(0,2)$. Gesucht ist in diesem Fall $P(Y \leq 4)$, also

$$F(4) = 1 - (1 - 0,2)^5 = 1 - 0,85 = 0,6723.$$

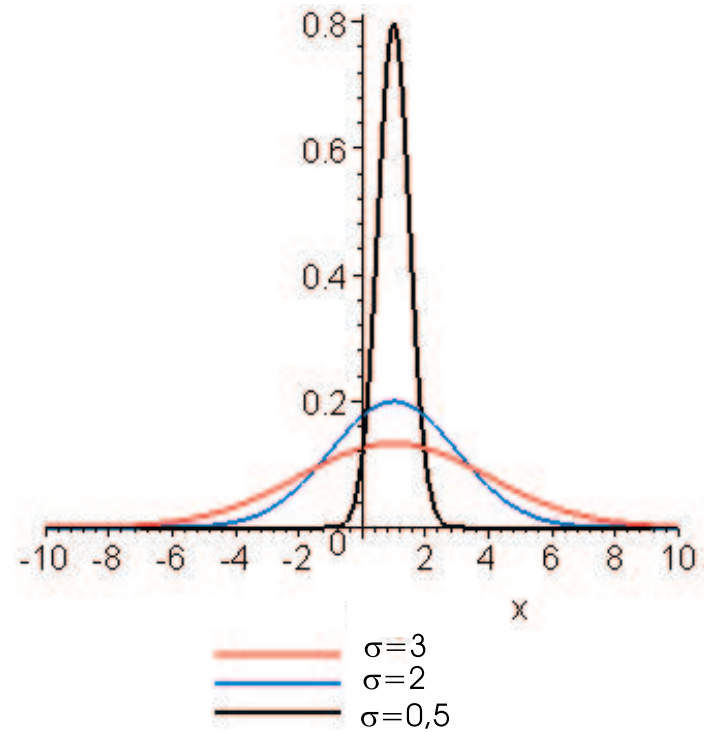
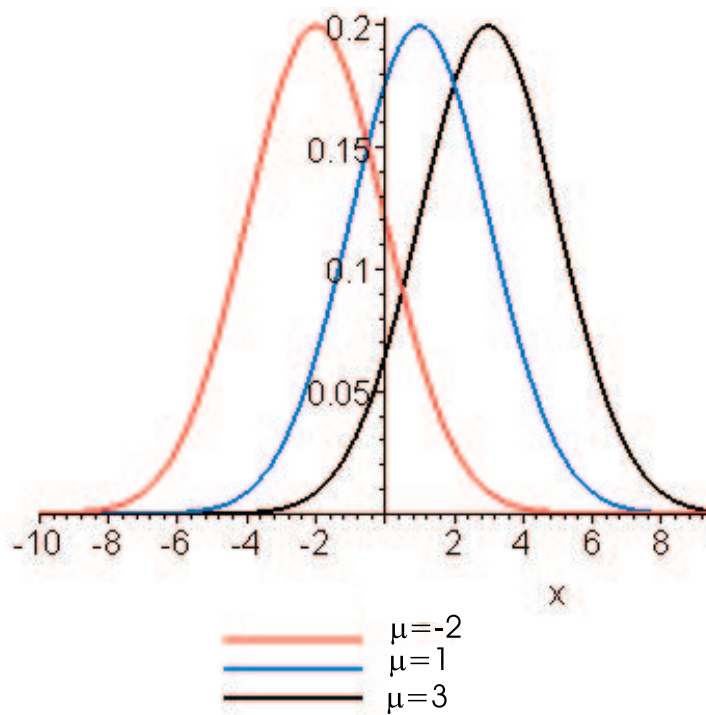
3. Eines Tages sieht Oskar staunend, dass sich eine sehr große Zahl von Anglern an seinem Teich versammelt hat. Aufgrund des Gedränges ist die Wahrscheinlichkeit für einen Erfolg pro Angler sehr klein geworden. Die mittlere Zahl der erfolgreichen Angler beträgt aber immerhin noch 8. Wie groß ist die Wahrscheinlichkeit, dass mindestens 10 Angler erfolgreich nach Hause gehen?

Die Information, dass die Anzahl der Angler sehr groß und die Erfolgswahrscheinlichkeit sehr klein geworden ist, lässt darauf schließen, dass die Zufallsvariable Z (die Anzahl der erfolgreichen Angler) poissonverteilt ist mit Erwartungswert 8, also $Z \sim P(8)$. Gesucht ist $P(Z \geq 10)$, also

$$P(Z \geq 10) = 1 - P(Z \leq 9) = 1 - F(9) = 1 - 0,7166 = 0,2834.$$

Normalverteilung $[X \sim N(\mu; \sigma^2)]$

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad E(X) = \mu \quad V(X) = \sigma^2$$



Standardnormalverteilung [$Z \sim N(0; 1)$]

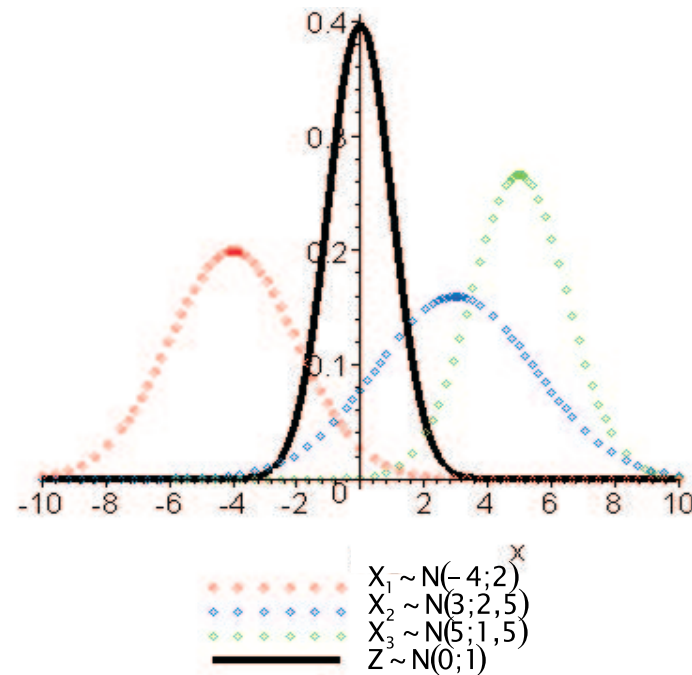
Eine normalverteilte Zufallsvariable mit Erwartungswert Null und Varianz Eins heißt *standardnormalverteilt*. Jede normalverteilte Zufallsvariable $X \sim N(\mu; \sigma^2)$ kann durch geeignete Transformation

$$Z = \frac{X - \mu}{\sigma} = -\frac{\mu}{\sigma} + \frac{1}{\sigma}X = a + bX$$

in eine standardnormalverteilte Zufallsvariable $Z \sim N(0; 1)$ überführt werden. Dabei handelt es sich um eine Lineartransformation der Zufallsvariable X mit $a = -\mu/\sigma$ und $b = 1/\sigma$. Für den Erwartungswert und die Varianz gilt dann:

$$E(Z) = a + bE(X) = -\frac{\mu}{\sigma} + \frac{1}{\sigma}\mu = 0$$

$$V(Z) = b^2V(X) = \frac{1}{\sigma^2}\sigma^2 = 1$$



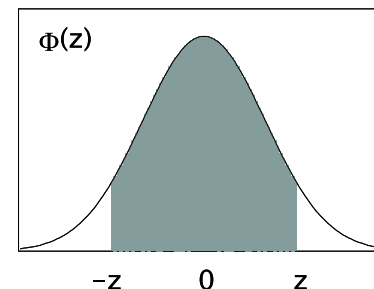
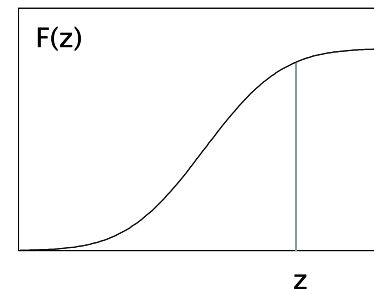
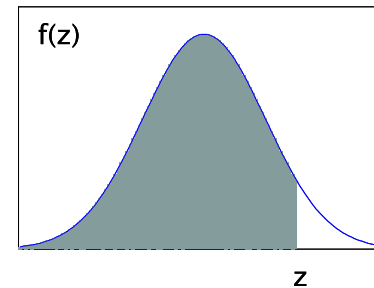
Standardnormalverteilung

$$X \sim N(\mu; \sigma^2) \Rightarrow Z = \frac{X - \mu}{\sigma} \Rightarrow Z \sim N(0; 1)$$

Tabelle 3: Standardnormalverteilung (Auszug)

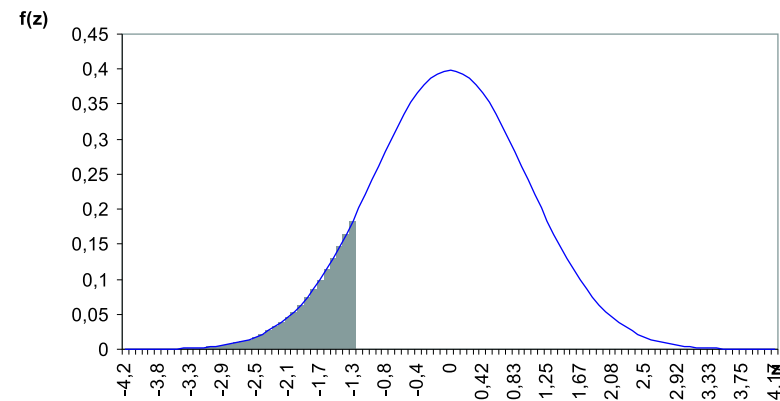
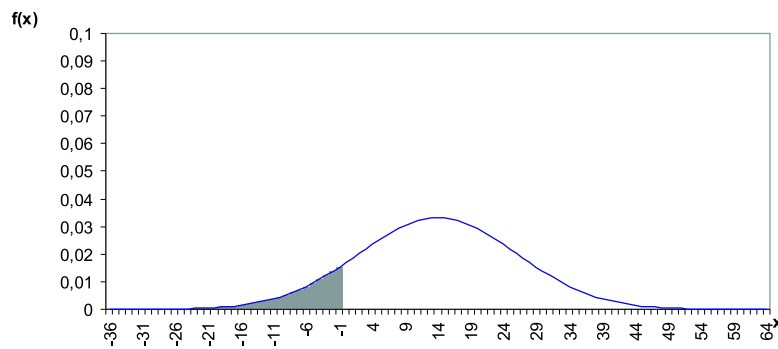
z	f(z)	F(z)	φ(z)
0,0	0,3989	0,5000	0,0000
0,1	0,3970	0,5398	0,0797
0,2	0,3910	0,5793	0,1585
0,3	0,3814	0,6179	0,2358
0,4	0,3683	0,6554	0,3108
0,5	0,3521	0,6915	0,3829
0,6	0,3332	0,7257	0,4515
0,7	0,3123	0,7580	0,5161
0,8	0,2897	0,7881	0,5763
0,9	0,2661	0,8159	0,6319
1,0	0,2420	0,8413	0,6827
1,1	0,2179	0,8643	0,7287
1,2	0,1942	0,8849	0,7699
1,3	0,1714	0,9032	0,8064
1,4	0,1497	0,9192	0,8385
1,5	0,1295	0,9332	0,8664
1,6	0,1109	0,9452	0,8904
1,7	0,0940	0,9554	0,9109
1,8	0,0790	0,9641	0,9281
1,9	0,0656	0,9713	0,9426

Quelle: v.d. Lippe (1999): Induktive Statistik – Formeln, Aufgaben, Klausurtraining, S. 119.



Beispiel zur Z-Transformation

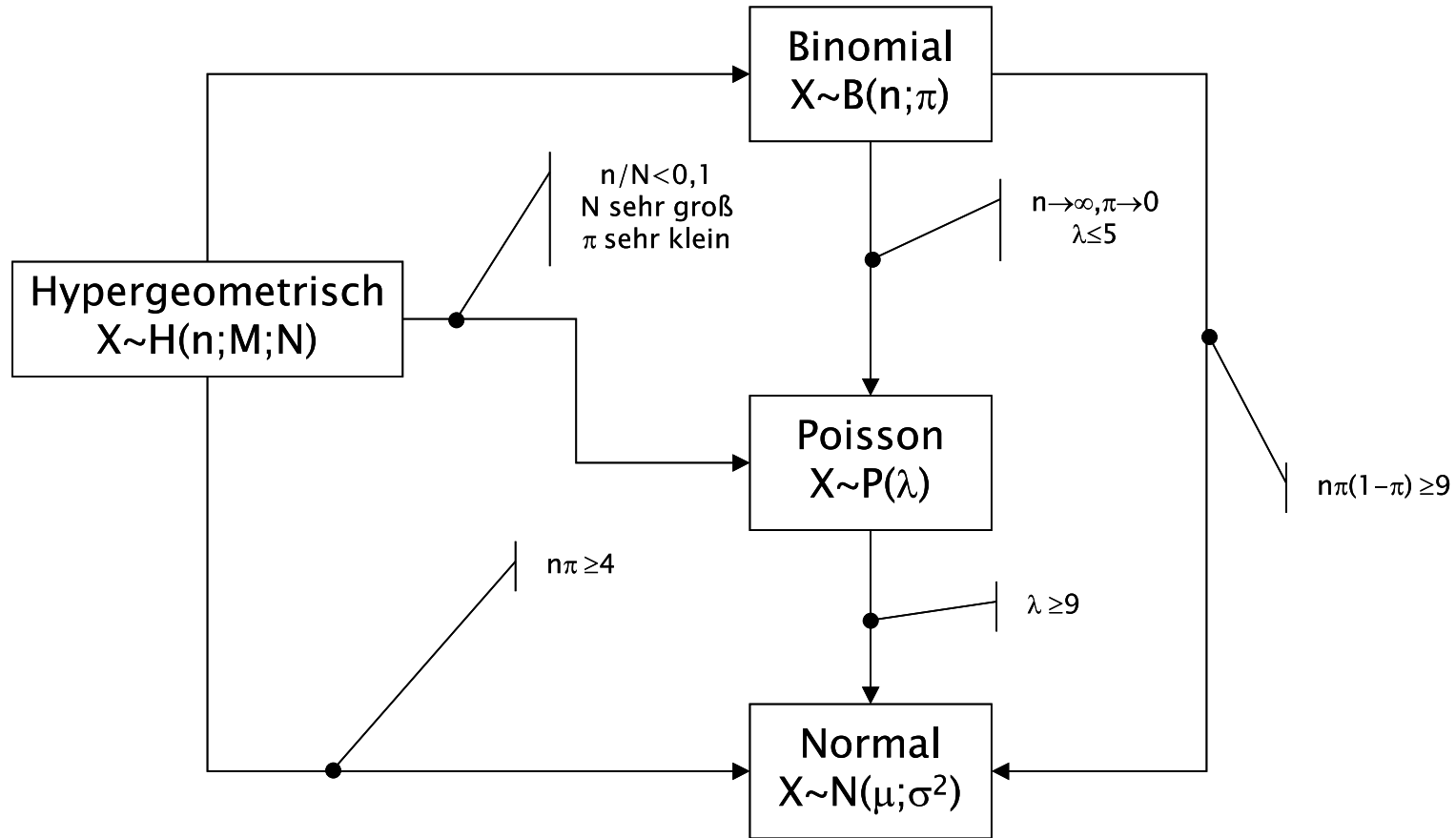
$$X \sim N(14; 144) \Rightarrow Z = \frac{X - \mu}{\sigma} \Rightarrow Z \sim N(0; 1)$$



$$\begin{aligned} P(X \leq -1) &= P\left(Z \leq \frac{-1 - 14}{12}\right) = P(Z \leq -1,25) \\ &= 0,1056 \end{aligned}$$

Die Wahrscheinlichkeit, dass eine $N(14; 144)$ -verteilte Zufallsvariable Werte kleiner oder gleich -1 annimmt, entspricht somit der Wahrscheinlichkeit, dass eine standardnormalverteilte Zufallsvariable $[N(0; 1)]$ Werte kleiner oder gleich $-1,25$ annimmt! (Die beiden schraffierten Flächen sind gleich groß!!)

Approximationsregeln



Approximation der Binomialverteilung durch die Normalverteilung

Auf einer Maschine wird mit einem Ausschußanteil von 10% produziert. Wie groß ist die Wahrscheinlichkeit, dass bei einer Produktionsserie von 120 Stück höchstens 14 Ausschußstücke enthalten sind?

Die Zufallsvariable X (Anzahl der Ausschußstücke) ist grundsätzlich binomialverteilt mit $n = 120$ und $\pi = 0,1$. Da aber die Varianz dieser Verteilung $n\pi(1 - \pi) = 120 \cdot 0,1 \cdot 0,9 = 10,8$ beträgt und dieser Wert größer als 9 ist, kann diese Verteilung durch die Normalverteilung approximiert werden. Dazu muß der Erwartungswert und die Varianz bestimmt werden. Diese betragen:

$$E(X) = n\pi = 120 \cdot 0,1 = 12$$

$$V(X) = n\pi(1 - \pi) = 120 \cdot 0,1 \cdot 0,9 = 10,8$$

Die Anzahl der Ausschußstücke ist somit approximativ normalverteilt:

$$X \sim N(12; 10,8).$$

Gesucht ist die Wahrscheinlichkeit

$$P(X \leq 14) = P\left(Z \leq \frac{14 - 12}{\sqrt{10,8}}\right) = P(Z \leq 0,6) = F(0,6) = 0,7257.$$

Approximation der Hypergeometrischen Verteilung durch die Normalverteilung

Unter den 500 Studenten eines Fachbereichs finanzieren 150 ihr Studium selbst. Wie groß ist die Wahrscheinlichkeit, dass unter 50 zufällig ausgewählten Studenten höchstens 19 zu dieser Gruppe zählen?

Eigentlich handelt es sich hierbei um eine hypergeometrische Verteilung mit $N = 500$, $M = 150$ und $n = 50$ und somit $\pi = \frac{150}{500} = 0,3$. In diesem Fall ist aber $n\pi = 50 \cdot 0,3 = 15$. Dieser Wert ist größer als 4, weshalb auch hier eine Approximation durch die Normalverteilung vorgenommen werden kann. Erwartungswert und Varianz betragen:

$$E(X) = n\pi = 50 \cdot 0,3 = 15$$

$$V(X) = n\pi(1 - \pi) \frac{N - n}{N - 1} = 50 \cdot 0,3 \cdot 0,7 \cdot \frac{450}{499} = 9,4689$$

Es gilt also:

$$X \sim N(15; 9,4689)$$

Die gesuchte Wahrscheinlichkeit beträgt dann

$$P(X \leq 19) = P\left(Z \leq \frac{19 - 15}{\sqrt{9,4689}}\right) = P(Z \leq 1,3) = F(1,3) = 0,9032.$$

Approximation der Poissonverteilung durch die Normalverteilung

An einem Bankschalter sei die Zahl der pro Viertelstunde eintreffenden Kunden poissonverteilt mit $\lambda = 10$. Wie groß ist die Wahrscheinlichkeit, dass mindestens 15 Kunden kommen?

Da der Parameter λ der Poissonverteilung größer als 9 ist, kann auch diese Verteilung durch die Normalverteilung approximiert werden. Da λ für diese Verteilung sowohl den Erwartungswert als auch die Varianz repräsentiert, ist die approximative Normalverteilung gegeben mit:

$$X \sim N(10; 10)$$

Die gesuchte Wahrscheinlichkeit beträgt dann

$$P(X \geq 15) = 1 - P(X \leq 15) = 1 - P\left(Z \leq \frac{15 - 10}{\sqrt{10}}\right) = 1 - P(Z \leq 1,6) = 1 - F(1,6) = 1 - 0,9452 = 0,0548.$$

Konvergenz von Folgen — Gesetz der großen Zahl

Bei stochastischer Konvergenz strebt die *Wahrscheinlichkeit*, dass sich X_n in einer ϵ -Umgebung vom Grenzwert c befindet, mit zunehmendem Stichprobenumfang gegen 1, also

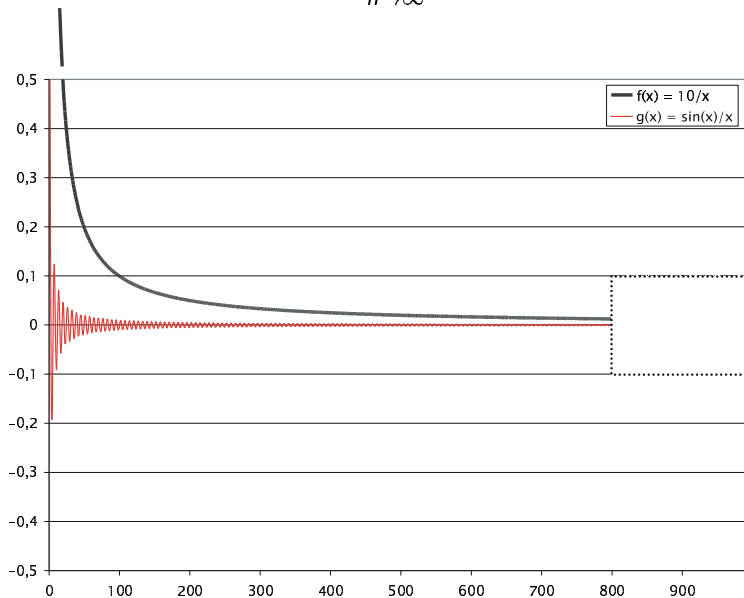
$$\lim_{n \rightarrow \infty} P(|X_n - c| < \epsilon) = 1 \quad \Rightarrow \quad \text{plim } X_n = c.$$

Bei mathematischer Konvergenz gilt dies mit *Sicherheit!*

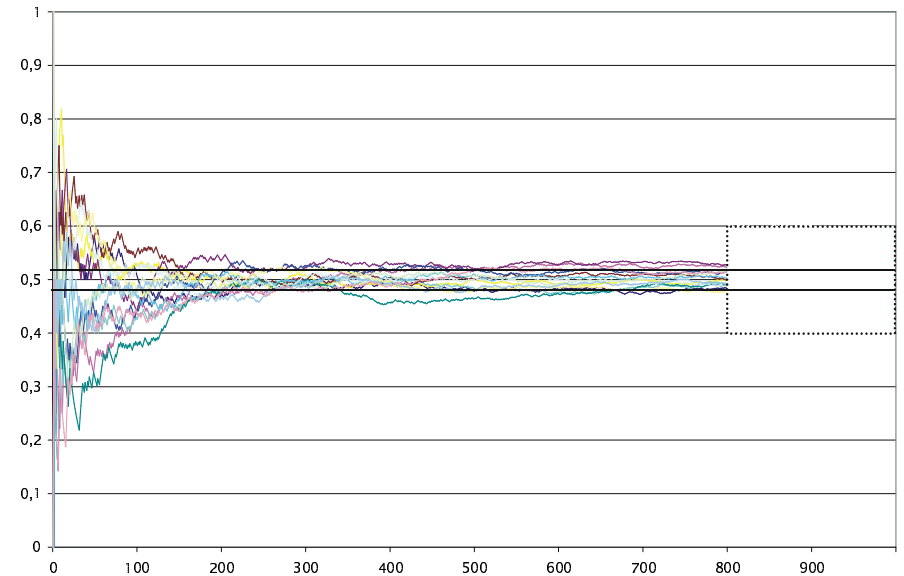
$$\lim_{n \rightarrow \infty} X_n = c.$$

Mathematische vs. Stochastische Konvergenz

$$\lim_{n \rightarrow \infty} X_n = c$$



$$\lim_{n \rightarrow \infty} P(|X_n - c| < \epsilon) = 1 \quad \Rightarrow \quad \text{plim } X_n = c$$



Beispiel — Gesetz der großen Zahl (\leadsto Aufgabe 7.1.12)

Eine 'echte' Münze (d.h. $\pi = 0,5$) wird wiederholt geworfen. Gesucht wird die Wahrscheinlichkeit dafür, dass der Anteil der Wappenwürfe zwischen 0,25 und 0,75 liegt. Dabei werden die Stichprobenumfänge $n = 2, 4, 8$ betrachtet.

Die Zufallsvariable X sei definiert als Anzahl der Wappenwürfe, also $X = X_1 + X_2 + \dots + X_n$. Der Anteil der Wappenwürfe ist dann X/n . Gesucht wird die Wahrscheinlichkeit

$$P\left(0,25 \leq \frac{X}{n} \leq 0,75\right) = P\left(\left|\frac{X}{n} - \pi\right| < 0,25\right)$$

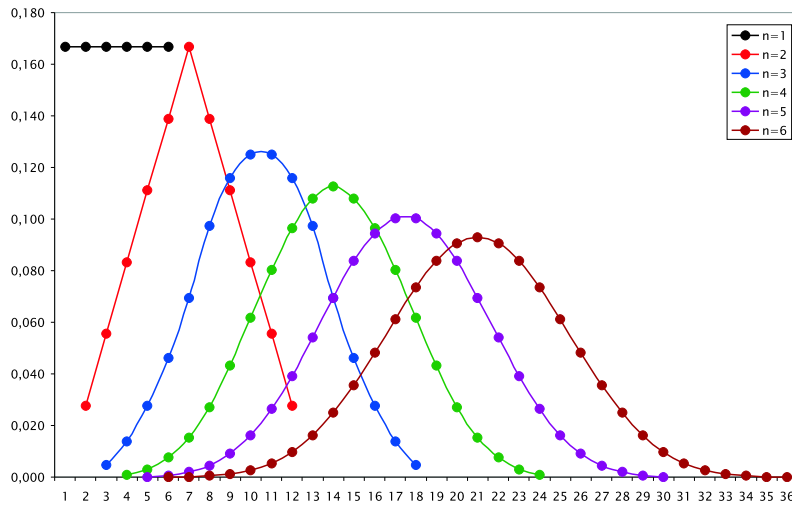
Betrachtet wird die gesuchte Intervallwahrscheinlichkeit für die unterschiedlichen Stichprobenumfänge:

$n = 2$	$n = 4$	$n = 8$																																																															
<table border="1" style="margin: auto; border-collapse: collapse;"> <thead> <tr> <th>X</th> <th>$\frac{X}{n}$</th> <th>$P\left(\frac{X}{n}\right)$</th> </tr> </thead> <tbody> <tr> <td>0</td> <td>0</td> <td>0,25</td> </tr> <tr> <td>1</td> <td>0,5</td> <td>0,5</td> </tr> <tr> <td>2</td> <td>1</td> <td>0,25</td> </tr> </tbody> </table>	X	$\frac{X}{n}$	$P\left(\frac{X}{n}\right)$	0	0	0,25	1	0,5	0,5	2	1	0,25	$P(\cdot) = 0,5$	<table border="1" style="margin: auto; border-collapse: collapse;"> <thead> <tr> <th>X</th> <th>$\frac{X}{n}$</th> <th>$P\left(\frac{X}{n}\right)$</th> </tr> </thead> <tbody> <tr> <td>0</td> <td>0</td> <td>0,0625</td> </tr> <tr> <td>1</td> <td>0,25</td> <td>0,25</td> </tr> <tr> <td>2</td> <td>0,5</td> <td>0,375</td> </tr> <tr> <td>3</td> <td>0,75</td> <td>0,25</td> </tr> <tr> <td>4</td> <td>1</td> <td>0,0625</td> </tr> </tbody> </table>	X	$\frac{X}{n}$	$P\left(\frac{X}{n}\right)$	0	0	0,0625	1	0,25	0,25	2	0,5	0,375	3	0,75	0,25	4	1	0,0625	$P(\cdot) = 0,875$	<table border="1" style="margin: auto; border-collapse: collapse;"> <thead> <tr> <th>X</th> <th>$\frac{X}{n}$</th> <th>$P\left(\frac{X}{n}\right)$</th> </tr> </thead> <tbody> <tr> <td>0</td> <td>0</td> <td>0,0039</td> </tr> <tr> <td>1</td> <td>0,125</td> <td>0,0313</td> </tr> <tr> <td>2</td> <td>0,25</td> <td>0,1094</td> </tr> <tr> <td>3</td> <td>0,375</td> <td>0,2188</td> </tr> <tr> <td>4</td> <td>0,5</td> <td>0,2734</td> </tr> <tr> <td>5</td> <td>0,625</td> <td>0,2188</td> </tr> <tr> <td>6</td> <td>0,75</td> <td>0,1094</td> </tr> <tr> <td>7</td> <td>0,875</td> <td>0,0313</td> </tr> <tr> <td>8</td> <td>1</td> <td>0,0039</td> </tr> </tbody> </table>	X	$\frac{X}{n}$	$P\left(\frac{X}{n}\right)$	0	0	0,0039	1	0,125	0,0313	2	0,25	0,1094	3	0,375	0,2188	4	0,5	0,2734	5	0,625	0,2188	6	0,75	0,1094	7	0,875	0,0313	8	1	0,0039	$P(\cdot) = 0,9298$
X	$\frac{X}{n}$	$P\left(\frac{X}{n}\right)$																																																															
0	0	0,25																																																															
1	0,5	0,5																																																															
2	1	0,25																																																															
X	$\frac{X}{n}$	$P\left(\frac{X}{n}\right)$																																																															
0	0	0,0625																																																															
1	0,25	0,25																																																															
2	0,5	0,375																																																															
3	0,75	0,25																																																															
4	1	0,0625																																																															
X	$\frac{X}{n}$	$P\left(\frac{X}{n}\right)$																																																															
0	0	0,0039																																																															
1	0,125	0,0313																																																															
2	0,25	0,1094																																																															
3	0,375	0,2188																																																															
4	0,5	0,2734																																																															
5	0,625	0,2188																																																															
6	0,75	0,1094																																																															
7	0,875	0,0313																																																															
8	1	0,0039																																																															

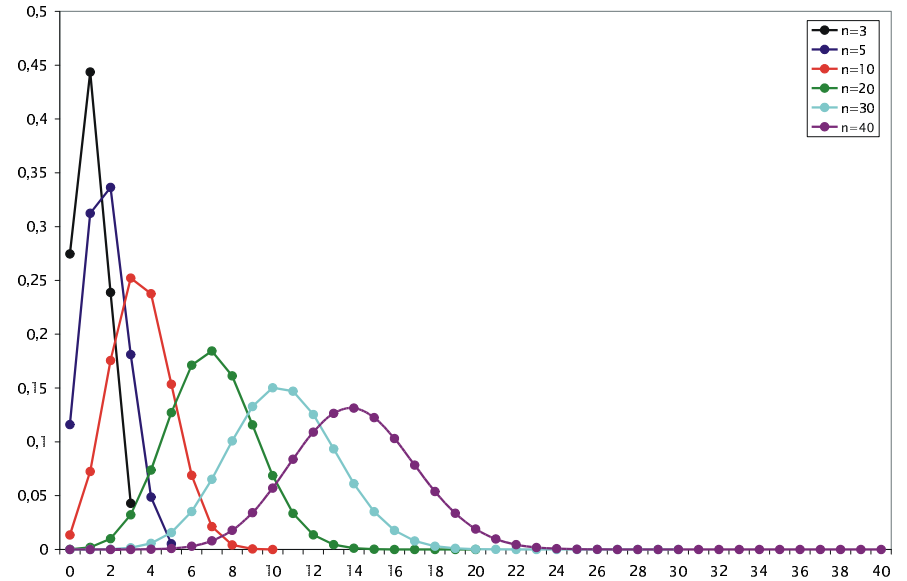
Mit wachsendem Stichprobenumfang ($n = 2, 4, 8, \dots$) strebt die Wahrscheinlichkeit, dass der Anteil der Wappenwürfe in einer ϵ -Umgebung um π liegt (hier: $\epsilon = 0,25$), gegen eins.

Konvergenz von Verteilungen — Grenzwertsätze

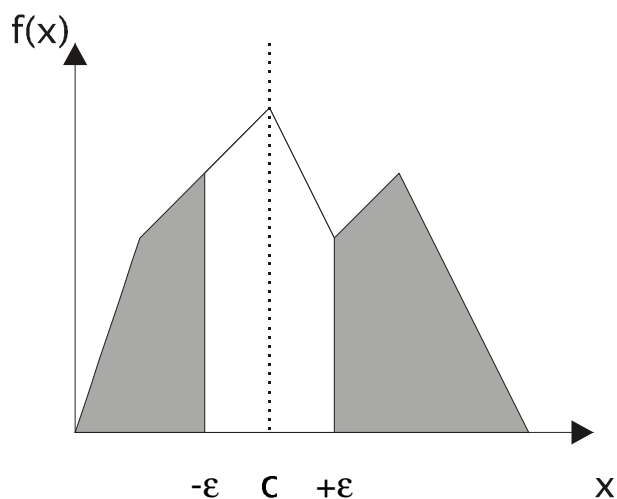
Wiederholter Würfelwurf; ZV X: Augensumme



Binomialverteilung mit $\pi = 0,35$: ZV X: Anzahl der Erfolge



Ungleichung von Tschebyscheff



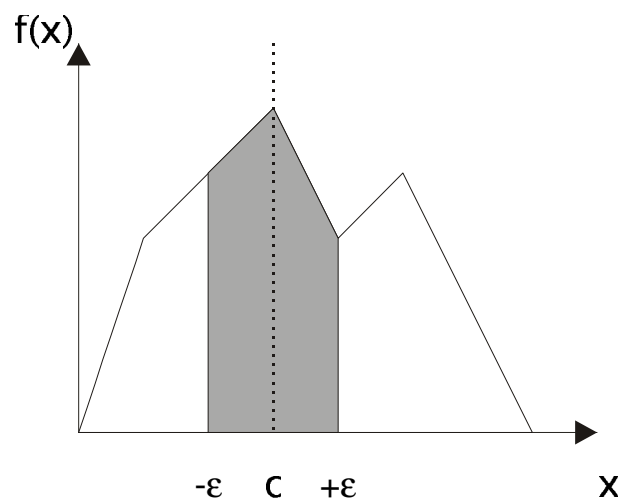
$$P(|X - c| \geq \epsilon) \leq \frac{1}{\epsilon^2} E(X - c)^2$$

$$c = \mu \Rightarrow E(X - \mu)^2 = \sigma^2$$

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

$$\epsilon = t \cdot \sigma \Rightarrow \epsilon^2 = t^2 \cdot \sigma^2$$

$$P(|X - \mu| \geq t \cdot \sigma) \leq \frac{1}{t^2}$$



$$P(|X - c| < \epsilon) \geq 1 - \frac{1}{\epsilon^2} E(X - c)^2$$

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

$$P(|X - \mu| < t \cdot \sigma) \geq 1 - \frac{1}{t^2}$$

Eigenschaften von Schätzfunktionen (1)

Erwartungstreue

$$E(\hat{\Theta}) = \Theta$$

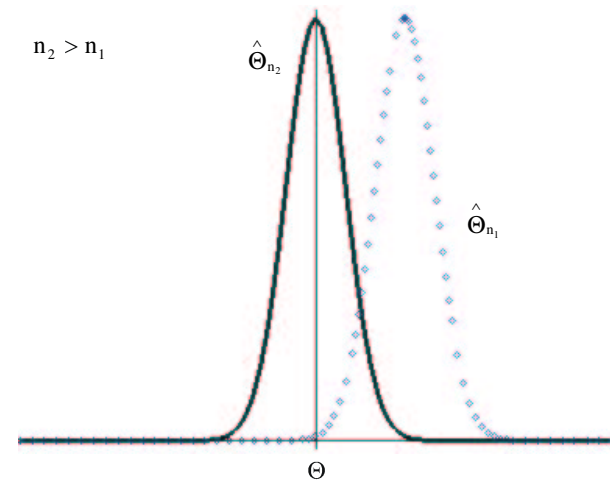
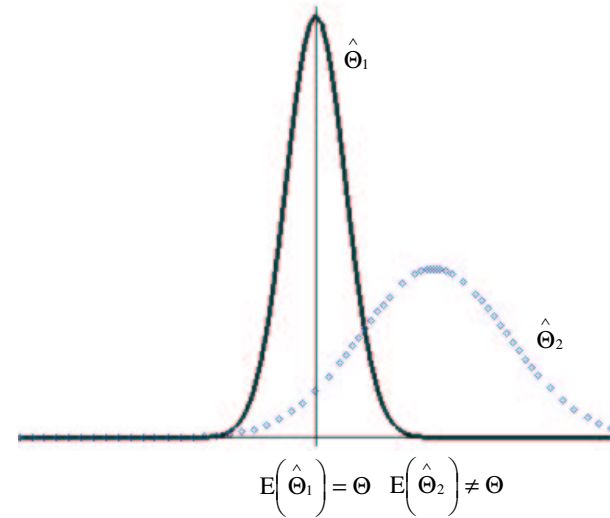
$$\text{Bias} = B = E(\hat{\Theta}) - \Theta = 0$$

unverzerrter Schätzer: $B = 0$

verzerrter Schätzer: $B \neq 0$

asymptotische Erwartungstreue

$$\lim_{n \rightarrow \infty} E(\hat{\Theta}) = \Theta \text{ bzw. } \lim_{n \rightarrow \infty} B = 0$$



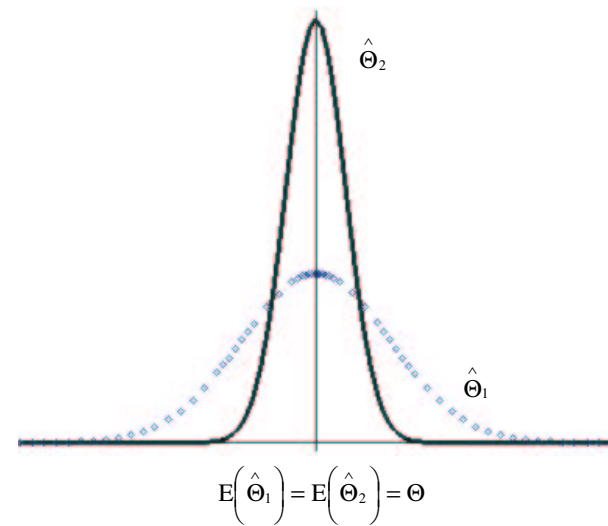
Eigenschaften von Schätzfunktionen (2)

Effizienz

$$V(\hat{\Theta}_1, n) < V(\hat{\Theta}_2, n)$$

Voraussetzungen:

- beide Schätzer sind erwartungstreu
- gleicher Stichprobenumfang

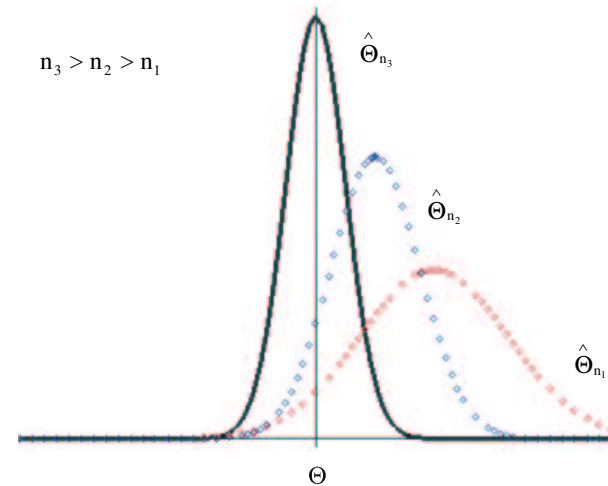


Konsistenz

$$\text{MSE} = E(\hat{\Theta} - \Theta)^2 = V(\hat{\Theta}) + B^2$$

$$\lim_{n \rightarrow \infty} \text{MSE} = 0 \quad \Rightarrow \quad \lim_{n \rightarrow \infty} V(\hat{\Theta}) = \lim_{n \rightarrow \infty} B = 0$$

- Voraussetzung: mindestens asymptotisch erwartungstreu
Folge: ein Mehraufwand bei der Erhebung führt zu besseren (sichereren) Ergebnissen.



Beispiel: (Einstichproben–Test, Homograder Fall)

Ein Pharmakonzern hat ein neues Medikament auf den Markt gebracht. Der Konzern behauptet, dass die Erfolgswahrscheinlichkeit für dieses Medikament gleich 0,75 ist.

Diese Aussage ist eine statistische Hypothese, welche durch eine Stichprobe überprüft werden kann, d.h. eine bestimmte Anzahl von Patienten wird mit dem Medikament behandelt und es wird überprüft, bei wieviel Personen das Medikament wirksam ist. Da das Medikament nicht an allen potentiellen Patienten getestet werden kann, sondern nur an den in der Stichprobe enthaltenen, kann man auch nicht mit vollkommener Sicherheit sagen, ob die aufgestellte Hypothese richtig oder falsch ist.

In einem Versuch wurden von $n = 40$ Versuchspersonen $x = 25$ Patienten geheilt. ($p = 25/40 = 0,625$)

Aufgrund der vorliegenden Stichprobe wird nun ein Hypothesentest durchgeführt, an dessen Ende die Entscheidung für oder gegen die Hypothese des Pharmakonzerns steht.

1. Formulierung der Hypothese bzw. der Gegenhypothese:

$$H_0 : \pi \geq 0,75 \quad \text{vs.} \quad H_1 : \pi < 0,75$$

2. Aufstellen der Teststatistik:

Es wird so getan, als sei die Hypothese des Herstellers wahr, d.h. es gelte $\pi = 0,75$. Zur Ermittlung der Teststatistik wird die Übersicht 9.3 auf S. 97 verwendet: homograder Fall, zms (Umfang N der Grundgesamtheit ist nicht bekannt) \Rightarrow Fall (3)

$$T = \frac{p - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}} = \frac{0.625 - 0.75}{\sqrt{\frac{0.75(1-0.75)}{40}}} = -1.8257.$$

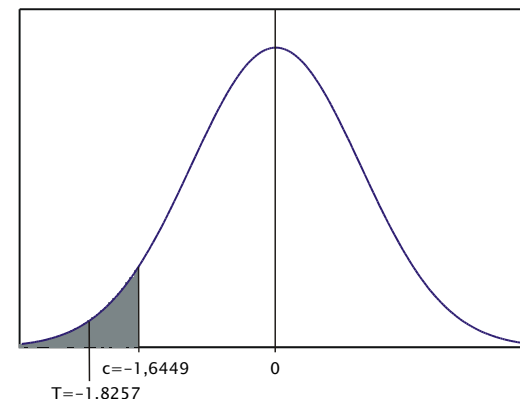
3. Ermittlung des kritischen Wertes:

Bei der Wahl für ein Signifikanzniveau (hier: $\alpha = 0,05$) gelangt man zu einem kritischen Wert, der in diesem Fall der Normalverteilung entnommen wird, da die Anzahl der Erfolge für genügend große Stichproben asymptotisch normalverteilt ist. Die Anpassungsvorschrift ($n \cdot \pi \cdot (1 - \pi) > 9$) wird erfüllt (hier: $40 \cdot 0,625 \cdot 0,375 = 9,375$). Der kritische Wert lautet – da es sich um einen einseitigen Test nach unten handelt – demnach:

$$c = -1,6449.$$

4. Entscheidung:

Da der Wert der Teststatistik kleiner als der kritische Wert ist ($T < c$), kann die Nullhypothese zum Niveau $\alpha = 0,05$ verworfen werden. Damit ist jedoch nicht bewiesen, dass die Aussage des Pharmaherstellers falsch ist. Das Stichprobenergebnis weicht jedoch so deutlich (signifikant) von der Aussage des Herstellers ab, dass vermutet werden kann, dass seine Behauptung nicht korrekt ist.



5. Fehler zweiter Art (β -Fehler):

Angenommen, in der Stichprobe wären 27 von 40 Personen geheilt worden. Der Anteil hätte dann ($p = 27/40 = 0,675$) betragen. Die Durchführung des obigen Tests hätte dazu geführt, dass die Null-Hypothese des Herstellers nicht verworfen werden kann ($T = -1,0954$). Einen Fehler zweiter Art zu begehen bedeutet dann, dass man die Null-Hypothese (des Herstellers) nicht verwirft, obwohl diese falsch ist. Die Wahrscheinlichkeit dafür beträgt (höchstens) β :

Vorab wird der kritische Anteilswert berechnet, d.h., der Wert, bei dem – bei Gültigkeit der Nullhypothese – keine Entscheidung getroffen werden kann. In diesem Fall muß gelten: $T = c$.

$$T = c \quad \Rightarrow \quad \frac{p_c - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}} = -1,6449 \quad \Rightarrow \quad p_c = 0,6374.$$

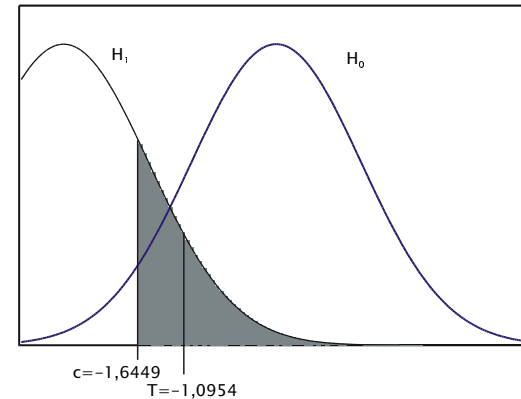
Wenn nun der Anteilswert in der Stichprobe kleiner als p_c ist, wird H_0 verworfen, ist er größer als p_c , entscheidet man sich für H_0 .

Die Wahrscheinlichkeit β für den Fehler zweiter Art berechnet sich als bedingte Wahrscheinlichkeit für die Annahme der Null-Hypothese unter der Bedingung, dass ein anderer Parameter gültig ist. In unserem Fall bedeutet das:

$$P(p > p_c | H_1) = P(T > c | H_1) \leq \beta.$$

Für $H_1 : \pi = 0,5$ erhält man dann:

$$P\left(T_{|H_1} > \frac{0,6374 - 0,5}{\sqrt{\frac{0,5 \cdot 0,5}{40}}}\right) = 1 - P(T \leq 1,7378) = 0,0406.$$



Kontakt: Michael Westermann

Anschrift: Universität Duisburg–Essen
— Campus Essen —
Fachbereich Wirtschaftswissenschaften
Universitätsstr. 12
45112 Essen

e–Post: westermann@vwl.uni-essen.de

Homepage: www.vwl.uni-essen.de/westermann/