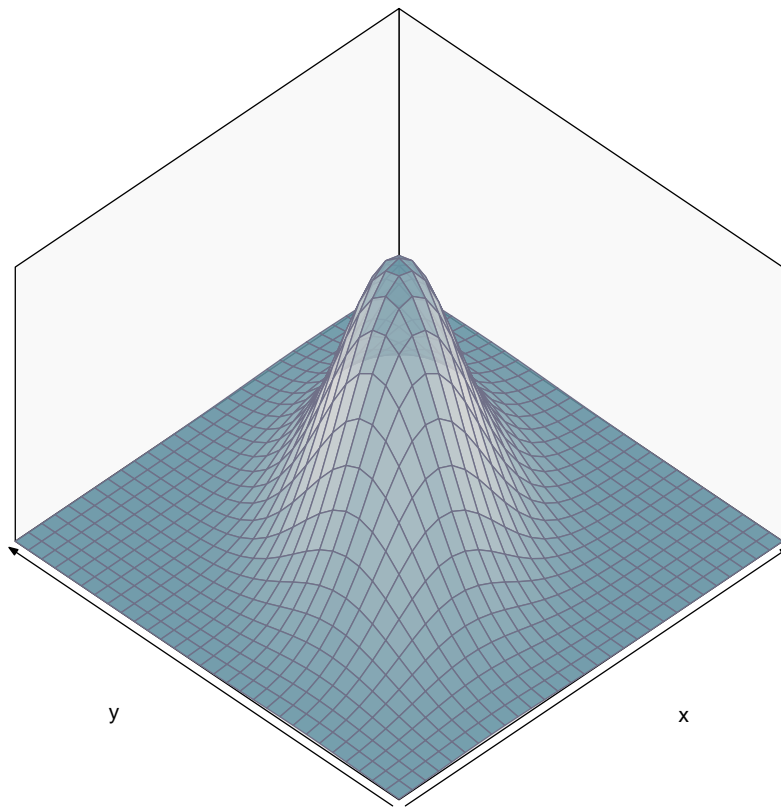


Wahrscheinlichkeitstheorie

Zusatz für die Interessierten, Modul Angewandte Statistik, HS 25

André Meichtry

15. September 2025



Open-source software for typesetting and for statistical computing was used, that is \LaTeX for typesetting, \R for statistical computing and the **knitr**-package for literate programming.

Inhaltsverzeichnis

| | | |
|----------|---|-----------|
| 1 | Algebra review | 2 |
| 1.1 | Wichtige Aspekte | 2 |
| 1.2 | Formelsammlung | 6 |
| 1.2.1 | Termumformungen | 6 |
| 1.2.2 | Brüche | 6 |
| 1.2.3 | Potenzen und Wurzeln | 7 |
| 1.2.4 | Logarithmen | 8 |
| 2 | Wahrscheinlichkeit | 9 |
| 2.1 | Grundbegriffe | 9 |
| 2.2 | Laplace-Wahrscheinlichkeit, objektive und subjektive Wahrscheinlichkeit . | 14 |
| 2.3 | Bedingte Wahrscheinlichkeit | 16 |
| 2.4 | Unabhängigkeit | 20 |
| 2.5 | Der Satz von Bayes | 22 |
| 2.6 | Bayesianische Statistik*. | 28 |
| 3 | Zufallsvariablen und ihre Verteilung | 29 |
| 3.1 | Zufallsvariable | 29 |
| 3.2 | Verteilungen | 32 |
| 3.3 | Wichtige Verteilungen | 35 |
| 3.3.1 | Binomialverteilung | 36 |
| 3.3.2 | Poisson-Verteilung | 37 |
| 3.3.3 | Normalverteilung | 38 |
| 3.4 | Andere Verteilungen | 42 |
| 3.5 | Verteilungen mit R | 45 |
| 3.6 | Erwartungswerte | 47 |
| 3.6.1 | Erwartungswert | 47 |
| 3.6.2 | Varianz und Standardabweichung | 48 |
| 3.6.3 | Erwartungswerte von wichtigen parametrischen Verteilungen. . . . | 50 |
| 3.7 | Gemeinsame Verteilungen | 51 |
| 3.8 | Kovarianz und Korrelation | 53 |
| 3.8.1 | Rechenregeln | 54 |
| | Literaturverzeichnis | 56 |

Abschnitte mit (*) können weggelassen werden ohne Verlust der Kontinuität.

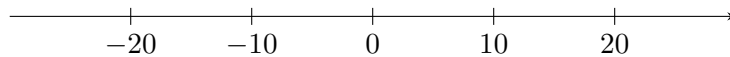
Kapitel 1

Algebra review

1.1 Wichtige Aspekte

Mengen. Eine Menge stellen wir mit geschweiften Klammern dar: $\{\dots\}$. Für eine Menge von x , die die Bedingung y erfüllen, schreiben wir $\{x \mid y\}$. Die Notation $x \in y$ bedeutet: x ist Element von y . Wichtige Zahlenmengen sind nun:

- \mathbb{N} : Menge der natürlichen Zahlen $\{1, 2, 3, \dots\}$
- \mathbb{Z} : Menge der ganzen Zahlen $\{\dots, -2, -1, 0, 1, 2, \dots\}$
- \mathbb{Q} : Menge der rationalen Zahlen, Brüche: $\{\frac{a}{b} \mid a \in \mathbb{Z}, b \in \mathbb{N}, b \neq 0\}$
- \mathbb{R} : Menge der reellen Zahlen. Zusätzlich zu den rationalen Zahlen kommen die irrationalen Zahlen wie z.B.
 - $\pi = 3.14159265358979$
 - $\sqrt{2} = 1.4142135623731$
 - $e = 2.71828182845905$



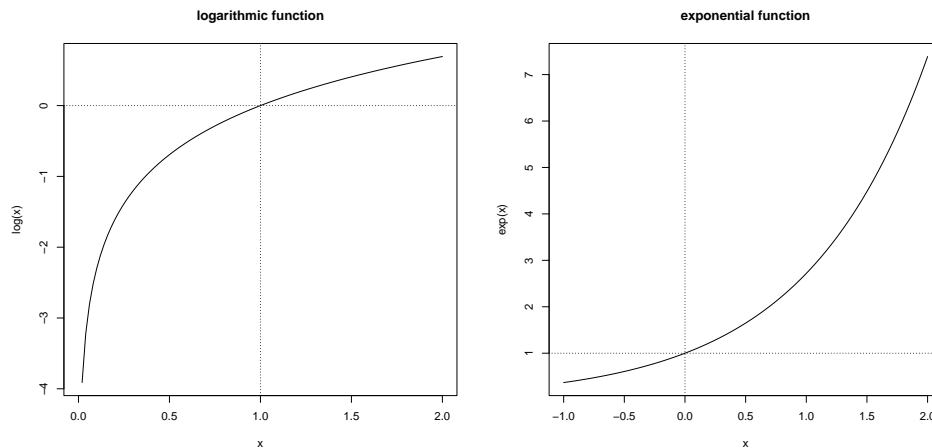
Algebra. Wichtige Aspekte sind:

- Symbole als Variablen verstehen, und für was sie stehen.
- Rechnen mit Klammern
- Ausklammern
- Negation ausklammern
- Brüche kürzen und erweitern
- Addition, Subtraktion, Multiplikation and Division von Brüchen
- Doppelbrüche
- Potenzieren

- Potenzieren mit ganzen Zahlen und mit Brüchen im Exponenten: $a^{1/n} = \sqrt[n]{a}$
- Potenzieren mit Summe im Exponenten: $a^{m+n} = a^m \times a^n$
- Potenzieren mit Differenz im Exponenten : $a^{m-n} = a^m / a^n$
- Potenzieren mit negativem Exponenten: $a^{-m} = a^{0-m} = \frac{a^0}{a^m} = 1/a^m$.
- Wenn $a = e = 2.718$, schreiben wir häufig $\exp(x)$ für e^x
- Spezielle Exponenten: $a^0 = 1$ und $a^1 = a$
- n -te Wurzeln
 - $\sqrt[n]{a} = a^{1/n}$
 - $\sqrt[n]{\frac{1}{a}} = (\frac{1}{a})^{1/n} = (a^{-1})^{1/n} = a^{-\frac{1}{n}}$
- Logarithmus
 - log zur Basis b von x , $y = \log_b x$, ist die Zahl y so dass $b^y = x$
 - $\log_b b = 1$
 - $\log_b b^x = x \log_b b = x$
 - $\log_b a^x = x \log_b a$
 - $\log_b a^{-x} = -x \log_b a$
 - $\log_b(xy) = \log_b x + \log_b y$
 - $\log_b \frac{x}{y} = \log_b x - \log_b y$
 - Basiswechsel: $\log_a b = \frac{\log_e b}{\log_e a}$
 - Wenn $b = e = 2.718$, ist die Basis der natürliche Logarithmus. $\log_e x$ schreiben wir kurz $\log x$ oder $\ln x$
 - $\log e = 1$
- Die Umkehrfunktion vom log zur Basis b von x ist b^x
- Logarithmus $\log(\cdot)$ und Exponentialfunktion $\exp(\cdot)$ Funktion sind zentral in der Wissenschaft und in der Statistik.
 - $\log(x) : \mathbb{R}^+ \rightarrow \mathbb{R} : y = \log(x)$
 - $\exp(x) : \mathbb{R} \rightarrow \mathbb{R}^+ : y = \exp(x)$

```

curve(log(x), from = 0, to = 2, main = "logarithmic function")
abline(v = 1, h = 0, lty = 3)
curve(exp(x), from = -1, to = 2, main = "exponential function")
abline(v = 0, h = 1, lty = 3)
    
```

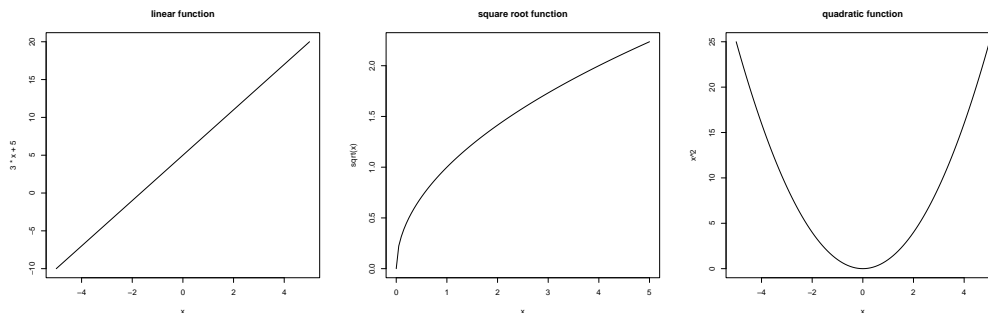


- Funktionen verstehen wie $y = a + bx$
- Indikatorvariablen: $I_{x=3}$: wahr, wenn $x = 3$, sonst falsch (oder 1 wenn $x = 3$, 0 sonst),

$$I_{x=3} = \begin{cases} 1 & x = 3 \\ 0 & x \neq 3. \end{cases}$$

- Lineare Gleichungen $ax + b = 0$
- Quadratische Gleichungen $ax^2 + bx + c = 0$
- Funktionen darstellen. (Mit R wird das einfach sein.)

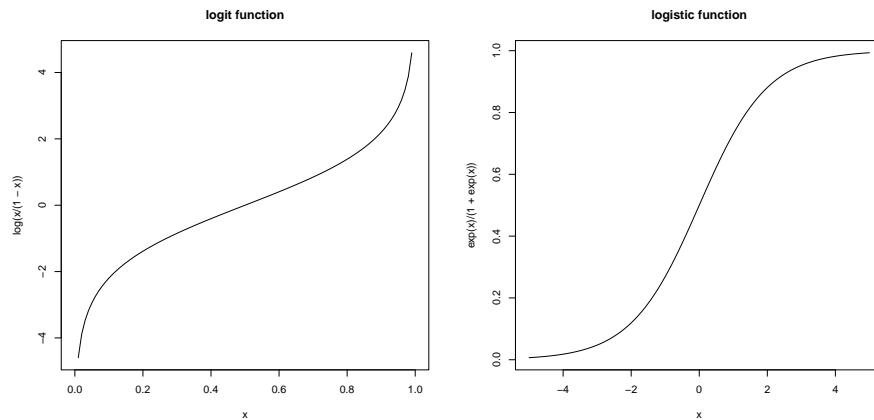
```
curve(3 * x + 5, from = -5, to = 5, main = "linear function")
curve(sqrt(x), from = 0, to = 5, main = "square root function")
curve(x^2, from = -5, to = 5, main = "quadratic function")
```



- Die Logistische (`logistic(.)`) und Logit-Funktion (`logit(.)`) werden wir häufig antreffen. Eine ist die Umkehrfunktion der anderen.

$$\begin{aligned} - \text{logit}(x) &: [0, 1] \rightarrow \mathbb{R}, \quad x \mapsto \log \frac{x}{1-x}. \\ - \text{logistic}(x) &: \mathbb{R} \rightarrow [0, 1], \quad x \mapsto \frac{\exp(x)}{\exp(x)+1}. \end{aligned}$$

```
curve(log(x/(1 - x)), main = "logit function")
curve(exp(x)/(1 + exp(x)), from = -5, to = 5, main = "logistic function")
```



- Wenn wir zum linearen Modell kommen (am Ende des Moduls und im nächsten Modul), wird Notation mit Vektor/Linearer Algebra wichtig sein:

– Sei \mathbf{x} ein (Kolonnen)vektor mit den Grössen x_1, x_2, \dots, x_p , also $\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{pmatrix}$

(z.B. die Werte von p Variablen einer Person wie Alter, Blutdruck, usw.)

– Sei $\boldsymbol{\beta}$ ein (Kolonnen)vektor mit den Grössen $\beta_1, \beta_2, \dots, \beta_p$, also $\boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}$

(Gewichte/Regressionskoeffizienten)

– \mathbf{x}^T ist dann ein Zeilenvektor (T : “transponiert”) und

– $\mathbf{x}^T \boldsymbol{\beta} = (x_1, x_2, \dots, x_p) \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_p \end{pmatrix} = \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p$ ist dann ein

Skalarprodukt (eine Zahl) und heisst linearer Prädiktor, die mit den Gewichten $\beta_1, \beta_2, \dots, \beta_p$ gewichtete Summe der x_1, x_2, \dots, x_p .

1.2 Formelsammlung

1.2.1 Termumformungen

| Kommutativgesetz | Assoziativgesetz |
|--|--|
| $a + b = b + a$ $ab = ba$ | $a + (b + c) = (a + b) + c$ $a(bc) = (ab)c$ |
| Distributivgesetz | Ausklammern |
| $a(b + c) = ab + ac$ | $ab + ac = a(b + c)$ |
| Binomische Formeln | |
| $(x + y)^2 = x^2 + 2xy + y^2$ $(x - y)^2 = x^2 - 2xy + y^2$ $x^2 - y^2 = (x + y)(x - y)$ | |

Tabelle 1.1: Termumformungen

1.2.2 Brüche

| Kürzen | Erweitern |
|---|---|
| $\frac{ax}{ay} = \frac{x}{y}$ | $\frac{x}{y} = \frac{ax}{ay}$ |
| Addieren und Subtrahieren | |
| $\frac{a}{b} \pm \frac{c}{d} = \frac{ad \pm cb}{bd}$ | |
| Multiplizieren | Dividieren |
| $\frac{a}{b} \frac{c}{d} = \frac{ac}{bd}$ | $\frac{a}{b} : \frac{c}{d} = \frac{a}{b} \frac{d}{c} = \frac{ad}{bc}$ |
| Doppelbrüche | |
| $\frac{a/b}{c/d} = \frac{a}{b} : \frac{c}{d} = \frac{ad}{bc}$ | |

Tabelle 1.2: Brüche

1.2.3 Potenzen und Wurzeln

| | |
|--|---|
| $\underbrace{a \cdot a \cdot \dots \cdot a}_{n \text{ mal}} = a^n$ | $\sqrt[n]{a} = a^{\frac{1}{n}}$ |
| Rechenregeln | |
| $a^m a^n = a^{m+n}$ | $\sqrt[m]{a} \sqrt[n]{a} = a^{\frac{m+n}{mn}}$ |
| $\frac{a^m}{a^n} = a^{m-n}$ | $\frac{\sqrt[n]{a}}{\sqrt[n]{b}} = \sqrt[n]{\frac{a}{b}}$ |
| $(a^m)^n = a^{mn}$ | $(\sqrt[n]{a})^m = \sqrt[n]{a^m}$ |
| $a^n b^n = (ab)^n$ | $\sqrt[m]{\sqrt[n]{a}} = \sqrt[mn]{a}$ |
| $\frac{a^n}{b^n} = \left(\frac{a}{b}\right)^n$ | $a^{p/q} = \sqrt[q]{a^p}$ |
| $a^{-n} = \left(\frac{1}{a}\right)^n$ | $a^{-1/n} = \sqrt[n]{\frac{1}{a}}$ |
| Spezielle Exponenten | |
| $a^0 = 1$ | $a^1 = a$ |

Tabelle 1.3: Potenzen und Wurzeln

$9^{(1/2)}$ #square root
 $(27)^{(1/3)}$ #3th root

1.2.4 Logarithmen

| |
|--|
| $x = \log_a(b) \Leftrightarrow a^x = b$ |
| Rechenregeln |
| $\log(ab) = \log(a) + \log(b)$ |
| $\log(a/b) = \log(a) - \log(b)$ |
| $\log(a^x) = x \log(a)$ |
| Identitäten |
| $a^{\log_a(b)} = b$ |
| $\log_a(a^b) = b$ |
| Spezielle Basen |
| $\log_{10}(x) = \lg(x)$ |
| $\log_e(x) = \log(x) = \ln(x)$ |
| Basiswechsel |
| $\log_a(b) = \frac{\log_c(b)}{\log_c(a)}$ |
| $\ln(b) = \log(b) = \frac{\log_{10}(b)}{\log_{10}(e)}$ |

Tabelle 1.4: Logarithmen

```
a <- 10
b <- 30
log(a * b)
log(a) + log(b)
log(a/b)
log(a) - log(b)
log(a^5)
5 * log(a)
```

```
e <- exp(1) #Eulersche Zahl e
e
log(20) #natürlicher Log mit Basis e
log10(20)/log10(e) #Basiswechsel
```

Kapitel 2

Wahrscheinlichkeit

In diesem Kapitel führen wir den Begriff der *Wahrscheinlichkeit* ein. Die *Wahrscheinlichkeitstheorie* befasst sich mit Aspekten, in denen der Zufall eine Rolle spielt. Wahrscheinlichkeitsrechnung ist ein mächtiges Werkzeug, das man in der Statistik, in der Epistemologie und Wissenschaftsphilosophie braucht.

Was ist Wahrscheinlichkeit? Alltagssätze haben die Struktur “Die Wahrscheinlichkeit von A ist p ”, mit A als einem *Ereignis* oder einer *Aussage* und p als einer Quantität des *Grades der Überzeugung* in A . Das Verständnis der mathematischen Definition von Wahrscheinlichkeit ist äusserst wichtig in der Statistik.

Am Ende dieses Abschnittes werden wir sehen, dass Wahrscheinlichkeiten numerische positive Quantitäten sind – definiert auf einer *Menge von Ergebnissen* – die additiv sind über sich gegenseitig ausschliessende Ergebnisse und sich auf eins summieren über alle möglichen sich gegenseitig ausschliessenden Ergebnisse.

Grundidee. *Zufallsexperimente* sind Experimente, deren Ergebnisse nicht immer exakt vorausgesagt werden können. Wir möchten dafür ein mathematisches Modell.

2.1 Grundbegriffe

Notation. $\Pr(A)$ steht für die Wahrscheinlichkeit des Ereignisses A . Der Begriff vom Ereignis wird unten definiert. Eine Menge stellen wir mit geschweiften Klammern dar: $\{\dots\}$. Für eine Menge von x , die die Bedingung y erfüllen, schreiben wir $\{x \mid y\}$. \mathbb{N} steht für die Menge der natürlichen Zahlen, \mathbb{Z} für die Menge der ganzen Zahlen, \mathbb{Q} für die Menge der rationalen Zahlen, \mathbb{R} für die Menge der reellen Zahlen und \mathbb{R}_+ für die Menge der nicht-negativen reellen Zahlen. A^C : Komplement von A . \in : Element von, \subseteq, \subset : Teilmenge von, \emptyset : leere Menge, \cup : ODER, \cap : UND, $|A|$: Anzahl Elemente in Menge A , Σ : Summe, \prod : Produkt.

Definition. Der *Ergebnisraum* oder *Stichprobenraum* Ω ist die Menge aller möglichen *Ergebnisse* des betrachteten Zufallsexperimentes. Die Elemente $\omega \in \Omega$ heissen *Elementarereignisse*.

1. Beim Werfen eines Würfels ist $\Omega = \{1, 2, \dots, 6\}$
2. Für die Heilungszeit eines Patienten ist $\Omega = \{t \mid t \geq 0\} = \mathbb{R}_+$
3. Für die (ganzzahligen) Noten bei einer Prüfung ist $\Omega = \{1, 2, 3, \dots, 6\}$
4. Für eine Kraftmessung K ist $\Omega = \{k \mid k \geq 0\} = \mathbb{R}_+$
5. Heilung eines Patienten: $\Omega = \{+, -\}$
6. Heilung von zwei Patienten: $\Omega = \{++, +-, -+, --\}$

Definition. Ein *Ereignis* ist eine Teilmenge $A \subseteq \Omega$, also eine Kollektion von Elementarereignissen. Die Klasse aller *beobachtbaren Ereignisse* ist der *Ereignisraum* \mathcal{F} . Dieser ist eine Teilmenge der *Potenzmenge* $2^\Omega = \mathcal{P}(\Omega)$ von Ω . Die Potenzmenge ist die *Menge aller Teilmengen* von Ω .

Ist Ω endlich oder abzählbar (Beispiele 1, 3, 5 und 6), so wählt man oft als Ereignisraum \mathcal{F} die Potenzmenge 2^Ω . Ist Ω überabzählbar (wie in Beispiel 2 und 4), so muss man als \mathcal{F} eine echte Teilklasse von 2^Ω nehmen (Wir gehen hier nicht auf die Details ein, weil sie für die Praxis unwesentlich sind). In jedem Fall muss der Ereignisraum \mathcal{F} folgende Bedingungen erfüllen:

1. $\Omega \in \mathcal{F}$.
2. Wenn $A_1, A_2, \dots \in \mathcal{F}$, dann $\bigcup A_i \in \mathcal{F}$. (“abgeschlossen unter Vereinigung”)
3. Wenn $A \in \mathcal{F}$, dann $A^C \in \mathcal{F}$. (“abgeschlossen unter Komplementbildung”)

Dasselbe Experiment kann man in der Regel durch verschiedene (Ω, \mathcal{F}) beschreiben:

Beispiel. Jemand macht eine Prüfung, teilt aber nur mit, ob das Ereignis bestanden oder nicht bestanden eingetreten ist.

- Erste Beschreibung: $\Omega_1 = \{pass, fail\}$ und $\mathcal{F}_1 = \{\emptyset, \Omega_1, \{pass\}, \{fail\}\}$.
- Zweite Beschreibung: $\Omega_2 = \{1, 2, 3, \dots, 6\}$; dann darf aber nicht $\mathcal{F}_2 = 2^{\Omega_2}$, weil wir ja die exakte Note nicht beobachten können. Hier wäre dann $\mathcal{F}_2 = \{\emptyset, \Omega_2, \{1, 2, 3\}, \{4, 5, 6\}\}$.

Man beachte, dass \mathcal{F}_1 und \mathcal{F}_2 dieselbe Anzahl von Mengen (beobachtbare Ereignisse) enthalten.

Wir stellen uns allgemein vor, dass wir bei unserem Zufallsexperiment genau ein Elementarereignis ω erhalten. Wir sagen, dass das Ereignis A *eintritt*, falls das realisierte Elementarereignis ω in A liegt, d.h. $\omega \in A$. Mit Hilfe von Mengenoperationen können wir dann aus $A, B \in \mathcal{F}$ neue Ereignisse bilden. Abbildungen 2.1 und 2.2 stellen Ereignisse als *Mengen* dar.

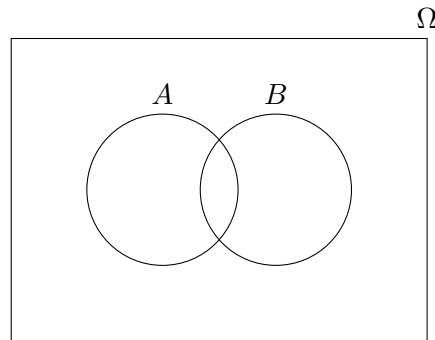


Abbildung 2.1: Venn diagram: Stichprobenraum Ω , Ereignisse A und B .

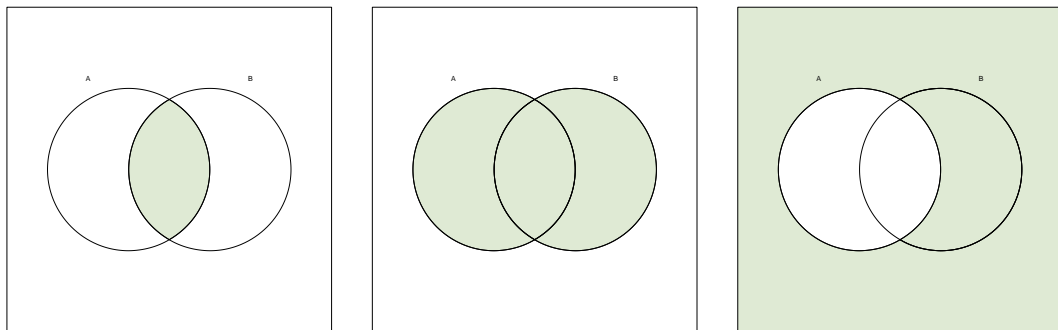
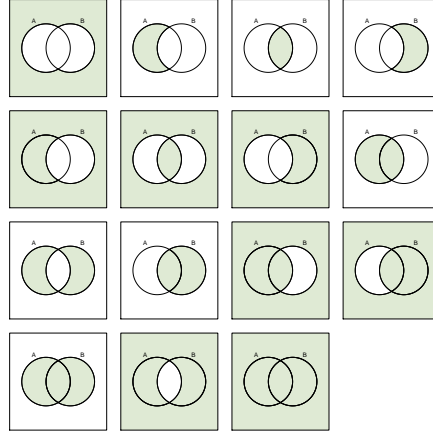


Abbildung 2.2: Schnittmenge $A \cap B$, Vereinigungsmenge $A \cup B$ und Komplement A^C

- Die *Schnittmenge* von zwei Ereignissen A und B besteht aus den Elementen, die sowohl zu A und zu B gehören. Wir notieren diese Menge $A \cap B$. Das ist das Ereignis, dass A und B eintreten.
- Die *Vereinigungsmenge* von A und B , bestehend aus den Elementen, die zu A oder B gehören, notieren wir mit $A \cup B$. Das ist das Ereignis, dass A oder B (oder beide) eintreten.
- Das Komplement von A , notiert mit A^C ist das Ereignis, dass A nicht eintritt.

Dank der Definition von \mathcal{F} liegen all diese Mengen auch wieder in \mathcal{F} .

Konstruktion von einem Ereignisraum*. Abbildung 2.3 illustriert den kleinsten Ereignisraum (das kleinste \mathcal{F}), der die zwei Ereignisse A und B beinhaltet. Es gibt $2^4 = 16$ Elemente in diesem Raum, *alle Vereinigungen* der Mengen in der ersten Reihe, $A^C \cap B^C, B \cap A^C, A \cap B^C, A \cap B$, die eine *Partition* von Ω darstellt. Unten rechts wäre die leere Menge.



Abbildungung 2.3: Ereignisraum \mathcal{F} aufbauend auf den Ereignissen A und B , die erste Zeile stellt eine Partition von Ω dar, die anderen Mengen sind Vereinigungen der Elemente der Partition.

Definition. Ein *Wahrscheinlichkeitsmass* ist eine Abbildung $\mathcal{F} \rightarrow [0, 1]$. Jedem Ereignis im Ereignisraum \mathcal{F} wird eine Zahl zwischen 0 und 1 zugeordnet. Für $A \in \mathcal{F}$ nennen wir $\Pr(A) \in [0, 1]$ die Wahrscheinlichkeit, dass A eintritt.

Bezüglich der Wahrscheinlichkeitsrechnung hat Kolmogoroff (1933) folgende *Axiome* aufgestellt. Diese Axiome gelten für alle Interpretationen von Wahrscheinlichkeit¹:

- Axiom 1: Die Wahrscheinlichkeit ist nicht negativ

$$\Pr(A) \geq 0 \quad (2.1.1)$$

- Axiom 2: Die Wahrscheinlichkeit des *sicheren* Ereignisses Ω ist 1

$$\Pr(\Omega) = 1 \quad (2.1.2)$$

- Axiom 3: Die Wahrscheinlichkeit der Vereinigung von *disjunkten* (sich gegenseitig ausschliessenden) Ereignissen A und B ist gleich der Summe der Wahrscheinlichkeiten der einzelnen Ereignisse. $A \cap B = \emptyset$, so

$$\Pr(A \cup B) = \Pr(A) + \Pr(B) \quad (2.1.3)$$

- Axiom 3b: Allgemein: Für jede Sequenz von paarweise disjunkten Ereignissen A_1, A_2, \dots (i.e., $A_i \cap A_j = \emptyset$, für $i \neq j$) gilt

$$\Pr\left(\bigcup A_i\right) = \sum \Pr(A_i) \quad (2.1.4)$$

¹objektive oder subjektive Wahrscheinlichkeit, dazu mehr später

Additionsregel. Aus den drei angeführten Axiomen folgen weitere wichtige Sätze der Wahrscheinlichkeitstheorie. Verwandt mit dem 3. Axiom (2.1.3) ist das sogenannte *Additionstheorem*. Es geht dabei um die allgemeine *Additionsregel* für beliebige A und B (die also nicht disjunkt sein müssen).

$$\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B). \quad (2.1.5)$$

Wir können (2.1.5) mit Hilfe der Abbildung 2.1 nachvollziehen. In der Sprache der Venn-Diagramme ist $\Pr(A) = \frac{|A|}{|\Omega|}$. Die Notation $|\cdot|$ steht für die *Mächtigkeit* der Menge, der Anzahl Elemente in der Menge. Die Additionsregel kann man nun auch allgemein für mehrere Ereignisse verallgemeinern. Hat man zum Beispiel drei Ereignisse A , B und C gegeben, so gilt

$$\Pr(A \cup B \cup C) = \Pr(A) + \Pr(B) + \Pr(C) - \Pr(A \cap B) - \Pr(A \cap C) - \Pr(B \cap C) + \Pr(A \cap B \cap C). \quad (2.1.6)$$

Die allgemeine Formel für n Ereignisse sieht dann aber auch nicht mehr so nett aus, deswegen lassen wir sie mal weg. Auf weitere Folgerungen gehen wir in 2.3 und 2.5 ein.

Am Ende dieses Abschnitts führen wir noch den Begriff der *Chance* oder *Odds* ein.

Definition. Eine Wahrscheinlichkeit p kann auch als *Chance* (oder englisch *Odds*) dargestellt werden. Chancen oder Odds von einem Ereignis sind definiert als das Verhältnis von der Wahrscheinlichkeit zur Gegenwahrscheinlichkeit von einem Ereignis,

$$Odds = \frac{p}{1-p}. \quad (2.1.7)$$

Umgekehrt gilt²

$$p = \frac{Odds}{1 + Odds}. \quad (2.1.8)$$

Odds können nicht-negative reelle Zahlen (Zahlen in \mathbb{R}_+) annehmen. Der Logarithmus der Odds heisst *logit*-Funktion,

$$y = \log \frac{p}{1-p} = \text{logit}(p). \quad (2.1.9)$$

Logits können Werte auf der ganzen reellen Zahlenachse \mathbb{R} annehmen. Die Umkehrfunktion von 2.1.9 nennt man die *logistische* Funktion,³

$$p = \frac{\exp(y)}{1 + \exp(y)}. \quad (2.1.10)$$

Diese zwei Funktionen werden wir später antreffen, wenn wir z.B. Wahrscheinlichkeiten (Zahlen im Intervall $[0, 1]$) auf den reellen Zahlenbereich $(-\infty, \infty)$ abbilden wollen und

² $O = p/(1-p) \Rightarrow O - Op = p \Rightarrow p + Op = O \Rightarrow p(O+1) = O \Rightarrow p = O/(1+O)$

³ $y = \log \frac{p}{1-p} \Rightarrow \exp(y) = \frac{p}{1-p} \Rightarrow p = \frac{\exp(y)}{1+\exp(y)}$

umgekehrt. Diese beiden Funktionen sind in der Abbildung 2.4 dargestellt.

```
curve(log(x/(1 - x)), from = 0, to = 1, xlab = "p", ylab = "logit", main = "logit-Funktion")
curve(exp(x)/(1 + exp(x)), from = -4, to = 4, xlab = "logit", ylab = "p", main = "logistische Funktion")
```

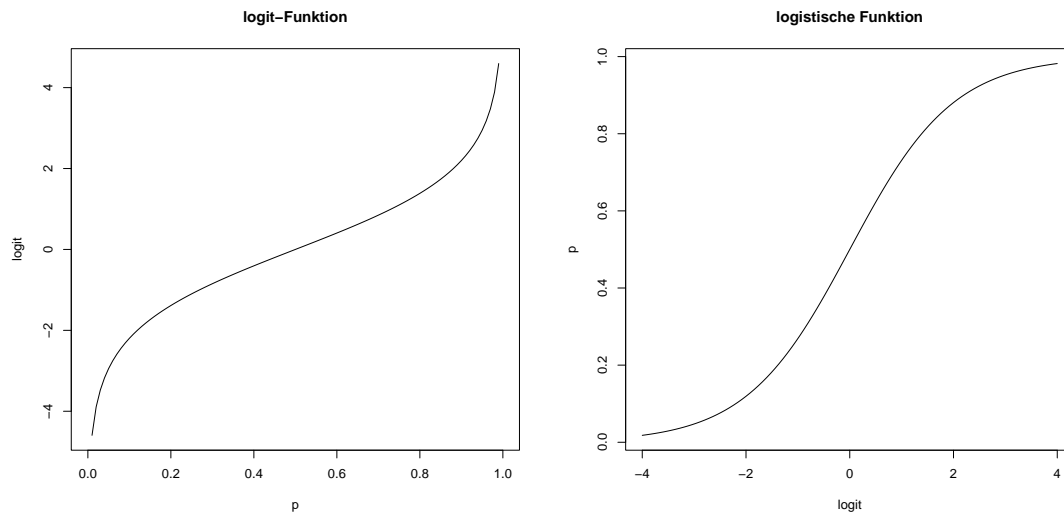


Abbildung 2.4: Die logit-Funktion (links) und die logistische Funktion (rechts).

2.2 Laplace-Wahrscheinlichkeit, objektive und subjektive Wahrscheinlichkeit

Was ist nun die *Wahrscheinlichkeit* $\Pr(A)$ von einem Ereignis A ? Wenn man davon ausgeht, dass alle möglichen Ergebnisse w_i *gleichwahrscheinlich* sind, dann stellt $\Pr(A)$ das Verhältnis der Anzahl Elemente in Teilmenge A relativ zur Anzahl der Elemente in der Menge Ω dar. Dies entspricht dem Verhältnis der *günstigen* zu den *möglichen* Ereignissen. Diese Wahrscheinlichkeit kennen wir aus der Schule, man nennt sie auch *Laplace-Wahrscheinlichkeit*. Sie ist in dem Sinne subjektiv, als man *a priori*, bevor man Daten sammelt, die Elementarereignisse als gleichwahrscheinlich annimmt.

Wenn wir z.B. eine Münze werfen, dann glauben wir *vor* dem Experiment an diese Gleichwahrscheinlichkeit von “Kopf” und “Zahl”. Dasselbe gilt beim Würfeln: Glauben wir an einen fairen Würfel, so ist die Wahrscheinlichkeit jedes Elementarereignisses $\{w_i\}$ aus $\Omega = \{1, 2, 3, 4, 5, 6\}$ a priori $\Pr(w_i) = 1/6$.

Laplace-Wahrscheinlichkeit:

$$\Pr(A) = \frac{\text{Anzahl günstiger Ergebnisse für } A}{\text{Anzahl möglicher Ergebnisse}}. \quad (2.2.1)$$

Die Laplace-Wahrscheinlichkeit geht also von der a priori postulierten Gleichwahrscheinlichkeit der Elementarereignisse aus. Sind nun aber die Elementarereignisse nicht a priori gleichwahrscheinlich, wie im Beispiel oben mit den Schulnoten, wird der *frequentistische* Wahrscheinlichkeitsbegriff gebraucht. Das ist z.B. der Fall, wenn wir nicht an eine *faire* Münze beim Münzwurf glauben.

Dann brauchen wir eine *gemessene* Wahrscheinlichkeit, eine Wahrscheinlichkeit, die wir anhand von Daten quantifizieren. Die Wahrscheinlichkeit eines Ereignisses A wird dann anhand der *relativen Häufigkeit* vom Eintreten von A bei *häufigem Wiederholen* eines *Zufallsexperiments* beschrieben. Die Wahrscheinlichkeit ist der *Grenzwert* dieser *relativen Häufigkeit* bei sehr häufiger Wiederholung.

Auch wenn man nicht unendlich viele Wiederholungen eines Zufallsexperiments machen und so den Grenzwert wirklich bestimmen kann, so ist es eine empirische Tatsache, dass sich relative Häufigkeiten bei oftmaligem Wiederholen um eine Grösse stabilisieren; diese Grösse nennen wir dann Wahrscheinlichkeit.

Die klassische Statistik baut auf dieser *frequentistischen Interpretation* von Wahrscheinlichkeit oder auf dem sogenannten *objektiven* Wahrscheinlichkeitsbegriff auf. Der frequentistische Wahrscheinlichkeitsbegriff ist der vorherrschende Wahrscheinlichkeitsbegriff in den Wissenschaften. Die Wahrscheinlichkeit $\Pr(A)$ für ein Ereignis A wird hier also durch die *relative Häufigkeit* $f_n(A)$ des Eintreffens von A bei n Wiederholungen quantifiziert, wobei diese Schätzung um so genauer ausfällt, je grösser die Anzahl an Wiederholungen n ist:

$$f_n(A) = \frac{\text{Anzahl der Ergebnisse } A \text{ in } n \text{ Messungen}}{n}. \quad (2.2.2)$$

Objektive Wahrscheinlichkeit:

$$\Pr(A) = \lim_{n \rightarrow \infty} f_n(A) \quad (2.2.3)$$

Neben der Laplace-Wahrscheinlichkeit und der objektiven Wahrscheinlichkeit gibt es noch einen *subjektiven* Wahrscheinlichkeitsbegriff. Dieser sieht Wahrscheinlichkeit als eine persönliche, subjektive Überzeugung des Betrachters. Diese Wahrscheinlichkeit kann mit *Wetten* quantifiziert werden, dabei ist der *Wettquotient* gerade die subjektive Wahrscheinlichkeit für das Eintreten von A . Die subjektive Wahrscheinlichkeit spielt vor allem in der “radikaleren” Variante der Bayes-Statistik eine Rolle, siehe Abschnitt 2.5. Auch der subjektive Wahrscheinlichkeitsbegriff erfüllt die Axiome von Kolmogoroff (siehe 2.1).

Subjektive Wahrscheinlichkeit: Die subjektive Wahrscheinlichkeit $\Pr(A)$ ist ein Mass für die persönliche Überzeugung und entspricht dem Wettquotienten für das Eintreten von A .

2.3 Bedingte Wahrscheinlichkeit

Definition. Seien A und B Ereignisse und $\Pr(A) > 0$. Die *bedingte Wahrscheinlichkeit* von B unter der Bedingung, dass A eintritt (*gegeben A*) wird definiert durch

$$\Pr(B \mid A) = \frac{\Pr(B \cap A)}{\Pr(A)}. \quad (2.3.1)$$

(2.3.1) kann man mit der Abbildung 2.1 nachvollziehen. $\Pr(B \cap A)$ stellt das Verhältnis der Anzahl Elemente in $B \cap A$ zur Anzahl Elemente in Ω dar. Wir sehen in der Abbildung, dass $\Pr(B \mid A) = \frac{|B \cap A|}{|A|}$, da $\Pr(B \cap A) = \frac{|B \cap A|}{|\Omega|}$ und $\Pr(A) = \frac{|A|}{|\Omega|}$.

Beispiel. Beim Würfeln seien die Ereignisse

- $A = \{\text{gerade Augenzahl}\} = \{2, 4, 6\}$,
- $B = \{\text{Augenzahl} > 3\} = \{4, 5, 6\}$.

Meine Kollegin würfelt und sagt mir, dass A eingetreten sei. Mit dieser zusätzlichen Information berechne ich dann $\Pr(B \mid A)$, die bedingte Wahrscheinlichkeit, dass auch B eingetreten ist. Dann ist $A \cap B = \{4, 6\}$ und $\Pr(B \mid A) = \frac{\Pr(B \cap A)}{\Pr(A)} = \frac{2/6}{3/6} = \frac{2}{3}$.

In der Regel ist $\Pr(B \mid A) \neq \Pr(B)$, so ist im obigen Beispiel $\Pr(B) = 1/2$. Bedingte Wahrscheinlichkeiten $\Pr(\cdot \mid A)$ können als Wahrscheinlichkeiten auf einem neuen Ergebnisraum $\Omega^* = A$ aufgefasst werden.

Multiplikationsregel. Durch Umformen von 2.3.1 erhalten wir die sogenannte *Multiplikationsregel*. Diese behandelt die Wahrscheinlichkeit von Schnittmengen oder der *gemeinsamen* Wahrscheinlichkeit von Ereignissen.

$$\Pr(A \cap B) = \Pr(A \mid B) \cdot \Pr(B) \quad (2.3.2)$$

oder

$$\Pr(A \cap B) = \Pr(B \mid A) \cdot \Pr(A). \quad (2.3.3)$$

Mit Hilfe der Multiplikationsregel kann man oft Wahrscheinlichkeiten auf einfache Weise in mehreren Schritten berechnen. Die gemeinsame Wahrscheinlichkeit für mehrere Ereignisse $A_i, i = 1, \dots, n$ ist dann

$$\Pr(A_1 \cap A_2 \cap A_3 \cdots \cap A_n) = \Pr(A_1) \Pr(A_2 \mid A_1) \Pr(A_3 \mid A_1, A_2) \cdots \Pr(A_n \mid A_1, \dots, A_{n-1}) \quad (2.3.4)$$

Die allgemeine Formel ist dann (für Interessierte*)

$$\Pr\left(\bigcap_{k=1}^n A_k\right) = \prod_{k=1}^n \Pr\left(A_k \mid \bigcap_{j=1}^{k-1} A_j\right). \quad (2.3.5)$$

Schauen wir uns einige Folgerungen von (2.3.1) an:

- Wenn A und B disjunkt sind, dann $\Pr(A | B) = 0 / \Pr(B) = 0$.
- Wenn $A \subset B$ (A ist eine Untermenge von B), dann $\Pr(A | B) = \Pr(A) / \Pr(B) < 1$, d.h., “ B ist eine nötige, aber nicht hinreichende Bedingung für A ” ($B \not\Rightarrow A$), siehe Abbildung 2.5. Beispiel: A : “Kraft > 100 Newton”, B : “Kraft > 50 Newton”.

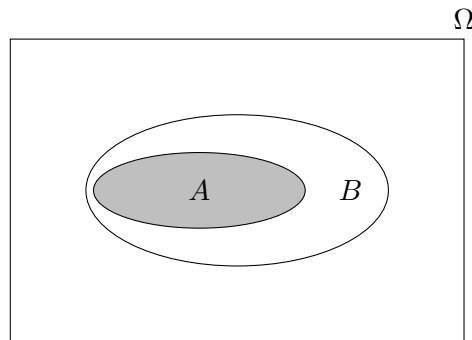


Abbildung 2.5: Ereignis $A \subset B$, $\Pr(A | B) < 1$, B impliziert nicht A .

- Wenn $B \subset A$, dann $\Pr(A | B) = \Pr(B) / \Pr(B) = 1$, d.h., A wird impliziert von B (B ist hinreichend für A) ($B \Rightarrow A$), siehe Abbildung 2.6. Beispiel: A : “ROM Knieflexion > 30 Grad”, B : “ROM Knieflexion > 90 Grad”.

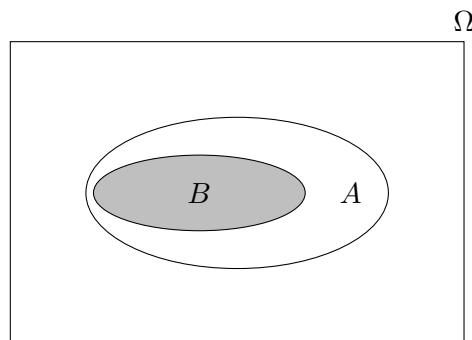


Abbildung 2.6: Ereignis $B \subset A$, $\Pr(A | B) = 1$, aus B folgt A .

- Die gemeinsame Wahrscheinlichkeit vom Ereignis $A \cap B$ kann nicht grösser sein als die Wahrscheinlichkeit von einem Ereignis A oder B . Bezogen auf Aussagen: *Eine Aussage kann nicht wahrscheinlicher sein als eine ihrer Folgerungen.* Beispiel: $A \cap B$: “Peter ist Sportler und seine Kraft ist grösser als 100 Newton”, A : “Peter ist Sportler”, B : “Kraft grösser als 100 Newton”.

Totale Wahrscheinlichkeit. Sei C_1, \dots, C_n eine Zerlegung (Partition) von Ω (in paarweise disjunkte Ereignisse, siehe Abbildung 2.7). Für beliebige Ereignisse A gilt dann

$$\Pr(A) = \sum_{i=1}^n \Pr(A \cap C_i) = \sum_{i=1}^n \Pr(A \mid C_i) \Pr(C_i). \quad (2.3.6)$$

Der Nutzen von 2.3.6 ist, dass manchmal die Berechnung von $\Pr(C_i)$ und $\Pr(A \mid C_i)$ einfacher ist als die direkte Berechnung von $\Pr(A)$.

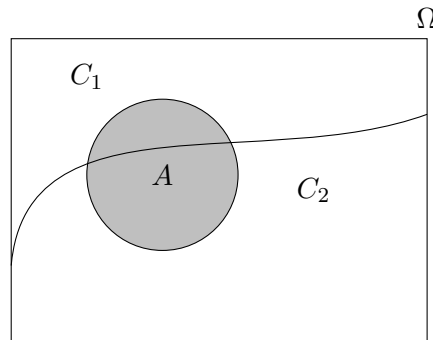


Abbildung 2.7: $\Pr(A) = \Pr(A \cap C_1) + \Pr(A \cap C_2) = \Pr(A \mid C_1) \Pr(C_1) + \Pr(A \mid C_2) \Pr(C_2)$

Beispiel. Bei einer Krankheitsdiagnose sind die folgenden Angaben bekannt:

- In der Bevölkerung sind 0.1% krank (*Prävalenz*).
- Von den kranken Personen werden 90% durch die Untersuchung entdeckt (*Sensitivität* des Tests).
- Von den gesunden Personen werden 99% durch die Untersuchung als gesund eingestuft (*Spezifität* des Tests).

Nun wird eine Person aus der Bevölkerung herausgegriffen, untersucht und als krank eingestuft. Wie wahrscheinlich ist es, dass das stimmt?

Sei A das Ereignis, dass eine zufällig gewählte Person krank ist, und B das Ereignis, dass die Untersuchung einer zufällig ausgewählten Person die Diagnose “krank” ergibt. Aus den Angaben haben wir

$$\begin{aligned} \Pr(A) &= 0.001, & \text{also } \Pr(A^C) &= 0.999 \\ \Pr(B \mid A) &= 0.9, & \text{also } \Pr(B^C \mid A) &= 0.1 \\ \Pr(B^C \mid A^C) &= 0.99, & \text{also } \Pr(B \mid A^C) &= 0.01 \end{aligned}$$

Die Wahrscheinlichkeiten $\Pr(B)$ und $\Pr(B^C)$ kennen wir (noch) nicht. Gesucht ist nun $\Pr(A \mid B)$, also die bedingte Wahrscheinlichkeit, dass unsere Person krank ist, gegeben,

dass sie als krank eingestuft wurde. Nach der Multiplikationsregel (2.3.2) gilt

$$\Pr(A \cap B) = \Pr(B \mid A) \Pr(A) = 0.9 \times 0.001 = 0.0009$$

und nach dem Satz der totalen Wahrscheinlichkeit (2.3.6)

$$\begin{aligned}\Pr(B) &= \Pr(B \mid A) \Pr(A) + \Pr(B \mid A^C) \Pr(A^C) \\ &= 0.9 \times 0.001 + 0.01 \times 0.999 \\ &= 0.01089\end{aligned}$$

Damit ist

$$\Pr(A \mid B) = \frac{\Pr(A \cap B)}{\Pr(B)} = \frac{0.0009}{0.01089} = 0.0826.$$

also sind 8.3% aller als krank diagnostizierten Personen tatsächlich krank. Eine Diagnose ist bei weitem nicht unfehlbar!

Verbessert man im obigen Beispiel die Diagnostik zu $\Pr(B \mid A^C) = 0.001$, d.h. zu weniger Falsch-Positiven Befunden, so ergibt sich analog $\Pr(A \mid B) = 0.4739$, d.h. die Vertrauenswürdigkeit einer Krank-Diagnose steigt von 8.3% auf 47.4%. Dieses Beispiel illustriert die Problematik der Fehldiagnosen bei eher seltenen Krankheiten.

Im obigen Beispiel haben wir bereits den Satz von *Bayes* benutzt, auf den wir bald zurückkommen (2.5). Das Beispiel ist in der Abbildung als *Baumdiagramm* 2.8 dargestellt.

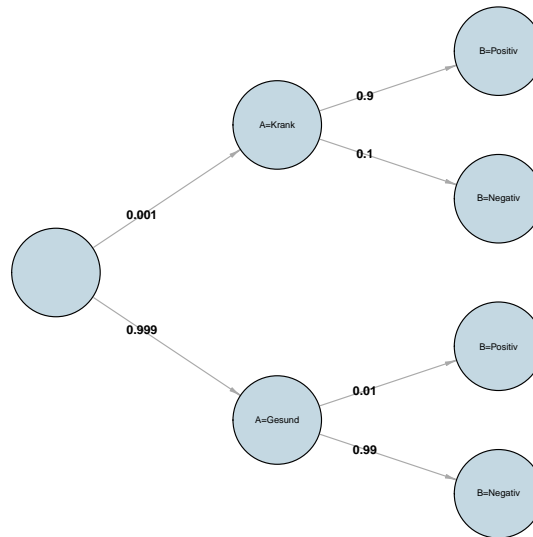


Abbildung 2.8: Baumdiagramm zum Beispiel der Diagnostik

2.4 Unabhängigkeit

Definition. Zwei Ereignisse A und B heißen (stochastisch) unabhängig, falls

$$\Pr(A \cap B) = \Pr(A) \cdot \Pr(B). \quad (2.4.1)$$

Bei Unabhängigkeit gilt

- $\Pr(A \mid B) = \Pr(A)$ (wenn $\Pr(B) \neq 0$)
- $\Pr(B \mid A) = \Pr(B)$ (wenn $\Pr(A) \neq 0$).

Unabhängigkeit bedeutet, dass das Eintreten eines Ereignisses keinen Einfluss hat auf die Wahrscheinlichkeit vom zweiten Ereignis. Bei Unabhängigkeit ist die Wahrscheinlichkeit für das gemeinsame Auftreten von A und B also gleich dem Produkt ihrer Einzelwahrscheinlichkeiten.

Beispiel. Abbildung 2.9 zeigt ein Beispiel für Unabhängigkeit. Sei A wieder das Ereignis, dass eine zufällig gewählte Person krank ist, und B das Ereignis, dass die

Untersuchung einer zufällig ausgewählten Person die Diagnose “krank” ergibt.

- $\Pr(A) = 0.3$
- $\Pr(B) = 0.9 \times 0.3 + 0.9 \times 0.7 = 0.9$
- $\Pr(A \cap B) = 0.3 \times 0.9 = \Pr(A) \cdot \Pr(B) = 0.3 \times 0.9$

Daraus folgt, dass $\Pr(A \mid B) = \Pr(A)$ und $\Pr(B \mid A) = \Pr(B)$, d.h. der Test bringt keine neue Information. Das Testresultat ist unabhängig vom Krankheitsstatus und der Krankheitsstatus ist unabhängig vom Testresultat!

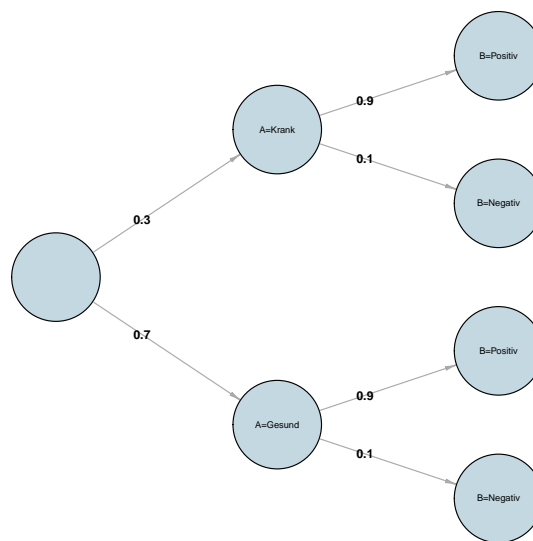


Abbildung 2.9: Unabhängigkeit von Krankheit und Testresultat

Definition. Die Ereignisse A_1, \dots, A_n heißen (stochastisch) unabhängig, wenn für jede endliche Teilfamilie die Produktformel gilt,

$$\Pr\left(\bigcap A_k\right) = \prod \Pr(A_k).$$

Unabhängige versus disjunkte Ereignisse. Wir haben oben gesehen: Wenn A und B disjunkt sind, dann $\Pr(A \mid B) = 0 / \Pr(B) = 0 \neq \Pr(A)$. Daraus folgt, dass disjunkte Ereignisse *nicht* unabhängig sind.

2.5 Der Satz von Bayes

In der sogenannten Bayes-Statistik braucht man z.T. einen subjektiven Wahrscheinlichkeitsbegriff, zumindest in der radikalen Variante des Bayesianismus. Wir wollen in diesem Kapitel nur die Grundlagen für die Bayes-Statistik betrachten.

Die bedingte Wahrscheinlichkeit $\Pr(A \mid B)$ haben wir schon kennengelernt. Diese Wahrscheinlichkeit ist die *bedingte* Wahrscheinlichkeit des Ereignisses A gegeben Ereignis B .

Betrachten wir nun ein bekanntes Gebiet mit bedingten Wahrscheinlichkeiten, nämlich das Gebiet der *Diagnostik*.

In der Diagnostik reden wir z.B. von der Wahrscheinlichkeit einer *Hypothese* H (z.B. Gegenwart von Krankheit) gegeben *Daten* E oder gegeben *Test* T . Dazu schauen wir uns das *Problem des Umkehrschlusses* an. Unter diesem Problem versteht man das irrtümliche Gleichsetzen der beiden bedingten Wahrscheinlichkeiten $\Pr(H \mid E)$ und $\Pr(E \mid H)$. Diese Wahrscheinlichkeiten sind aber i.A. nicht identisch, können aber durch das *Theorem von Bayes* verbunden werden. Dieses folgt direkt aus (2.3.2) und (2.3.3) und lautet

$$\Pr(H \mid E) = \frac{\Pr(E \mid H) \Pr(H)}{\Pr(E)}. \quad (2.5.1)$$

Das Theorem ist benannt nach dem englischen Mathematiker und Pfarrer THOMAS BAYES, der diesen Satz erstmals 1763 in einer posthum veröffentlichten Arbeit beschrieb [2]. In der wichtigen Gleichung (2.5.1) stellt $\Pr(H \mid E)$ die *a posteriori*-Wahrscheinlichkeit von H dar, $\Pr(H)$ stellt die *a priori*-Wahrscheinlichkeit (z.T. subjektive Überzeugung) von H dar (die Wahrscheinlichkeit der Hypothese, *bevor* man im Besitz von Daten ist). $\Pr(H)$ ist im besten Fall eine gute Schätzung der *Prävalenz* einer Krankheit, wenn z.B. H für “krank” steht. H^C bedeutet das Komplement von H , “gesund”.

Aufgrund der totalen Wahrscheinlichkeit (2.3.6) können wir (2.5.1) schreiben als

$$\Pr(H \mid E) = \frac{\Pr(E \mid H) \Pr(H)}{\Pr(E \mid H) \Pr(H) + \Pr(E \mid H^C) \Pr(H^C)}. \quad (2.5.2)$$

Definition. Die *Likelihood* einer Hypothese, $L(H)$, ist eine Funktion der Hypothese und stellt die Wahrscheinlichkeit der beobachteten Daten unter der Hypothese dar,

$$L(H) = \Pr(E \mid H). \quad (2.5.3)$$

Die Likelihood ist also ein Funktion des zweiten Arguments in $\Pr(E \mid H)$. Die Daten wurden beobachtet und die Likelihood der Hypothese ist gleich der Wahrscheinlichkeit der beobachteten Daten unter dieser Hypothese.

Bei zwei komplementären Hypothesen H und H^C ist

- die Likelihood von H : $L(H) = \Pr(E | H)$
- die Likelihood von H^C : $L(H^C) = \Pr(E | H^C)$

Teilen wir die a posteriori-Wahrscheinlichkeit von H in (2.5.1) durch diejenige von H^C . Dann kann man (2.5.1) auch in der *Odds-Form* schreiben.

$$\underbrace{\frac{\Pr(H | E)}{\Pr(H^C | E)}}_{\text{Posterior Odds}} = \underbrace{\frac{\Pr(E | H)}{\Pr(E | H^C)}}_{\text{Likelihood Ratio}} \times \underbrace{\frac{\Pr(H)}{\Pr(H^C)}}_{\text{Prior Odds}}. \quad (2.5.4)$$

Wir können das kurz schreiben, mit LR als der *Likelihood Ratio*,

$$Odds(H | E) = LR \times Odds(H). \quad (2.5.5)$$

Die a posteriori-Chance ist also das Produkt aus den a priori-Odds und dem durch die Daten bestimmten LR . Oder: Die LR transformiert die a priori-Chance von H zur a posteriori-Chance von H .

Wenn wir noch auf beiden Seiten den Logarithmus nehmen, dann können wir (2.5.5) additiv schreiben,

$$\boxed{\log Odds(H | E) = \log LR + \log Odds(H)}. \quad (2.5.6)$$

Die **philosophische Konsequenz** ist:

Erkenntnis in der Wissenschaft: A Posteriori = A Priori + empirische Daten

Betrachten wir als Spezialfall für die empirischen Daten E einen Test T ($T = 1$ für positives und $T = 0$ für ein negatives Resultat) sowie einer Hypothese der Krankheit ($H = 1$ für Krankheit versus $H = 0$ für Gesundheit). Dann sind folgende aus der Diagnostik bekannten Größen wichtig:

- $\Pr(T = 1 | H = 1)$: Sn des Tests (Richtig-Positiv-Rate, TPR)
- $\Pr(T = 0 | H = 0)$: Sp des Tests (Richtig-Negativ-Rate, TNR)
- $\Pr(T = 0 | H = 1)$: $1-Sn$ (Falsch-Negativ-Rate, FNR)
- $\Pr(T = 1 | H = 0)$: $1-Sp$ (Falsch-Positiv-Rate, FPR)
- $\Pr(H = 1 | T = 1)$: PPV des Tests
- $\Pr(H = 0 | T = 0)$: NPV des Tests
- $LR+ = \frac{L(H=1)}{L(H=0)} = \frac{\Pr(T=1|H=1)}{\Pr(T=1|H=0)} = \frac{TPR}{FPR} = \frac{Sn}{1-Sp}$: LR bei positivem Test
- $LR- = \frac{L(H=1)}{L(H=0)} = \frac{\Pr(T=0|H=1)}{\Pr(T=0|H=0)} = \frac{FNR}{TNR} = \frac{1-Sn}{Sp}$: LR bei negativem Test

Abkürzungen: Sn : Sensitivität, Sp : Spezifität, PPV : Positiv prädiktiver Wert, NPV : Negativ prädiktiver Wert, LR : Likelihood ratio.

Das Problem des Umkehrschlusses ist vor allem bei *seltenen Krankheiten* relevant. Ist die Prävalenz $\Pr(H = 1)$ einer Krankheit klein, dann kann $\Pr(H = 1 \mid T = 1)$ z.T. (viel) kleiner werden als $\Pr(T = 1 \mid H = 1)$.

Beispiel 1. Die Tabelle 2.2 zeigt die Resultate einer Studie, die die Sensitivität Sn und die Spezifität Sp , also die Gütekriterien bestimmen wollte für einen neuen Test für die Diagnose einer Krankheit. Mit den Daten dieser Studie können die Gütekriterien, prädiktiven Werte und Likelihood Ratios geschätzt werden: $Sn = 0.9$, $Sp = 0.9$.

| | | Wahrheit | | |
|------|---------|----------|---------|-------|
| | | $H = 1$ | $H = 0$ | Total |
| Test | $T = 1$ | 90 | 100 | 190 |
| | $T = 0$ | 10 | 900 | 910 |
| | Total | 100 | 1000 | 1100 |

Tabelle 2.1: Diagnostik bei einer seltenen Krankheit, $H = 1$: Krank, $H = 0$: Gesund, $T = 1$: Test positiv, $T = 0$: Test negativ. Die Berechnung der Gütekriterien ergibt $Sn = 0.9$ und $Sp = 0.9$.

Brauchen wir als (a priori) Prävalenz die beobachtete Prävalenz ($100/1100 = 0.091$), dann ist der $PPV = 0.474$ und der $NPV = 0.989$. Die Sensitivität Sn ist also viel grösser als der prädiktive Wert! Für den Patienten ist in diesem Fall ein positives Testresultat weniger dramatisch als man nach Betrachtung der hohen Sn und Sp meinen könnte.

Dieser Test eignet sich besser für den Ausschluss (weil NPV gross) und weniger für den Einschluss von Krankheit (PPV nicht gross). Für die LR 's ergibt sich $LR+ = 0.9/0.1 = 9$ und $LR- = 0.1/0.9 = 1/9$, die Wahrscheinlichkeit eines positiven Tests ist bei einem Kranken 9-mal grösser als bei einem Gesunden, und die Wahrscheinlichkeit eines negativen Tests ist bei einem Kranken 9-mal kleiner als bei einem Gesunden. (Allgemein ist natürlich $LR+$ nicht der Kehrwert von $LR-$).

Abbildung 2.10 zeigt für die Sn und Sp in diesem Beispiel, wie sich die Prävalenz auf den PPV und NPV auswirkt. Gestrichelt ist die beobachtete Prävalenz aus den erhobenen Daten.

```
p0obs <- 100/1100 #observed prevalence
p0 <- seq(0, by = 0.01, 1) #prevalences
Sn <- 0.9
Sp <- 0.9 #Sn and Sp
PPV <- (p0 * Sn)/(p0 * Sn + (1 - p0) * (1 - Sp)) #pos.pred. value
NPV <- (1 - p0) * Sp/((1 - p0) * Sp + p0 * (1 - Sn)) #neg. pred. value
plot(p0, PPV, xlab = "prevalence", ylab = "PPV", type = "l", ylim = c(0, 1))
abline(v = p0obs, h = (p0obs * Sn)/(p0obs * Sn + (1 - p0obs) * (1 - Sp)), lty = 2)
plot(p0, NPV, xlab = "prevalence", ylab = "NPV", type = "l", ylim = c(0, 1))
abline(v = p0obs, h = (1 - p0obs) * Sp/((1 - p0obs) * Sp + p0obs * (1 - Sn)), lty = 2)
```

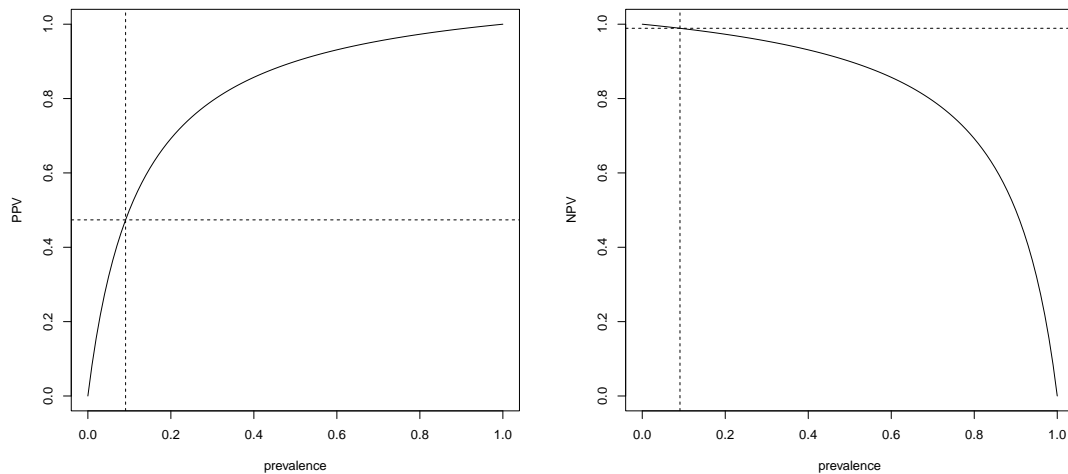


Abbildung 2.10: PPV und NPV als Funktion der Prävalenz, bei Test mit $Sn=0.9$, $Sp=0.9$. Gestrichelt: Beobachtete Prävalenz.

Bei niederprävalenten Merkmalen muss die Spezifität praktisch maximal sein (wenig Falsch-Positive), damit der positiv prädiktive Wert nicht absinkt. Denn schon wenige Falsch-Positive können den positiv prädiktiven Wert stark vermindern.

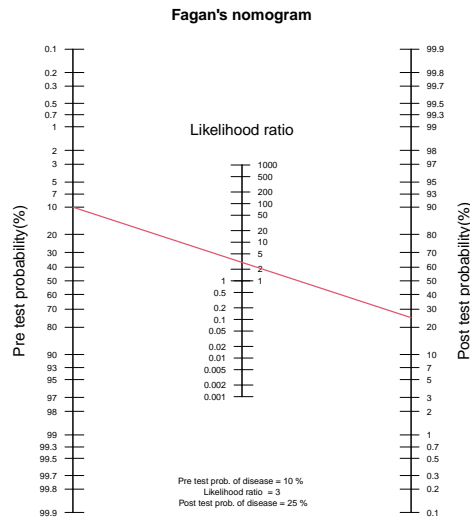
Will man also mit einem Test eine Krankheit (oder ein Merkmal wie “Blockade des Iliosakralgelenks”) *einschliessen*, muss man spezifisch sein, was gleichbedeutend ist mit wenig Falsch-Positiven und einem grossen positiven prädiktiven Wert. Ist dieses Merkmal zudem noch *selten*, muss die Spezifität fast maximal sein. Will man hingegen Krankheit *ausschliessen*, muss man sensitiv sein (wenig Falsch-Negative), ist zudem die Krankheit *häufig*, muss die Sensitivität fast maximal sein.

Beispiel 2. Im folgenden Beispiel veranschaulichen wir Bayes mit einem *Rechenschieber*, wie man ihn auch früher physisch zum Rechnen brauchte. Dieser basiert auf der additiven Bayes-Formel 2.5.6.

Wir nehmen eine Vortestwahrscheinlichkeit von Krankheit $\Pr(H = 1) = 0.1$ an, dies entspricht also einer prior Odds von 0.111. Zudem haben wir einen diagnostischen Test mit $Sn = \Pr(T = 1 \mid H = 1) = 0.3$ und $Sp = \Pr(T = 0 \mid H = 0) = 0.9$.

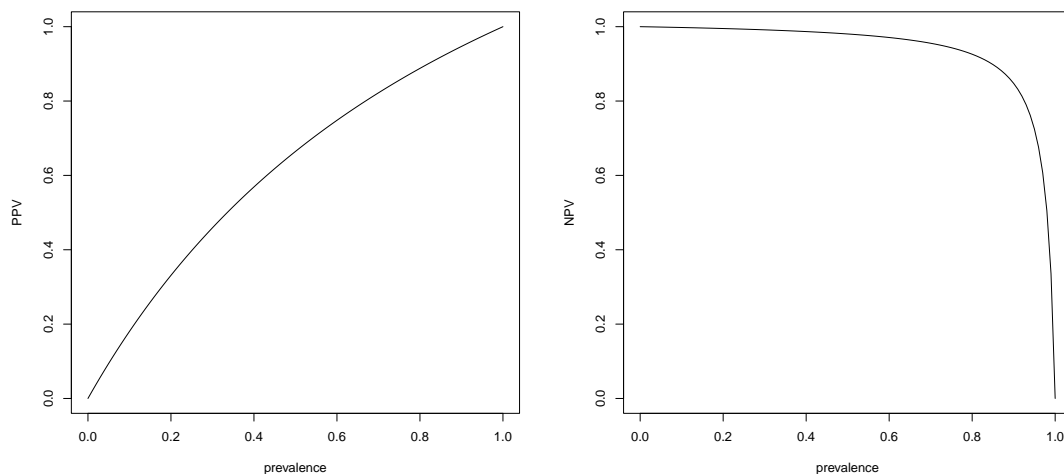
Die Likelihood ratio ($H = 1$ versus $H = 0$) **nach Beobachtung eines positiven Tests** (definiert als das Verhältnis der Wahrscheinlichkeit von einem positiven Test unter $H = 1$ zur Wahrscheinlichkeit eines positiven Tests unter $H = 0$) ist $LR+ = Sn/(1 - Sp) = 3$. Die resultierende Nachtestwahrscheinlichkeit ist $\Pr(H = 1 \mid T = 1) = 0.25$. Der Leser versuche, dies anhand (2.5.4) nachzurechnen.

```
## install.packages('TeachingDemos') ## auskommentieren für Installation der package, dann Zeile
## löschen!
library(TeachingDemos) ## package laden
p0 <- 0.1
sn <- 0.3
sp <- 0.9
fagan.plot(probs.pre.test = p0, LR = sn/(1 - sp))
```



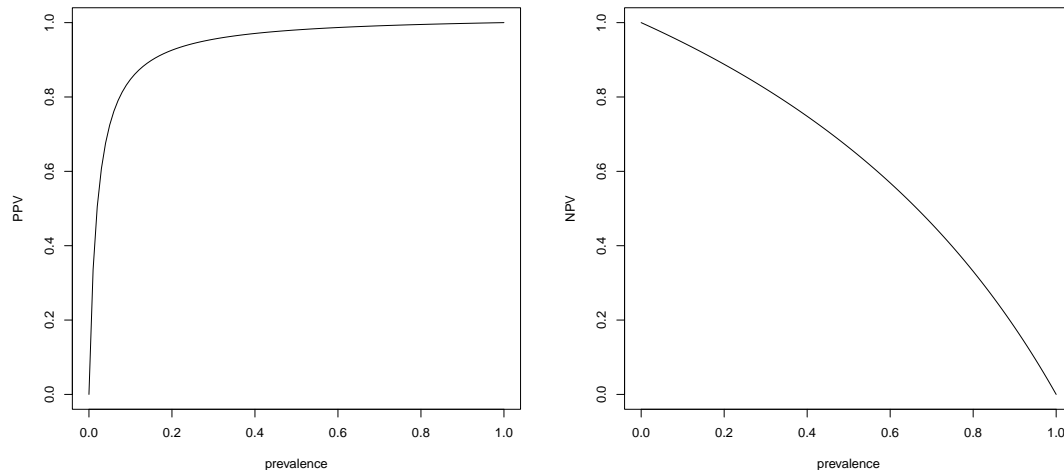
Mit `plotFagan2()` aus derselben package können die Vortestwahrscheinlichkeit, Sensitivität und Spezifität auch interaktiv eingegeben werden. Probiere es aus!

Visualisierung der Abhängigkeit von der Prävalenz bei unterschiedlichen S_n und S_p .
Beispiel 1: Prädiktive Werte bei hochsensitivem Test mit $S_n = 0.99$ und $S_p = 0.5$:



Test ist geeignet für das *Ausschliessen* von Krankheit, dazu brauchen wir eher einen sensitiven Test: “*SnNout*”, bei hoher *Sn*, wenn neg., dann ausschliessen

Beispiel 2: Prädiktive Werte bei hochspezifischem Test mit $Sn = 0.5$ und $Sp = 0.99$:



Test ist geeignet für das *Einschliessen* von Krankheit, dazu brauchen wir eher einen spezifischen Test: “*SpPin*”, bei hoher *Sp*, wenn pos., dann einschliessen.

Spezialfall: Gleichheit von a priori und a posteriori-Wahrscheinlichkeit. Wann sind diese beiden Wahrscheinlichkeiten gleich?

Aus (2.5.1) folgt: Damit $\Pr(H | T) = \Pr(H)$ gilt, muss $\Pr(T | H) = \Pr(T)$ gelten, d.h. die Wahrscheinlichkeit von T ist *unabhängig* von H . Das ist gleichbedeutend mit $LR = 1$ oder $\log LR = 0$:

- Positiver Test: $LR+ = \frac{\Pr(T=1|H=1)}{\Pr(T=1|H=0)} = \frac{TPR}{FPR} = 1 \Leftrightarrow TPR = FPR$.
- Negativer Test: $LR- = \frac{\Pr(T=0|H=1)}{\Pr(T=0|H=0)} = \frac{FNR}{TNR} = 1 \Leftrightarrow FNR = TNR$.

Spezialfall: Gleichheit von bedingten Wahrscheinlichkeiten. Wir haben gesehen, dass im Allgemeinen

$$\Pr(H | T) \neq \Pr(T | H).$$

Aus (2.5.1) folgt: Damit wir T und H in den bedingten Wahrscheinlichkeiten vertauschen dürfen, muss

$$\Pr(T) = \Pr(H)$$

gelten, wir haben dann *Symmetrie* bezüglich T und H . Das ist bei der Schätzung dann der Fall, wenn die Häufigkeiten symmetrisch sind, d.h. wenn wir gleich viele Falsch-Positive und Falsch-Negative Ereignisse haben. Das wäre der Fall bei der folgenden symmetrischen Vierfeldertafel (symmetrisch bezüglich der Hauptdiagonalen):

| | | Wahrheit | | |
|---------|---------|----------|----------|--------|
| | | $H = 1$ | $H = 0$ | Total |
| Empirie | $T = 1$ | a | b | a+b |
| | $T = 0$ | b | d | b+d |
| | Total | a+b | b+d | a+2b+d |

Tabelle 2.2: Symmetrie von T und H : $\Pr(H | T) = \Pr(T | H)$.

2.6 Bayesianische Statistik*.

Der Gebrauch vom Bayes-Theorem in der Diagnostik ist etabliert als *formalisierte klinische Argumentation*. Kontroverser ist die *Bayesianische Statistik*.

Wir werden später sehen, dass statistisches Schätzen und Testen eine Analogie hat zur Diagnostik.

Bei diagnostischen Tests ist die Hypothese wie oben z.B. “*Patient ist krank*”. In der schliessenden Statistik werden Hypothesen allgemeinere Aussagen sein, wie etwa Aussagen über einen unbekannten Parameter θ in einer Population, so wie “ $\theta \leq 177$ cm” für die unbekannte mittlere Körpergrösse in einer Population. Wir werden uns beim Schätzen (Kapitel ??) und Testen von Hypothesen (Kapitel ??) aber zunächst auf die klassische Statistik beschränken, auf die Quantifizierung der Wahrscheinlichkeit von empirischen Daten, gegeben eine Hypothese, also eigentlich auf obige Likelihood (2.5.3).

Die umgekehrte Wahrscheinlichkeit, die a posteriori-Wahrscheinlichkeit, ist nur über das Bayes-Theorem zu berechnen. Und damit wir die Wahrscheinlichkeit einer Hypothese berechnen können, brauchen wir grundsätzlich ein a priori. Die sogenannte Bayesianische Statistik braucht den Satz von Bayes in der statistischen Analyse, wo ein Parameter θ eine unbekannte Quantität ist (z.B. wie das obige θ). Die a priori Verteilung $p(\theta)$ muss dann spezifiziert werden. Dieser Schritt kann angesehen werden als eine natürliche Erweiterung auf die subjektive Interpretation von Wahrscheinlichkeit, siehe auch [24].

Nach diesem Einblick in das Bayes-Theorem wollen wir daher mit folgendem für jeden Praktiker heilsamen Satz schliessen:

Bayes: Ohne Subjektivität gibt es – strenggenommen – keine Erkenntnis.

Kapitel 3

Zufallsvariablen und ihre Verteilung

Meistens lässt sich ein *Zufallsexperiment* beschreiben durch eine Abbildung auf einem Grundraum Ω und durch die Wahrscheinlichkeiten, mit denen diese Abbildung die Werte im Wertebereich annimmt. Eine solche Abbildung nennt man *Zufallsvariable*, und die Wahrscheinlichkeiten werden durch die *Verteilung* der Zufallsvariable beschrieben.

3.1 Zufallsvariable

Definition. Eine *Zufallsvariable* ist eine Abbildung vom Ergebnisraum Ω auf die reellen Zahlen \mathbb{R} ,

$$X : \Omega \rightarrow \mathbb{R} \quad (3.1.1)$$

Eine Zufallsvariable stellt eine *unbekannte* Quantität dar, die einen Wert in einer Menge von Werten (Zahlen) einnehmen kann. Wir schreiben Zufallsvariablen mit grossen lateinischen Buchstaben, X, Y, \dots , vor deren Beobachtung. Nach Beobachtung brauchen wir kleine Buchstaben x, y, \dots für den spezifischen beobachteten Wert.

Kumulative Verteilungsfunktion. Die *kumulative Verteilungsfunktion* von X ist die Abbildung

$$P : \mathbb{R} \rightarrow [0, 1] \quad P(x) = \Pr(X \leq x). \quad (3.1.2)$$

Dabei ist $P(x) = \Pr(X \leq x) = \Pr(\{\omega \mid X(\omega) \leq x\})$, mit

- $\Pr(\{\omega \mid X(\omega) \leq x\})$: “Die Wahrscheinlichkeit des Ereignisses $\{\omega \mid X(\omega) \leq x\}$ ”.
- $\{\omega \mid X(\omega) \leq x\}$: “Die Menge aller Ergebnisse, für die die Zufallsvariable einen Wert kleiner gleich x annimmt”.

Bemerkung.* Bedingung an eine Zufallsvariable: Die Menge $\{X \leq x\} = \{\omega \mid X(\omega) \leq x\}$ muss für jedes x ein beobachtbares Ereignis, also in \mathcal{F} sein.

Beispiel 1. Wir betrachten die “Heilung” nach Behandlung von drei Patienten. Der Ergebnisraum ist $\Omega = \{GGG, G GK, GK K, K KG, K K K, K GG, K K G, K G K\}$. Ein

Beispiel für eine Zufallsvariable wäre die *Anzahl von Heilungen*.

| | | | | | | | | |
|-------------|-----|-----|-----|-----|-----|-----|-----|-----|
| ω | GGG | GGK | GKK | GKG | KKK | KGG | KKG | KGK |
| $X(\omega)$ | 3 | 2 | 1 | 2 | 0 | 2 | 1 | 1 |

Beispiel 2. Wir behandeln einen Patienten solange, bis er “geheilt” ist. Der Grundraum wäre dann $\Omega = \{G, KG, KKG, KKKG, KKKKG, \dots\}$. Die *Anzahl von Therapien bis zur Heilung* wäre dann eine mögliche Zufallsvariable:

| | | | | | | |
|-------------|---|----|-----|------|-------|--------|
| ω | G | KG | KKG | KKKG | KKKKG | KKKKKG |
| $X(\omega)$ | 1 | 2 | 3 | 4 | 5 | 6 |

In beiden Beispielen haben wir eine *diskrete* Zufallsvariable. Diskrete Zufallsvariable haben eine *abzählbare* Anzahl von Werten.

Wahrscheinlichkeitsmasse. Sei X eine diskrete Zufallsvariable¹. Dann ist die *Wahrscheinlichkeitsmasse* oder *Gewichtsfunktion*

$$p : \mathbb{R} \rightarrow [0, 1] \quad p(x) = \Pr(X = x). \quad (3.1.3)$$

Sie ordnet jedem Wert von X eine Wahrscheinlichkeit $p(x)$ zu. Im Gegensatz zur obigen Verteilungsfunktion geht es hier um die Wahrscheinlichkeit vom Ereignis $X = x$ und nicht vom Ereignis $X \leq x$.

Wahrscheinlichkeitsdichte. Wenn eine Zufallsvariable X mit beliebiger Präzision gemessen werden kann, hat sie in jedem noch so kleinen Intervall beliebig viele Werte, eine solche Variable nennen wir *kontinuierlich* oder *stetig*. Wir können dann nicht mehr von der Wahrscheinlichkeit *eines* ganz bestimmten Wertes reden. Die Wahrscheinlichkeit, dass z.B. eine Körpergröße *genau* 170.002019394932... cm ist, ist dann 0, also $\Pr(X = x) = 0$.

Wir können aber die Wahrscheinlichkeit angeben, mit der X in ein infinitesimal² kleines Intervall dx fällt,

$$\Pr(X \in [x, x + dx]) = p(x)dx. \quad (3.1.4)$$

Wahrscheinlichkeitsmasse und Wahrscheinlichkeitsdichte verhalten sich analog zu Masse und Dichte in der Physik. Also steht $p(x)dx$ für eine Wahrscheinlichkeitsmasse (“Dichte mal Volumen”) und die *Wahrscheinlichkeitsdichte* $p(x)$ ist dann

$$p : \mathbb{R} \rightarrow [0, \infty[\quad p(x) = \frac{\Pr(X \in [x, x + dx])}{dx} \quad (3.1.5)$$

¹Mit *abzählbarer* Anzahl von Werten.

²“ins unendlich Kleine gehend”

Dichten sind nicht mehr Zahlen zwischen 0 und 1, sondern Zahlen zwischen 0 und ∞ . Die Dichtefunktion ist also *fast* das Gegenstück zur Gewichtsfunktion für kontinuierliche Variablen. Wir werden **dieselbe Notation** $p(x)$ brauchen für Gewichtsfunktionen von diskreten Zufallsvariable und Dichtefunktionen von kontinuierlichen Zufallsvariable. Für die kumulative Verteilungsfunktion schreiben wir $P(x)$ für diskrete und kontinuierliche Zufallsvariable.

Abbildung 3.1 zeigt verschiedene Typen von Zufallsvariablen und ihre Verteilungen. Drei Verteilungen stellen die Gewichtsfunktion einer *diskreten* Zufallsvariablen dar, die unten rechts stellt die Dichtefunktion einer *kontinuierlichen* Zufallsvariablen dar.

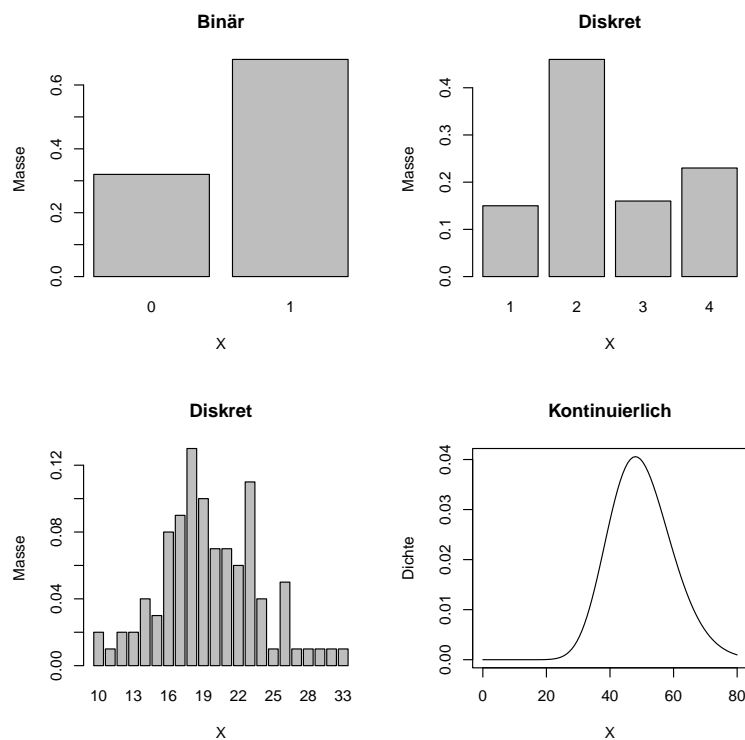


Abbildung 3.1: Typen von Verteilungen

3.2 Verteilungen

Diskrete Zufallsvariable. Abbildung 3.2 zeigt die Gewichtsfunktion $p(x)$ und die Verteilungsfunktion $P(x)$ für zwei diskrete Zufallsvariablen aus Abbildung 3.1. $P(x)$ ist eine Art Treppenfunktion, die an den Stellen x_i um den Wert p_i nach oben springt.

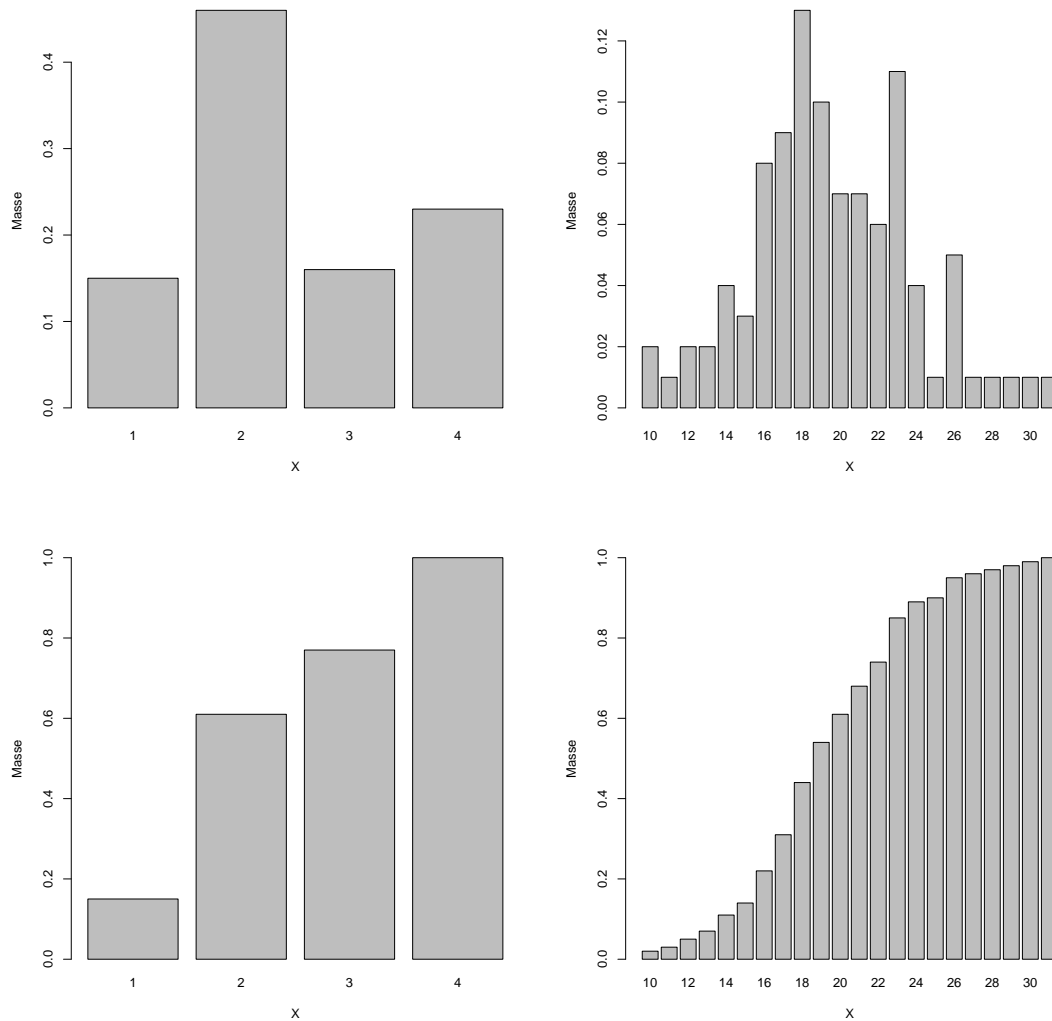


Abbildung 3.2: Diskrete Zufallsvariable. Gewichtsfunktion (oben) und kumulative Verteilungsfunktion (unten)

Stetige Zufallsvariable. Wenn wir eine Zufallsvariable X mit beliebiger Präzision messen können, ist Differential- und Integralrechnung nötig. Im Falle von stetigen Zufallsvariablen ist die Wahrscheinlichkeit für das Ereignis, dass X ein durch a und b begrenztes Intervall trifft, gegeben durch

$$\Pr(a \leq X \leq b) = \int_a^b p(x)dx = P(b) - P(a), \quad (3.2.1)$$

mit $p(x)$ als der oben eingeführten Wahrscheinlichkeitsdichte von X . Die Dichtefunktion wird zwischen den Grenzen a und b “aufsummiert” oder *integriert*. Abbildung 3.3 zeigt eine Dichtefunktion $p(x)$ zusammen mit der Wahrscheinlichkeit für das Ereignis, dass sich X in einem Intervall $[a, b]$ befindet. Wenn das Intervall $[a, b]$ den ganzen Wertebereich von X darstellt, dann ist die blaue Fläche=1. Es sei noch gesagt, dass bei kontinuierlichen Zufallsvariablen $a < X < b$ äquivalent ist mit $a \leq X \leq b$. Die Verteilungsfunktion bei kontinuierlicher Zufallsvariable ist

$$P(x) = \Pr(X < x) = \int_{-\infty}^x p(t)dt. \quad (3.2.2)$$

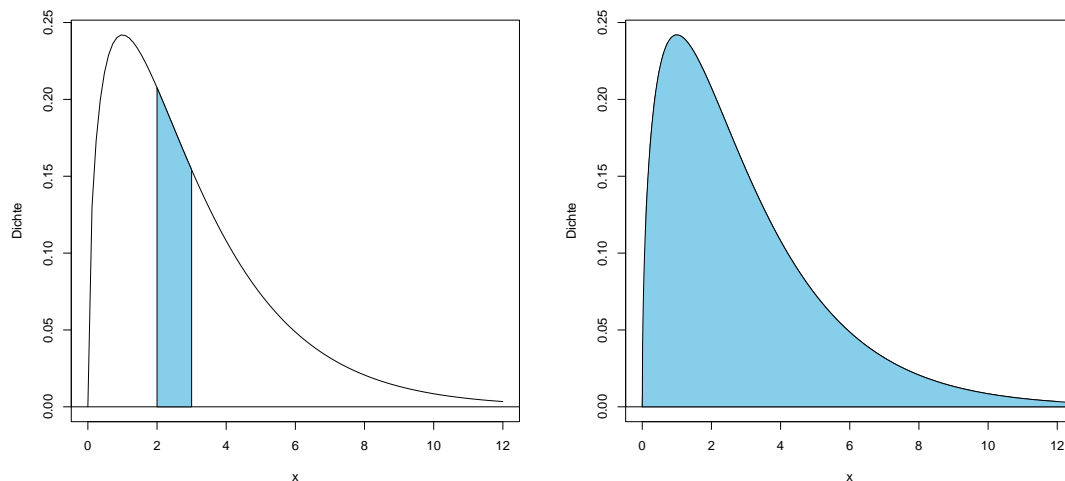


Abbildung 3.3: Kontinuierliche Zufallsvariable. Wahrscheinlichkeit für das Ereignis, dass $a < X < b$

Abbildung 3.4 zeigt die Dichtefunktion $p(x)$ und die Verteilungsfunktion $P(x)$ für eine kontinuierliche Zufallsvariable.

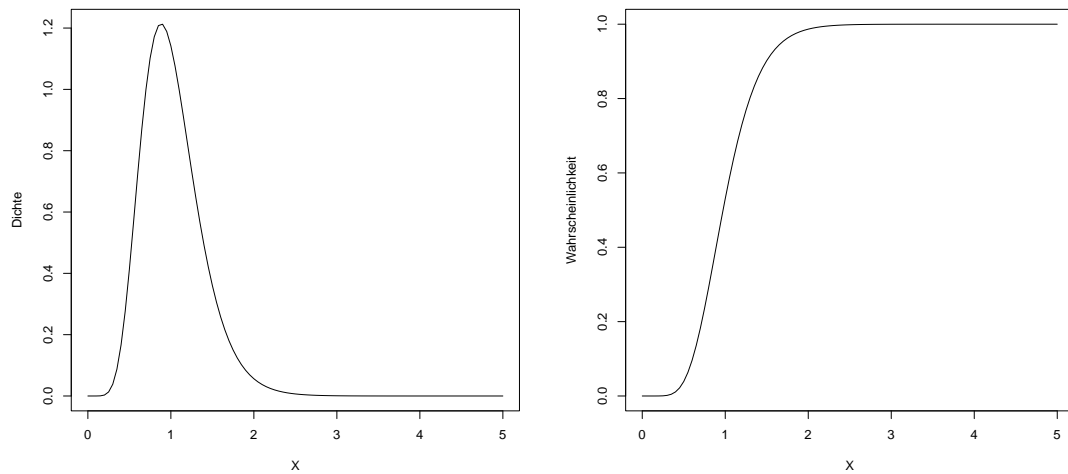


Abbildung 3.4: Kontinuierliche Zufallsvariable. Links: Dichtefunktion $p(x)$. Rechts: Kumulative Verteilungsfunktion $P(x)$

Abbildung 3.5 zeigt den Zusammenhang zwischen Dichtefunktion und kumulativer Verteilungsfunktion (3.2.2) für $X = 5$. Die Fläche unter der Dichtefunktion ist 0.828. Das können wir direkt in der Verteilungsfunktion ablesen.

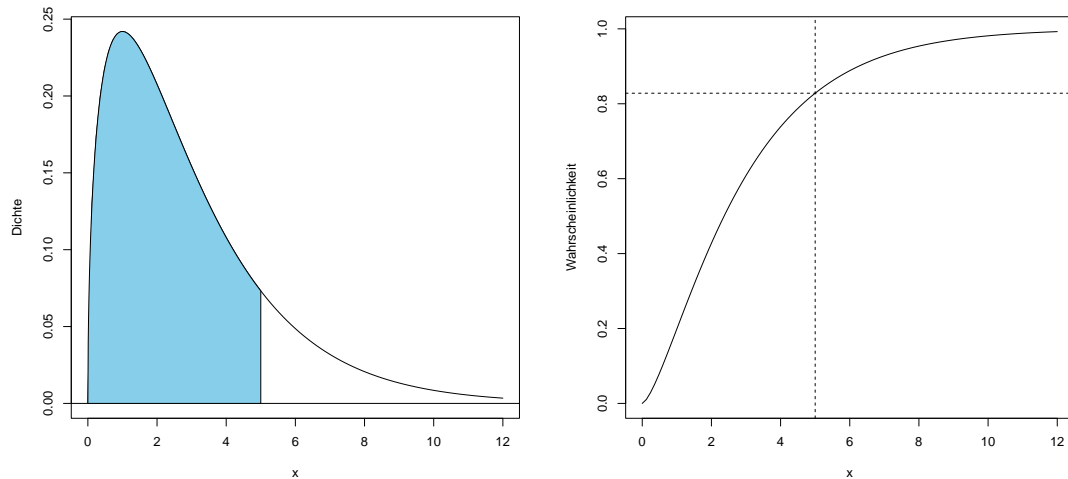


Abbildung 3.5: Kontinuierliche Zufallsvariable. Links: Integrierte Dichte bei $X = 5$. Rechts: Kumulative Verteilungsfunktion bei $X = 5$

Quantilfunktion. Die Umkehrfunktion der Verteilungsfunktion heisst *Quantilfunktion*. Sei $P(x) = \Pr(X \leq x)$ die kumulierte Wahrscheinlichkeit bei $X = x$. Wenn die Verteilung kontinuierlich ist und monoton steigend, dann ist das p -Quantil eindeutig.

$$Q_p = x \in \mathbb{R} \mid \Pr(X \leq x) = p \quad (3.2.3)$$

Besondere Quantile: $Q_{0.25}$ nennt man das erste *Quartil*, $Q_{0.5}$ den *Median* und $Q_{0.75}$ das dritte Quartil. $Q_{0.4}$ nennt man das vierte *Dezil*. $Q_{0.06}$ ist das sechste *Perzentil*. Abbildung 3.6 zeigt die Verteilungsfunktion und deren Umkehrfunktion (Quantilfunktion) bei einer kontinuierlichen Zufallsvariable X . Der Wert 0.524 entspricht z.B. der 70. Perzentile.

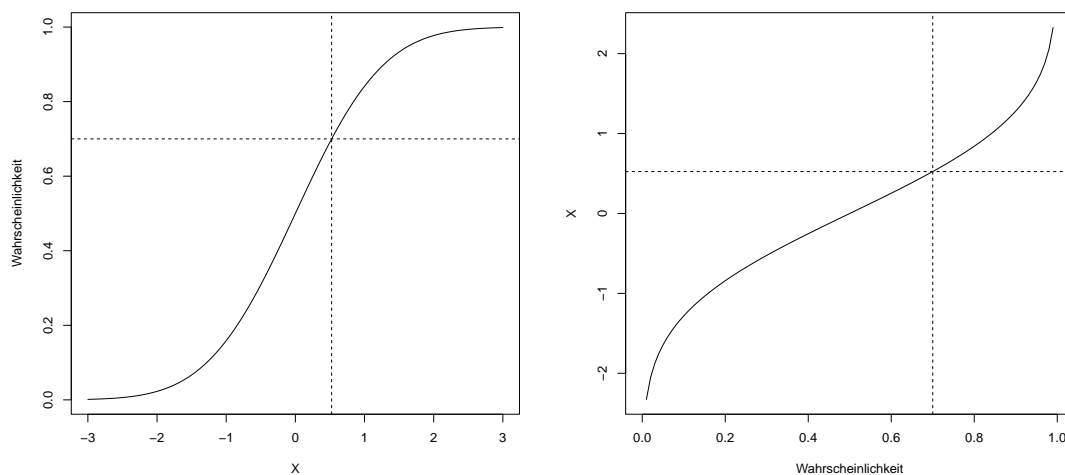


Abbildung 3.6: Links: Verteilungsfunktion. Rechts: Quantilfunktion

Bei diskreten Verteilungen wird es ein bisschen komplizierter³.

3.3 Wichtige Verteilungen

Es gibt – neben den sogenannten nichtparametrischen – viele parametrische Wahrscheinlichkeitsverteilungen. Wir lernen nun im Folgenden drei der wichtigsten Wahrscheinlichkeitsfunktionen kennen, die diskrete *Binomialverteilung*, die diskrete *Poisson-Verteilung* und die stetige *Normalverteilung*.

³Für die Interessierten*. Die Menge aller p -Quantile ist gegeben durch ein untere Grenze $\inf\{x \in \mathbb{R} \mid P(x) \geq p\}$ und einer oberen Grenze $\sup\{x \in \mathbb{R} \mid P(x) \leq p\}$ (*untere Schranke* von X mit kumulierter relativer Häufigkeit grösser gleich p respektive *obere Schranke* von X mit kumulierter relativer Häufigkeit kleiner gleich p). Beispiel Median: $p = 0.5$. Der *Untermedian* ist $\inf\{x \in \mathbb{R} \mid P(x) \geq 0.5\}$. Der *Obermedian* ist $\sup\{x \in \mathbb{R} \mid P(x) \leq 0.5\}$.

3.3.1 Binomialverteilung

Die *Binomialverteilung* ist eine diskrete Verteilung, die aufgrund wahrscheinlichkeitstheoretischer Überlegungen hergeleitet wird. Sie beruht auf dem einfachsten Zufallsexperiment. Wir betrachten n wiederholte Experimente mit dichotomem Ausgang. Zum Beispiel das Ereignis “Therapieerfolg beim i -ten Patienten”. π sei dann die – meist unbekannte – Eintretenswahrscheinlichkeit.

Wir nehmen dabei an, dass die Eintretenswahrscheinlichkeit π für jede Beobachtung gleich ist und dass die Beobachtungen voneinander *unabhängig* sind. (Das entspricht einem “Ziehen mit Zurücklegen”). Eine Zufallsvariable X heisst nun binomialverteilt mit den *Parametern* π und n , kurz $X \sim \text{Bin}(\pi, n)$, wenn die Gewichtsfunktion gegeben ist durch

$$p(x) = \binom{n}{x} \pi^x (1 - \pi)^{n-x}, \quad x = 0, 1, 2, \dots, n. \quad (3.3.1)$$

Dabei ist $\binom{n}{x} = \frac{n!}{(n-x)!x!}$ der *Binomialkoeffizient*. ($k!$ ist die *Fakultät* $k! = 1 \cdot 2 \cdot \dots \cdot k$). Dieser stellt die Anzahl der x -elementigen Teilmengen einer n -elementigen Menge dar (`choose(n=,k=)`). Diese Verteilung entsteht also allgemein, wenn X die Anzahl “Erfolge” in n Wiederholungen repräsentiert.

Abbildung 3.7 zeigt die Verteilung einer binomialverteilten Zufallsvariablen X bei $n = 15$ für 4 verschiedene Parameter π (Eintretenswahrscheinlichkeiten).

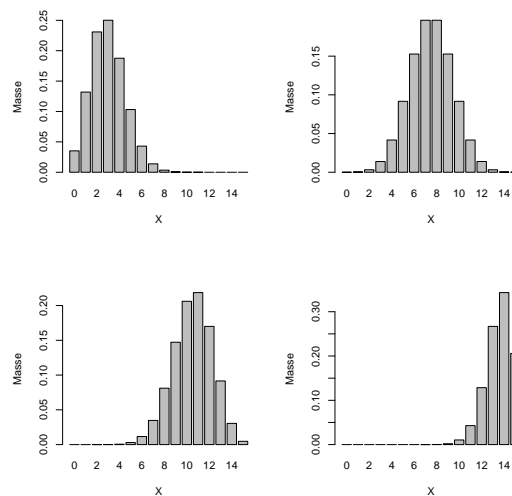


Abbildung 3.7: Binomialverteilung ($n = 15$) mit $\pi = 0.2, 0.5, 0.7$ and $\pi = 0.9$

Wenn $n = 1$, hat man den Spezialfall von einem Bernoulli trial, $X \sim \text{Bin}(\pi, 1)$. Abbildung 3.8 zeigt die Verteilung einer Bernoulli-verteilten Zufallsvariablen X für 4

verschiedene Parameter (Eintretenswahrscheinlichkeiten) π .

$$p(x) = \pi^x (1 - \pi)^{1-x}, \quad x = 0, 1. \quad (3.3.2)$$

Die Summe von n unabhängigen Bernoulli-verteilten Zufallsvariable mit identischem Parameter π ist dann binomialverteilt mit Parametern π und n .

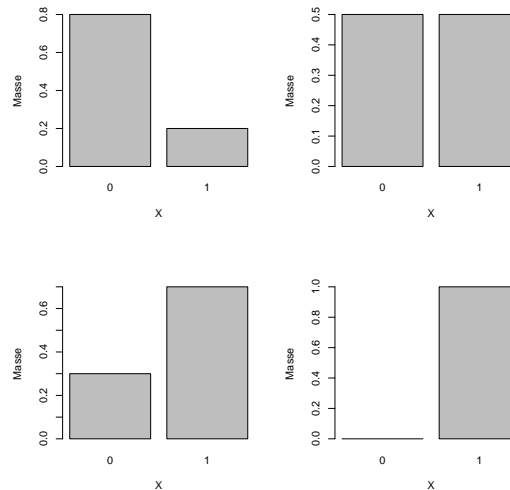


Abbildung 3.8: Bernoulli-Verteilung mit $\pi = 0.2, 0.5, 0.7$ und $\pi = 1$

3.3.2 Poisson-Verteilung

Die *Poisson-Verteilung* mit Parameter λ ist eine diskrete Verteilung mit Gewichtsfunktion

$$p(x) = \frac{\lambda^x}{x!} \exp(-\lambda), \quad x = 0, 1, 2, \dots \quad (3.3.3)$$

Wir schreiben kurz $X \sim \text{Pois}(\lambda)$.

Die Poisson-Verteilung ist nicht auf den Wertebereich $0, 1, 2, \dots, n$ beschränkt. Man erhält die Poisson-Verteilung durch den Grenzübergang aus der Binomialverteilung⁴.

Anzahlen von seltenen Ereignissen werden oft mit einer Poisson-Verteilung modelliert, z.b.

- die Anzahl Stürze in einer bestimmten Periode
- die Anzahl Erkrankungen in einer bestimmten Periode

⁴Wenn der Parameter n der Binomialverteilung gross und der zweite Parameter π klein wird, und wenn das Produkt $n\pi$ konstant bleibt, dann ist eine Approximation der Binomialverteilung durch die Poisson-Verteilung mit Parameter $\lambda = n\pi$ möglich.

Abbildung 3.9 zeigt die Masse von Poissonverteilungen für verschiedene Parameter λ .

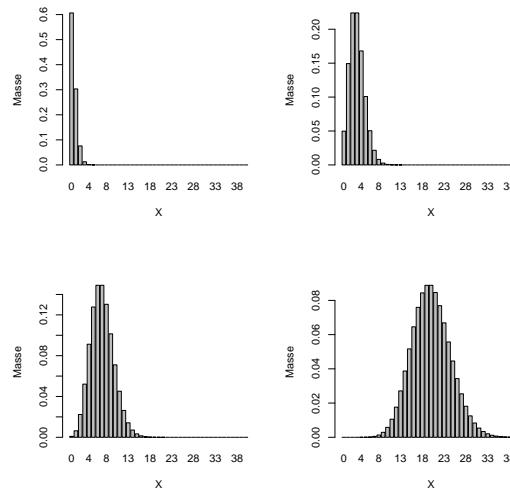


Abbildung 3.9: Poisson Masse mit $\lambda = 0.5, 3, 7, 20$

3.3.3 Normalverteilung

Lassen wir für die Binomialverteilung n gegen unendlich gehen, so wird daraus eine *Normalverteilung*. Wenn n gegen unendlich geht, wenn man also das Zufallsexperiment sehr oft wiederholt, gibt es immer mehr “mögliche Ausprägungsgrade” der Zufallsvariablen, der Wertebereich wird immer “stetiger”.

Aus der *diskreten* Binomialverteilung wird dann als Grenzfall eine *stetige* Normalverteilung. So kann z.B. das Merkmal $X = \text{Körpergröße}$ in einer Population normalverteilt sein. Die Normalverteilung hat zwei Parameter, μ und σ^2 . Wir werden später sehen, dass die Parameter dem sogenannten *Erwartungswert* respektive der sogenannten *Varianz* entsprechen. Man schreibt dann auch

$$X \sim \mathcal{N}(\mu, \sigma^2) \quad (3.3.4)$$

Die Normalverteilung ist überaus wichtig aus verschiedenen Gründen:

- Empirische Verteilungen haben häufig die Form einer Normalverteilung, viele Merkmale in der Natur sind normalverteilt. Das Grenzwerttheorem (siehe dazu ??) besagt, dass die Summe einer grossen Anzahl von Zufallsgrößen normalverteilt sind.
- Die Normalverteilung ist ein wichtiges Verteilungsmodell für statistische Kennwerte. So sind z.B. Mittelwerte und viele andere Größen, die man aus

Stichproben von wiederholten Stichprobennahmen berechnet, approximativ normalverteilt.

- Die Normalverteilung ist wichtig in der statistischen Fehlertheorie, z.B. werden Residuen aus Regressionsmodellen als normalverteilt mit Mittelwert 0 angenommen. Wir wissen aus dem Alltag, dass sich viele Zufallseinflüsse gegenseitig annullieren, wenn man sie mittelt.
- Die Normalverteilung ist eine mathematische Basisverteilung für viele andere Verteilungen, insbesondere für die t , die χ^2 und die F -Verteilung.

Eine normalverteilte Variable X mit Parametern μ und σ^2 hat die Dichte

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \quad x \in \mathbb{R}. \quad (3.3.5)$$

Standardnormalverteilung. Man kann zeigen, dass wenn X normalverteilt ist mit Parametern μ und σ^2 , also $X \sim \mathcal{N}(\mu, \sigma^2)$, dann ist *standardisierte* Variable $Z = \frac{X-\mu}{\sigma}$ *standardnormalverteilt*, $Z \sim \mathcal{N}(0, 1)$.

Für $\mu = 0$ und $\sigma^2 = 1$ erhält man also die Standardnormalverteilung mit der Dichte

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right). \quad (3.3.6)$$

Wollen wir die Wahrscheinlichkeit quantifizieren, mit der eine standardnormalverteilte Zufallsvariable X einen Wert x unterschreitet oder Werte zwischen zwei Grössen a und b annimmt, müssen wir die Dichte (Gleichung 3.3.6) *integrieren*, siehe (3.2.2). Dies ergibt die kumulative Verteilungsfunktion der Standardnormalverteilung,

$$\Phi(x) = \Pr(X < x) = \int_{-\infty}^x \phi(t) dt \quad (3.3.7)$$

Die Dichte $p(x)$ und kumulierte Verteilung $P(x)$ einer $\mathcal{N}(0, 1)$ -verteilten Zufallsvariablen wird mit $\phi(x)$ respektive $\Phi(x)$ notiert. Diese sind in Abbildung 3.10 dargestellt.

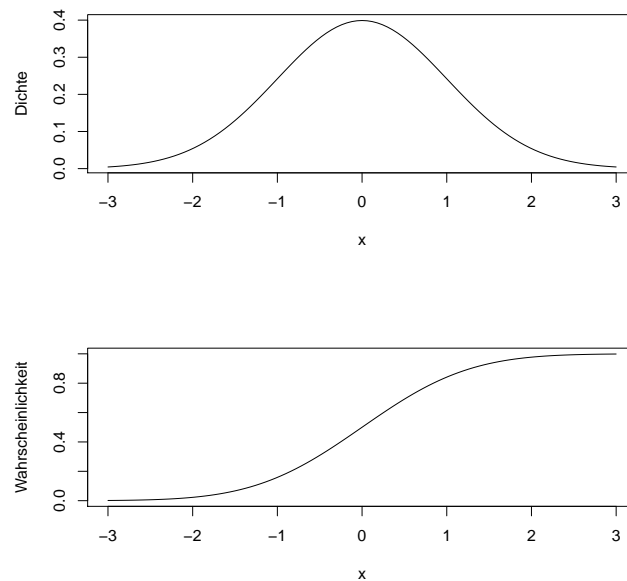


Abbildung 3.10: Wahrscheinlichkeitsdichte $\phi(x)$ (oben) und Verteilungsfunktion $\Phi(x)$ (unten) der Standardnormalverteilung.

Die Wahrscheinlichkeiten der Verteilungsfunktion kann man nicht aus einer geschlossenen Formel berechnen. Sie sind jedoch *tabelliert*. Heute werden sie aber eher in Computerprogrammen abgerufen, wir kommen bald darauf zurück. Aus den tabellierten Werten von $\Phi(x)$ kann man alle Wahrscheinlichkeiten berechnen, mit denen die standardnormalverteilte Variable X einen Wert zwischen a und b annimmt. Für jedes a und $b > a$ gilt

$$\Pr(a < X < b) = \Phi(b) - \Phi(a). \quad (3.3.8)$$

Die Wahrscheinlichkeiten von Gleichung 3.3.7 sind für verschiedene Quantile in der z -Tabelle 3.1 tabelliert. $\Phi(0) = 0.5$, $\Phi(1.96) = 0.975$, $\Phi(-1.96) = 1 - \Phi(1.96) = 0.025$. Umgekehrt gilt: $Q_{0.975} = 1.96$, $Q_{0.025} = -1.96$.

Abbildung 3.11 illustriert eine empirische (1000 simulierte standardnormalverteilte Zufallszahlen) und eine theoretische Normalverteilung. Abbildung 3.12 zeigt die Wahrscheinlichkeiten, mit der eine standardnormalverteilte Zufallsvariable Werte in gewissen Bereichen einnimmt.

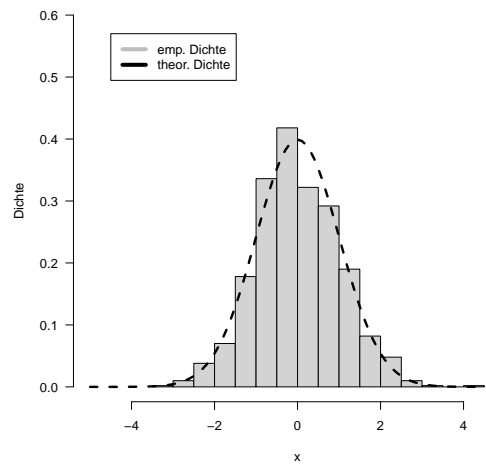


Abbildung 3.11: Empirische und theoretische Normalverteilung

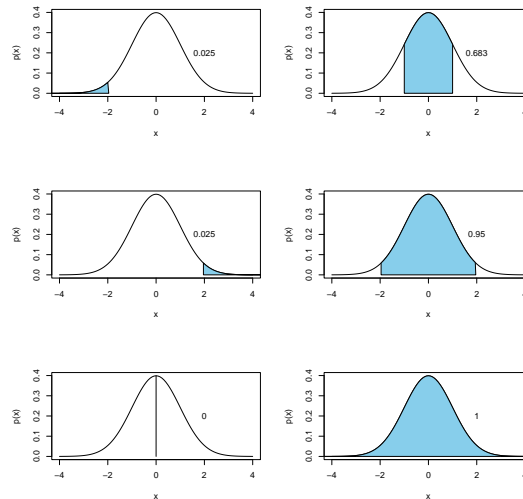


Abbildung 3.12: Wahrscheinlichkeit von verschiedenen Ereignissen

3.4 Andere Verteilungen

Es gibt sehr viele andere diskrete und stetige parametrische Verteilungen; wir wollen nachfolgend nur auf die wichtigsten eingehen.

Die χ^2 -, die t - und die F -Verteilung spielen eine wichtige Rolle, weil viele *Statistiken* χ^2 , t - oder F -verteilt sind. Diese Verteilungen bestehen aus ganzen Familien von Verteilungen, da sie im Gegensatz zur Normalverteilung noch durch sogenannte *Freiheitsgrade* (“Degrees of freedom”) spezifiziert sind. Das ist die Anzahl der Werte in der Berechnung einer Statistik, die frei variieren dürfen. Sie quantifizieren die Anzahl an *unabhängigen* Informationen, die für die Schätzung von Parametern zur Verfügung stehen.

χ^2 -Verteilung. Die Chi-Quadrat-Verteilung kann aus der Normalverteilung $\mathcal{N}(\mu, \sigma^2)$ abgeleitet werden. Hat man n *unabhängige standardnormalverteilte* i.i.d. (independent and identically distributed) Zufallsvariablen $Z_i \sim \mathcal{N}(0, 1)$, so ist die Chi-Quadrat-Verteilung mit n Freiheitsgraden definiert als die Verteilung von

$$Q_n = Z_1^2 + Z_2^2 + \cdots + Z_n^2. \quad (3.4.1)$$

Diese Grösse ist dann χ^2 -verteilt mit n Freiheitsgraden, $Q_n \sim \chi_n^2$. Solche Summen von quadrierten standardnormalverteilten Zufallsvariable werden wir später bei Schätzfunktionen wie der Stichprobenvarianz antreffen. Wir brauchen diese Verteilung bei sogenannten χ^2 -Verfahren. Die χ^2 -Verteilung hängt ebenfalls zusätzlich von einem Freiheitsgrad ab. Abbildung 3.13 zeigt einige Vertreter der χ^2 -Verteilung mit verschiedenen Freiheitsgraden.

t -Verteilung. Sind X und Y unabhängig mit $X \sim \mathcal{N}(0, 1)$ und $Y \sim \chi_n^2$, so ist

$$T = \frac{X}{\sqrt{\frac{Y}{n}}} \quad (3.4.2)$$

t -verteilt mit n Freiheitsgraden. Wir werden im Kapitel ?? sehen, dass die Zufallsvariable *Mittelwert* normalverteilter Daten nicht mehr normalverteilt, sondern t -verteilt ist, wenn die Varianz des Merkmals unbekannt ist und mit der Stichprobenvarianz geschätzt werden muss. Die Abbildung 3.14 zeigt einige Vertreter der t -Verteilung mit verschiedenen Freiheitsgraden. Die t -Verteilung kann bei Freiheitsgraden grösser als 30 praktisch durch die Normalverteilung approximiert werden. Die t -Verteilungen brauchen wir bei sogenannten t -Test-Verfahren. Diese zählen zu den bekanntesten Tests.

F-Verteilung. Die F -Verteilung kann aus der χ^2 -Verteilung abgeleitet werden, sie ist die Verteilung der Zufallsvariable

$$F_{n,m} = \frac{\chi_n^2/n}{\chi_m^2/m}, \quad (3.4.3)$$

mit χ_n^2 und χ_m^2 als unabhängigen Chi-Quadrat-verteilten Zufallsvariablen mit n bzw. m Freiheitsgraden (Beispiel für $n = 3$ und $m = 100$ in Abbildung 3.15). Diese Statistik spielt vor allem in klassischen *Varianzanalysen* eine wichtige Rolle.

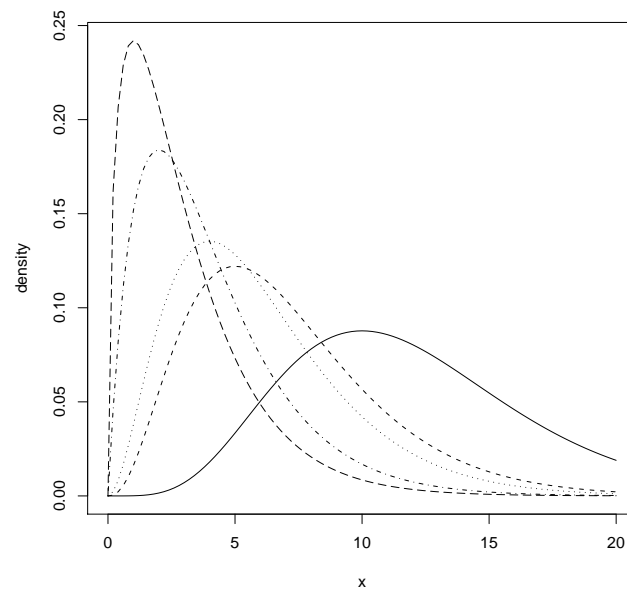


Abbildung 3.13: χ^2 -Verteilungen für $n = 12$ (—), $n = 7$ (--), $n = 6$ (\cdots), $n = 4$ (-.) und $n = 3$ (- -) Freiheitsgrade.

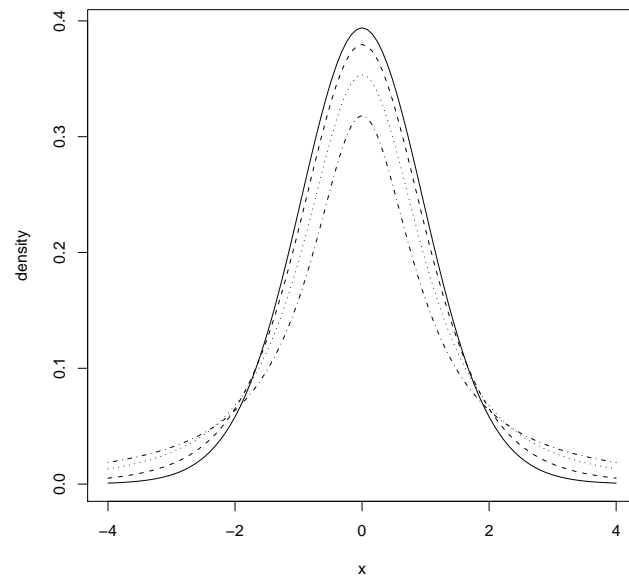


Abbildung 3.14: t -Verteilungen für $n = 1(\cdots)$, $n = 2(\cdots)$, $n = 5(---)$, $n = 20(—)$ Freiheitsgrade.

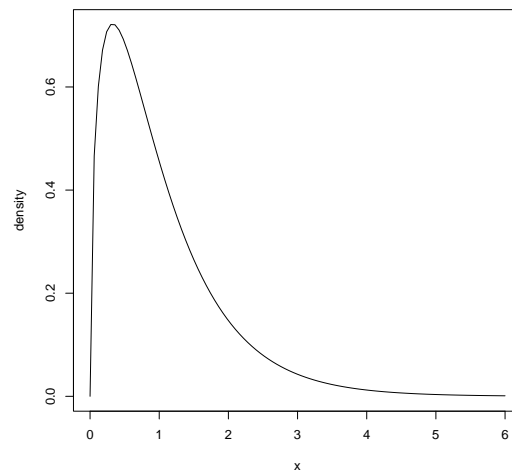


Abbildung 3.15: $F_{3,100}$ -Verteilung von $F = \frac{\frac{\chi_3^2}{3}}{\frac{\chi_{100}^2}{100}}$

3.5 Verteilungen mit R

Dieser Abschnitt ist für unseren Kurs zentral! Es gibt R-Funktionen, mit denen man

1. Zufallszahlen aus einer Verteilung ziehen kann
2. Die Dichte (oder Masse) bei $X = x$ angeben kann
3. Die kumulative Verteilung bei $X = x$ angeben kann
4. Quantile bei einer gewissen kumulierten Wahrscheinlichkeit herausgeben lassen kann

In R steht dann (unabhängig von der Verteilung)

- `r` für Zufallszahlen (random)
- `d` für Dichte **oder** Masse (density)
- `p` für Verteilungsfunktion (probability)
- `q` für die Quantilfunktion (quantile).

Für die Standardnormalverteilung z.B. bekommt man

- n Zufallszahlen mit `rnorm(n=,mean=0,sd=1)`
- Dichte d beim Quantil x mit `dnorm(x=,mean=0,sd=1)`
- Verteilungsfunktion p beim Quantil q mit `pnorm(q=,mean=0,sd=1)`
- Quantil q bei Verteilungsfunktion p mit `qnorm(p=,mean=0,sd=1)`.

Mit dem Kommando `?dnorm` oder `help(dnorm)` findet man sehr instruktive Hilfe zu den entsprechenden Funktionen.

Beispiele zur Normalverteilung. Gebe folgende Befehle in die R-Konsole ein:

- Kumulierte Wahrscheinlichkeit bei $X = 1.96$, wenn $X \sim \mathcal{N}(0, 1)$

```
pnorm(q = 1.96, mean = 0, sd = 1)
## [1] 0.975
```

- Das 0.975-Quantil von $X \sim \mathcal{N}(0, 1)$

```
qnorm(p = 0.975, mean = 0, sd = 1)
## [1] 1.96
```

- Dichte bei $X = 2$, wenn $X \sim \mathcal{N}(0, 1)$

```
dnorm(x = 2, mean = 0, sd = 1)
## [1] 0.054
```

- Wahrscheinlichkeit von $X < 1.96$, wenn $X \sim \mathcal{N}(0, 1)$

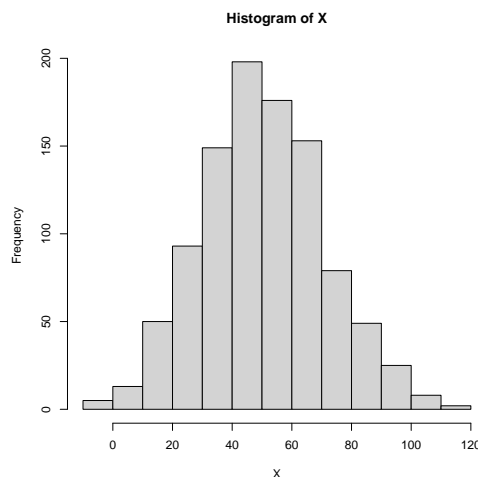

```
pnorm(q = 1.96, mean = 0, sd = 1)
## [1] 0.975
```

- Wahrscheinlichkeit von $-1.96 < X < 1.96$, wenn $X \sim \mathcal{N}(0, 1)$

```
pnorm(q = 1.96, mean = 0, sd = 1) - pnorm(q = -1.96, mean = 0, sd = 1)
## [1] 0.95
```

- Ziehen von $n = 1000$ Zufallszahlen aus $X \sim \mathcal{N}(50, 400)$

```
X <- rnorm(n = 1000, mean = 50, sd = 20) #Achtung: in R müssen wir sigma eingeben, nicht sigma^2
hist(X) #Wir kommen auf diesen Befehl in der deskriptiven Statistik zurück
```



Viele andere Verteilungen sind implementiert.

- `dbinom()`, `pbinom()`, `qbinom()`, `rbinom()` für Binomialverteilung
- `dpois()`, `ppois()`, `qpois()`, `rpois()` für Poissonverteilung
- `dt()`, `pt()`, `qt()`, `rt()` für t -Verteilung
- `dF()`, `pF()`, `qF()`, `rF()` für F -Verteilung
- `dchisq()`, `pchisq()`, `qchisq()`, `rchisq()` für χ^2 -Verteilung

Welche Argumente man in die Funktion einsetzen muss, kann man mit der Hilfefunktion erkennen. Schreibe `?dbinom` und `?dpois` in die Konsole und lese das Hilfe-File, vor allem die Teile **Description**, **Usage**, **Arguments** und **Details**.

Beispiele für andere Verteilungen.

- Masse bei $X = 2$, wenn $X \sim \text{Bin}(\pi = 0.6, n = 20)$

```
dbinom(x = 2, size = 20, prob = 0.6)
## [1] 0.0000047
```

- Wahrscheinlichkeit von $X \leq 1$, wenn $X \sim \text{Pois}(\lambda = 3)$

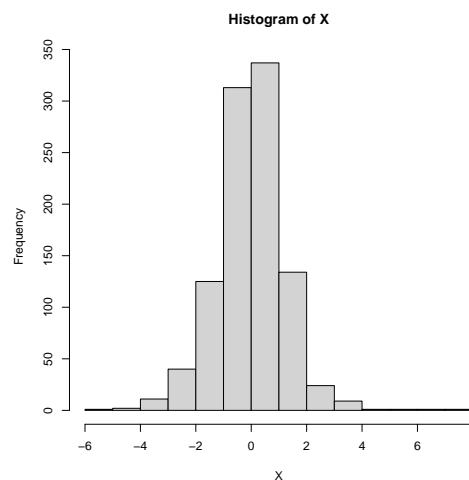
```
ppois(q = 1, lambda = 3)
## [1] 0.199
dpois(x = 0, lambda = 3) + dpois(x = 1, lambda = 3) ##Alternative
## [1] 0.199
```

- 0.5-Quantil von X , wenn $X \sim \text{Bin}(\pi = 0.5, n = 100)$

```
qbinom(p = 0.5, size = 100, prob = 0.5)
## [1] 50
```

- $n = 1000$ Zufallszahlen aus t -Verteilung mit 6 Freiheitsgraden.

```
X <- rt(n = 1000, df = 6)
hist(X)
```



3.6 Erwartungswerte

Grundidee. Wir möchten für eine Zufallsvariable X gewisse *Kennzahlen* finden, die in geeigneter Form das *durchschnittliche Verhalten* von X beschreiben.

3.6.1 Erwartungswert

Definition. Sei X eine diskrete Zufallsvariable mit Werten x_1, \dots, x_k mit Masse $p(x)$. Der *Erwartungswert* von X ist

$$E(X) = \sum_{i=1}^k x_i p(x_i). \quad (3.6.1)$$

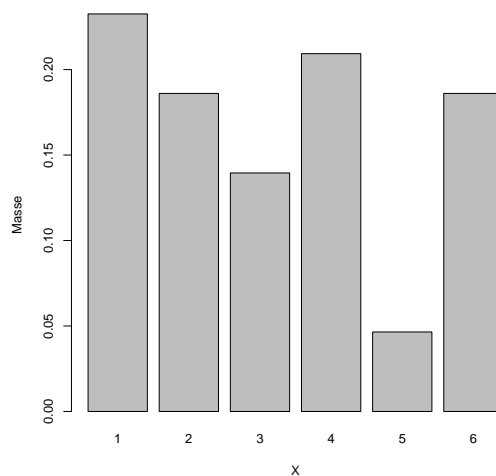
Wenn X kontinuierlich ist, dann ersetzen wir die Summe \sum durch ein Integral \int (und

die Masse durch Dichte.)

$$E(X) = \int xp(x)dx. \quad (3.6.2)$$

Beispiel. Eine diskrete Zufallsvariable und ihre Verteilung sei gegeben:

```
## X      p
## 1 0.23256
## 2 0.18605
## 3 0.13953
## 4 0.20930
## 5 0.04651
## 6 0.18605
```



Der Erwartungswert ist $1 \cdot 0.233 + 2 \cdot 0.186 + 3 \cdot 0.14 + 4 \cdot 0.209 + 5 \cdot 0.047 + 6 \cdot 0.186 = 3.209$.

3.6.2 Varianz und Standardabweichung

Definition. Sei X eine Zufallsvariable, dann heit

$$\text{Var}(X) = E[(X - E(X))^2] \quad (3.6.3)$$

die *Varianz* von X , und $\sqrt{\text{Var}(X)}$ heit *Standardabweichung* von X . Man schreibt auch $\text{sd}(X) = \sigma = \sqrt{\text{Var}(X)}$. Sowohl die Varianz als auch die Standardabweichung sind Kennzahlen fr die *Streuung* der Verteilung von X . In Worten ist die Varianz ein *erwarteter quadrierter Abstand vom Erwartungswert* oder – leicht weniger przise – ein “mittlerer quadrierter Abstand”.

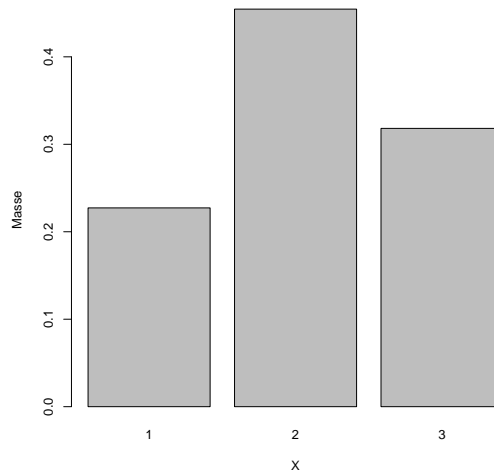
Man kann zeigen, dass

$$\text{Var}(X) = E(X^2) - E(X)^2. \quad (3.6.4)$$

Mit dieser Regel kann oft einfacher gerechnet werden.

Beispiel. Eine diskrete Zufallsvariable und ihre Verteilung sei gegeben:

```
## X      p
## 1 0.227
## 2 0.455
## 3 0.318
```



Der Erwartungswert ist $1 \cdot 0.227 + 2 \cdot 0.455 + 3 \cdot 0.318 = 2.091$. Die Varianz ist $\text{Var}(X) = E(X^2) - E(X)^2 = 1 \cdot 0.227 + 4 \cdot 0.455 + 9 \cdot 0.318 - 2.091^2 = 0.537$.

Varianz einer linearen Funktion einer Zufallsvariable. Wenn $Y = a + bX$ eine lineare Funktion der Zufallsvariable X ist, dann ist

$$\text{Var}(Y) = \text{Var}(a + bX) = b^2 \text{Var}(X) \quad (3.6.5)$$

Beachte das b^2 !

Beispiel Normalverteilung: $X \sim \mathcal{N}(0, 1)$ sei standardnormalverteilt mit Varianz 1. Dann ist $Y = 10 + 12 \times X$ normalverteilt mit Varianz 144, $Y \sim \mathcal{N}(10, 144)$.

```
X <- rnorm(1000)
var(X)

## [1] 0.997

Y <- 10 + 12 * X
var(Y)

## [1] 144
```

3.6.3 Erwartungswerte von wichtigen parametrischen Verteilungen.

Ohne Beweise geben wir in der Tabelle 3.2 Erwartungswert und Varianz an für wichtige parametrische Verteilungen.

| Verteilung | Notation | Erwartungswert | Varianz |
|---------------------------|-------------------------------------|----------------|-------------------------------------|
| Bernoulli-Verteilung | $X \sim \text{Bin}(\pi, 1)$ | π | $\pi(1 - \pi)$ |
| Binomial-Verteilung: | $X \sim \text{Bin}(\pi, n)$ | πn | $n\pi(1 - \pi)$ |
| Poisson-Verteilung: | $X \sim \text{Pois}(\lambda)$ | λ | λ |
| Normalverteilung: | $X \sim \mathcal{N}(\mu, \sigma^2)$ | μ | σ^2 |
| Standardnormalverteilung: | $X \sim \mathcal{N}(0, 1)$ | 0 | 1 |
| t_n -Verteilung: | $X \sim t_n$ | 0 | $\frac{n}{n-2}$ |
| χ_n^2 -Verteilung: | $X \sim \chi_n^2$ | n | $2n$ |
| $F_{n,m}$ -Verteilung: | $X \sim F_{n,m}$ | $m/(m-2)$ | $\frac{2m^2(n+m-2)}{n(m-2)^2(m-4)}$ |

Tabelle 3.2: Erwartungswert und Varianz von wichtigen Verteilungen

Beispiel. Eine Zufallsvariable sei t -verteilt mit 14 Freiheitsgraden, $X \sim t_{14}$. Wir simulieren eine grosse Zufallsstichprobe aus dieser Verteilung und berechnen den (empirischen) Durchschnitt und die (empirische) Varianz⁵. Ersterer sollte nahe bei dem Erwartungswert 0, letzterer nahe bei der Varianz $14/12 = 1.167$ sein.

```
X <- rt(10000, df = 14)
mean(X)

## [1] 0.00216

var(X)

## [1] 1.17
```

⁵Mehr dazu in der beschreibenden Statistik.

3.7 Gemeinsame Verteilungen

Grundidee. Wir möchten das gemeinsame Verhalten von und die Zusammenhänge zwischen mehreren Zufallsvariablen beschreiben und untersuchen. Das erlaubt uns, den Begriff der Unabhängigkeit von Ereignissen auf Zufallsvariablen zu verallgemeinern.

Dazu schauen wir uns direkt eine sogenannte *bivariate* Verteilung an.

Bivariate diskrete Verteilung. Eine gemeinsame diskrete Verteilung von X und Y wäre

```
##      Y=0   Y=1   Y=2   Y=3   Sum
## X=1 0.125 0.250 0.125 0.000 0.500
## X=2 0.000 0.125 0.250 0.125 0.500
## Sum 0.125 0.375 0.375 0.125 1.000
```

Hier könnte z.B. X für “Interventionsart” stehen und Y für eine vierwertige Variable auf dem Konstrukt “Zufriedenheit”.

Gemeinsame Verteilung. Die 2×4 Wahrscheinlichkeiten $p(x, y) = \Pr(X = x, Y = y)$ stellen die *gemeinsame Verteilung* dar. So ist $p(1, 2) = 0.125$.

Randverteilungen. $p(x)$ stellt die *Randverteilung* von X dar, $p(y)$ stellt die Randverteilung von Y dar

- $p_X(1) = 0.5, p_X(2) = 0.5$.
- $p_Y(0) = 0.125, p_Y(1) = 0.375, p_Y(2) = 0.375, p_Y(3) = 0.125$.

Bedingte Verteilung. Die Verteilung $p(x | y)$ heisst *bedingte Verteilung* von X gegeben $Y = y$. Diese ist

$$p(x | y) = \frac{p(x, y)}{p(y)} \quad (3.7.1)$$

So ist z.B. für $Y = 2$: $p(1 | 2) = 0.125/0.375 = 0.333, p(2 | 2) = 0.25/0.375 = 0.667$.

Unabhängigkeit. X und Y sind genau dann unabhängig, wenn für alle y gilt

$$p(x | y) = p(x). \quad (3.7.2)$$

So sind A und B im folgenden Beispiel – im Gegensatz zu obigem Beispiel – unabhängig.

```
##      B=1   B=2   Sum
## A=1 0.25 0.25 0.50
## A=2 0.25 0.25 0.50
## Sum 0.50 0.50 1.00
```

Gemeinsame Normalverteilung. Eine grundlegende mehrdimensionale Verteilung ist die *zweidimensionale Normalverteilung*. Diese hat zusätzlich zu μ_x, μ_y, σ_x^2 und σ_y^2 noch einen weiteren Parameter ρ (eine Zahl zwischen -1 und 1). Man kann zeigen, dass ρ gerade die *Korrelation* zwischen X und Y darstellt, die wir unten einführen. Die Abbildung 3.16 zeigt die gemeinsame Verteilung von $X \sim \mathcal{N}(10, 9)$ und $Y \sim \mathcal{N}(10, 16)$ mit $\rho = 0$ (Links), $\rho = 0.5$ (Mitte) und $\rho = 0.8$ (Rechts). Die Höhe repräsentiert die gemeinsame Dichte $p(x, y)$. Das Volumen unter der Dichtefunktion ist gleich Eins.

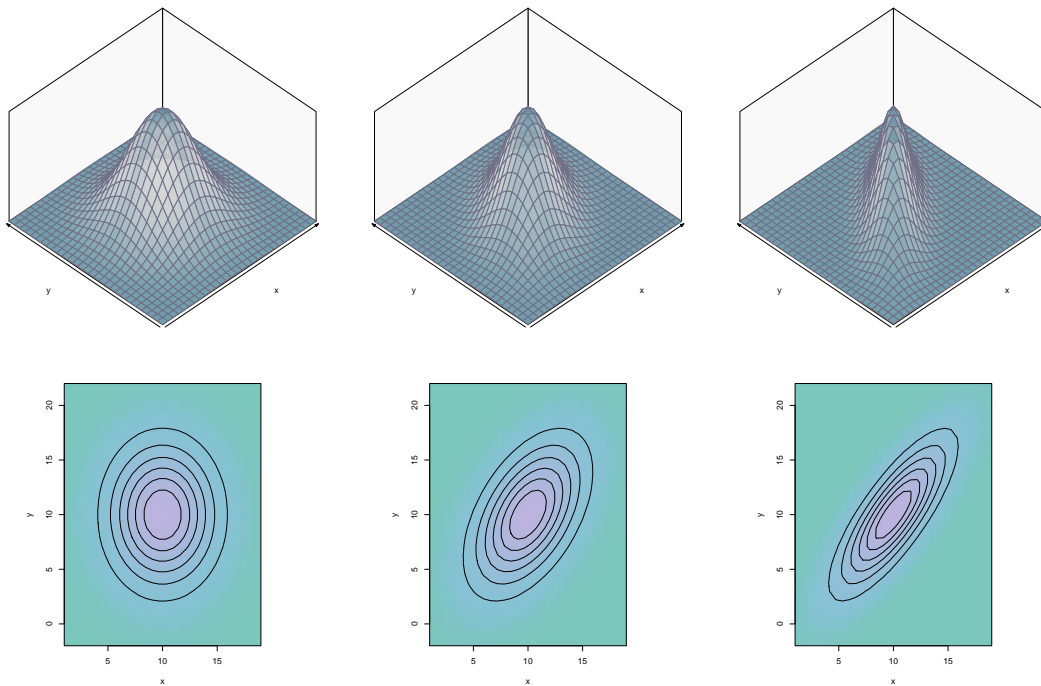


Abbildung 3.16: Dichte von gemeinsamer Normalverteilung

3.8 Kovarianz und Korrelation

Definition. Seien X und Y Zufallsvariablen. Dann heisst

$$\begin{aligned}\text{Cov}(X, Y) &= E[(X - E(X))(Y - E(Y))] \\ &= E[(X - \mu_X)(Y - \mu_Y)]\end{aligned}\quad (3.8.1)$$

die *Kovarianz* von X und Y . Die Kovarianz misst, ob – im Schnitt – die Abweichungen der Variable $X - \mu_X$ “zusammen gehen” mit den Abweichungen der Variable $Y - \mu_Y$. Dann ist

$$\rho_{X,Y} = \text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y} \quad (3.8.2)$$

die *Korrelation* von X und Y . Die Korrelation zwischen X und Y ist ein standardisiertes Mass für die Stärke des *linearen Zusammenhangs* zwischen den beiden Zufallsvariablen. Ist $\rho = 0$, so heissen X und Y unkorreliert.

- Korrelationen sind Zahlen zwischen -1 und 1, $|\text{Corr}(X, Y)| \leq 1$.
- $|\text{Corr}(X, Y)| = 1$ genau dann, wenn $Y = a + bX$ ist, mit $b \neq 0$ (Wenn also Y eine lineare Funktion von X ist⁶). Mann nennt dann X und Y perfekt korreliert.
- Eine Kovarianz ist also eine mehrdimensionale Verallgemeinerung der Varianz. Die Kovarianz einer Variablen mit sich selber ist gerade die Varianz.

Wichtig. Unabhängige Zufallsvariablen X und Y sind immer unkorreliert, das Umgekehrte gilt aber in der Regel nicht. Wenn aber X und Y gemeinsam normalverteilt sind, dann impliziert Unkorreliertheit auch Unabhängigkeit.

Wie bei der Varianz (3.6.4) gibt es bei der Kovarianz eine Rechenregel. Man kann zeigen, dass (bei diskreten Zufallsvariablen)

$$\text{Cov}(X, Y) = E(XY) - E(X)E(Y) = \sum_i \sum_j x_i y_j p_{ij} - \sum_i x_i p_i \times \sum_j y_j p_j. \quad (3.8.3)$$

Beispiel. Kommen wir zurück auf obiges Beispiel der diskreten bivariaten Verteilung.

| ## | Y=0 | Y=1 | Y=2 | Y=3 | Sum |
|--------|-------|-------|-------|-------|-------|
| ## X=1 | 0.125 | 0.250 | 0.125 | 0.000 | 0.500 |
| ## X=2 | 0.000 | 0.125 | 0.250 | 0.125 | 0.500 |
| ## Sum | 0.125 | 0.375 | 0.375 | 0.125 | 1.000 |

Die Kovarianz zwischen X und Y ist dann gemäss (3.8.3): $\text{Cov}(X, Y) = 0.125 \cdot 1 \cdot 0 + 0 \cdot 2 \cdot 0 + \dots + 0.125 \cdot 2 \cdot 3 - (0.5 \cdot 1 + 0.5 \cdot 2)(0.125 \times 0 + 0.375 \times 1 + \dots + 0.125 \cdot 3) = 0.25$. Aus den Randverteilungen kommen wir auf $\sigma_X = 0.5$ und $\sigma_Y = 0.866$, und $\rho_{X,Y} = 0.577$.

⁶Aus der Schule kennen wir vielleicht noch die einfache lineare Funktion, die durch $Y = a + bX$ gegebene Gerade.

3.8.1 Rechenregeln

Für Zufallsvariable X, Y, Z und Zahlen $a, b, c, d \in \mathbb{R}$ gilt:

1. $\text{Var}(a + bX) = b^2 \text{Var}(X)$
2. $\text{Cov}(X, X) = \text{Var}(X)$
3. $\text{Cov}(X, Y + Z) = \text{Cov}(X, Y) + \text{Cov}(X, Z)$
4. $\text{Cov}(X, a) = 0$
5. $\text{Cov}(X, bY) = b \text{Cov}(X, Y)$
6. $\text{Cov}(X, Y) = \text{Cov}(Y, X)$
7. $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2 \text{Cov}(X, Y)$
8. $\text{Cov}(a + \sum_i b_i X_i, c + \sum_j d_j Y_j) = \sum_i \sum_j b_i d_j \text{Cov}(X_i, Y_j)$
9. $\text{Var}(a + \sum_i b_i X_i) = \sum_{i,j} b_i b_j \text{Cov}(X_i, X_j)$
10. Sind X und Y *unabhängig*, dann ist $\text{Cov}(X, Y) = 0$
11. Sind X_1, \dots, X_n *unabhängig*, dann $\text{Var}(a + \sum_i b_i X_i) = \sum_i b_i^2 \text{Var}(X_i)$

Zum Schluss zwei wichtige Beispiele:

Beispiel 1. Varianz der Summe und vom Durchschnitt von unabhängigen Variablen X_1, X_2, X_3 mit $\text{Var}(X_1) = 10$, $\text{Var}(X_2) = 30$ und $\text{Var}(X_3) = 50$ sowie $\text{Cov}(X_i, X_j) = 0$ für $i \neq j$. Dann ist

$$\text{Var}(X_1 + X_2 + X_3) = \text{Var}(X_1) + \text{Var}(X_2) + \text{Var}(X_3) = 90.$$

Die Varianz vom *Durchschnitt* ist dann

$$\text{Var}\left(\frac{X_1 + X_2 + X_3}{3}\right) = \frac{1}{9}(\text{Var}(X_1) + \text{Var}(X_2) + \text{Var}(X_3)) = \frac{1}{9}(10 + 30 + 50) = 10$$

Beispiel 2. Wir möchten die Varianz vom Durchschnitt von n i.i.d. Zufallsvariablen (“independent and identically distributed”). Wenn X_1, X_2, \dots, X_n i.i.d., dann ist

$$\boxed{\text{Var}\left(\frac{X_1 + X_2 + \dots + X_n}{n}\right) = \frac{n \text{Var}(X)}{n^2} = \frac{1}{n} \text{Var}(X)} \quad (3.8.4)$$

Dieses Resultat werden wir sehr oft brauchen, insbesondere im Zusammenhang mit dem *Standardfehler* einer Schätzung. Es besagt, dass Durchschnitte weniger variieren als Einzelmessungen und dass man deren Varianz beliebig klein machen kann (wegmitteln), wenn n gross wird. Das entspricht auch unserer Alltagserfahrung.

| z | 0 | 0.01 | 0.02 | 0.03 | 0.04 | 0.05 | 0.06 | 0.07 | 0.08 | 0.09 |
|-----|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 0 | 0.500 | 0.504 | 0.508 | 0.512 | 0.516 | 0.520 | 0.524 | 0.528 | 0.532 | 0.536 |
| 0.1 | 0.540 | 0.544 | 0.548 | 0.552 | 0.556 | 0.560 | 0.564 | 0.567 | 0.571 | 0.575 |
| 0.2 | 0.579 | 0.583 | 0.587 | 0.591 | 0.595 | 0.599 | 0.603 | 0.606 | 0.610 | 0.614 |
| 0.3 | 0.618 | 0.622 | 0.626 | 0.629 | 0.633 | 0.637 | 0.641 | 0.644 | 0.648 | 0.652 |
| 0.4 | 0.655 | 0.659 | 0.663 | 0.666 | 0.670 | 0.674 | 0.677 | 0.681 | 0.684 | 0.688 |
| 0.5 | 0.691 | 0.695 | 0.698 | 0.702 | 0.705 | 0.709 | 0.712 | 0.716 | 0.719 | 0.722 |
| 0.6 | 0.726 | 0.729 | 0.732 | 0.736 | 0.739 | 0.742 | 0.745 | 0.749 | 0.752 | 0.755 |
| 0.7 | 0.758 | 0.761 | 0.764 | 0.767 | 0.770 | 0.773 | 0.776 | 0.779 | 0.782 | 0.785 |
| 0.8 | 0.788 | 0.791 | 0.794 | 0.797 | 0.800 | 0.802 | 0.805 | 0.808 | 0.811 | 0.813 |
| 0.9 | 0.816 | 0.819 | 0.821 | 0.824 | 0.826 | 0.829 | 0.831 | 0.834 | 0.836 | 0.839 |
| 1 | 0.841 | 0.844 | 0.846 | 0.848 | 0.851 | 0.853 | 0.855 | 0.858 | 0.860 | 0.862 |
| 1.1 | 0.864 | 0.867 | 0.869 | 0.871 | 0.873 | 0.875 | 0.877 | 0.879 | 0.881 | 0.883 |
| 1.2 | 0.885 | 0.887 | 0.889 | 0.891 | 0.893 | 0.894 | 0.896 | 0.898 | 0.900 | 0.901 |
| 1.3 | 0.903 | 0.905 | 0.907 | 0.908 | 0.910 | 0.911 | 0.913 | 0.915 | 0.916 | 0.918 |
| 1.4 | 0.919 | 0.921 | 0.922 | 0.924 | 0.925 | 0.926 | 0.928 | 0.929 | 0.931 | 0.932 |
| 1.5 | 0.933 | 0.934 | 0.936 | 0.937 | 0.938 | 0.939 | 0.941 | 0.942 | 0.943 | 0.944 |
| 1.6 | 0.945 | 0.946 | 0.947 | 0.948 | 0.949 | 0.951 | 0.952 | 0.953 | 0.954 | 0.954 |
| 1.7 | 0.955 | 0.956 | 0.957 | 0.958 | 0.959 | 0.960 | 0.961 | 0.962 | 0.962 | 0.963 |
| 1.8 | 0.964 | 0.965 | 0.966 | 0.966 | 0.967 | 0.968 | 0.969 | 0.969 | 0.970 | 0.971 |
| 1.9 | 0.971 | 0.972 | 0.973 | 0.973 | 0.974 | 0.974 | 0.975 | 0.976 | 0.976 | 0.977 |
| 2 | 0.977 | 0.978 | 0.978 | 0.979 | 0.979 | 0.980 | 0.980 | 0.981 | 0.981 | 0.982 |
| 2.1 | 0.982 | 0.983 | 0.983 | 0.983 | 0.984 | 0.984 | 0.985 | 0.985 | 0.985 | 0.986 |
| 2.2 | 0.986 | 0.986 | 0.987 | 0.987 | 0.987 | 0.988 | 0.988 | 0.988 | 0.989 | 0.989 |
| 2.3 | 0.989 | 0.990 | 0.990 | 0.990 | 0.990 | 0.991 | 0.991 | 0.991 | 0.991 | 0.992 |
| 2.4 | 0.992 | 0.992 | 0.992 | 0.992 | 0.993 | 0.993 | 0.993 | 0.993 | 0.993 | 0.994 |
| 2.5 | 0.994 | 0.994 | 0.994 | 0.994 | 0.994 | 0.995 | 0.995 | 0.995 | 0.995 | 0.995 |
| 2.6 | 0.995 | 0.995 | 0.996 | 0.996 | 0.996 | 0.996 | 0.996 | 0.996 | 0.996 | 0.996 |
| 2.7 | 0.997 | 0.997 | 0.997 | 0.997 | 0.997 | 0.997 | 0.997 | 0.997 | 0.997 | 0.997 |
| 2.8 | 0.997 | 0.998 | 0.998 | 0.998 | 0.998 | 0.998 | 0.998 | 0.998 | 0.998 | 0.998 |
| 2.9 | 0.998 | 0.998 | 0.998 | 0.998 | 0.998 | 0.998 | 0.998 | 0.999 | 0.999 | 0.999 |
| 3 | 0.999 | 0.999 | 0.999 | 0.999 | 0.999 | 0.999 | 0.999 | 0.999 | 0.999 | 0.999 |

Tabelle 3.1: Standardnormalverteilung, Verteilungsfunktion von z .

Literaturverzeichnis

- [1] A. Agresti. *Categorical Data Analysis*. Wiley Series in Probability and Statistics. Wiley-Interscience, 2nd edition, 2002.
- [2] T. Bayes. An essay towards solving a problem in the doctrine of chances. *Phil. Trans. of the Royal Soc. of London*, 53:370–418, 1763.
- [3] J. Bortz. *Statistik für Human- und Sozialwissenschaftler*. Springer, Berlin; Heidelberg; New York, 2005.
- [4] J. Bortz and N. Döring. *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler*. Springer, Heidelberg, 2006.
- [5] R. Carnap. *Logical Foundations of Probability*. University of Chicago Press, 1962.
- [6] Ronald Christensen. Testing Fisher, Neyman, Pearson, and Bayes. *The American Statistician*, 59(2):121–126, 2005.
- [7] C.S. Davis. *Statistical Methods for the Analysis of Repeated Measurements*. Springer Texts in Statistics. Springer, 2002.
- [8] B. De Finetti. *Funzione caratteristica di un fenomeno aleatorio*. Mem. della R. accad. naz. dei Lincei, classe di sci. fis., vi, 4. Società anonima tipografica, 1930.
- [9] H.C.W. de Vet, C.B. Terwee, L.B. Mokkink, and D.L. Knol. *Measurement in Medicine: A Practical Guide*. Practical Guides to Biostatistics and Epidemiology. Cambridge University Press, 2011.
- [10] PJ Diggle, P. Heagerty, KY Liang, and SL Zeger. *Analysis of Longitudinal Data*. Oxford University Press, second edition, 2002.
- [11] L. Fahrmeir, R. Künstler, I. Pigeot, and G. Tutz. *Statistik, Der Weg zur Datenanalyse*. Springer-Lehrbuch. Springer, Heidelberg, 6rd edition, 2007.
- [12] R.A. Fisher. *Statistical methods for research workers*. Edinburgh Oliver & Boyd, 1925.
- [13] A. Gelman, J.B. Carlin, H.S. Stern, D.B. Dunson, A. Vehtari, and D.B. Rubin. *Bayesian Data Analysis, Third Edition*. Chapman & Hall/CRC Texts in Statistical Science. Taylor & Francis, 2013.
- [14] S. Greenland and K.J. Rothman. *Modern Epidemiology*. Lippincott-Raven, 1998.

- [15] G. Grimmett and D. Stirzaker. *Probability and Random Processes*. Oxford University Press, 2001.
- [16] L. Held. *Methoden der statistischen Inferenz: Likelihood und Bayes*. Springer, Heidelberg, 2008.
- [17] H. Jeffreys. *Theory of Probability*. Oxford, Oxford, England, third edition, 1961.
- [18] Donald E. Knuth. Literate programming. *Comput. J.*, 27(2):97–111, May 1984.
- [19] John K. Kruschke. *Doing Bayesian Data Analysis: A Tutorial with R and BUGS*. Academic Press, 1st edition, 2010.
- [20] John K. Kruschke and Mike Meredith. *BEST: Bayesian Estimation Supersedes the t-Test*, 2015. R package version 0.3.0.
- [21] Paul E. Meehl. Theory-Testing in Psychology and Physics: A Methodological Paradox. *Philosophy of Science*, 34(2):103–115, 1967.
- [22] Paul E Meehl. Theoretical risks and tabular asterisks: Sir karl, sir ronald, and the slow progress of soft psychology. *Journal of consulting and clinical Psychology*, 46(4):806, 1978.
- [23] A. Meichtry. *Statistik: Handbuch für Therapeuten*. Thieme, 2017.
- [24] A. Meichtry, J. Kool, A. Schämamm, and R. Hilfiker. Therapieeffekte: Beurteilung der empirischen Evidenz. *Physioscience*, 4:184–193, 2008.
- [25] S. Morkved, K. Bo, B. Schei, and K. A. Salvesen. Pelvic floor muscle training during pregnancy to prevent urinary incontinence: a single-blind randomized controlled trial. *Obstet Gynecol*, 101(2):313–319, Feb 2003.
- [26] J. Neyman and E. S. Pearson. On the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society of London*, 231:289–337, 1933.
- [27] J. Pearl. *Causality: Models, Reasoning and Inference*. Cambridge Univ Press, 2000.
- [28] C. S. Peirce. *Collected Papers*. Harvard University Press, Cambridge, 1931–1935.
- [29] K.R. Popper. *The Logic of Scientific Discovery*. Routledge Classics. Routledge, 2002.
- [30] K.R. Popper and T.E. Hansen. *Die beiden Grundprobleme der Erkenntnistheorie*. Die Einheit der Gesellschaftswissenschaften. Mohr, 1979.
- [31] L.G. Portney and M.P. Watkins. *Foundations of Clinical Research: Applications to Practice*. Pearson/Prentice Hall, 2009.

- [32] S. Senn. *Dicing with Death: Chance, Risk and Health*. Cambridge University Press, 2003.
- [33] David J Spiegelhalter, Keith R Abrams, and Jonathan P Myles. *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*. Wiley, Chichester, 2004.
- [34] W. Stegmüller. *Jenseits von Popper und Carnap*. Personelle und Statistische Wahrscheinlichkeit. Springer, 1973.
- [35] D.L. Streiner, G.R. Norman, and J. Cairney. *Health Measurement Scales: A practical guide to their development and use*. OUP Oxford, 2014.
- [36] Stephen T. Ziliak and Deirdre N. McCloskey. *The Cult of Statistical Significance*. University of Michigan Press, 1st edition, 2 2008.