

Induktive Statistik

Prof. Dr. Christoph Hanck

Universität Duisburg-Essen

Sommersemester 2025

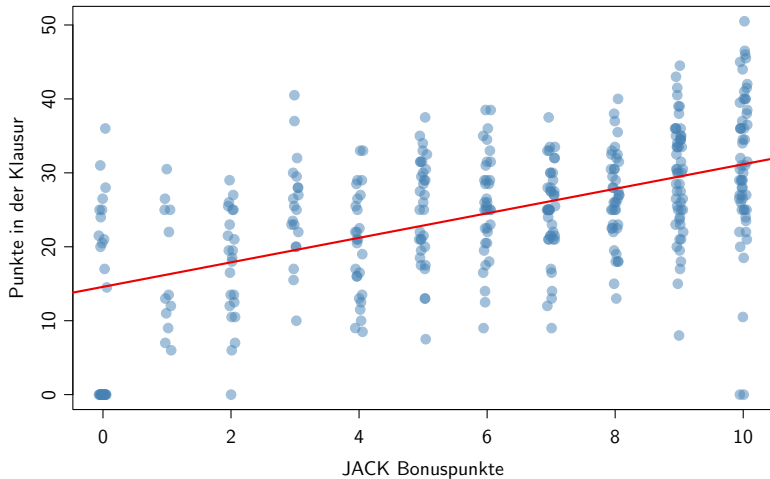


- 1 Einleitung
- 2 Grundlagen
- 3 Eindimensionale Zufallsvariablen
- 4 Ausgewählte theoretische Verteilungen
- 5 Grundzüge der Stichprobentheorie
- 6 Statistische Schätzverfahren
- 7 Statistische Testverfahren
- 8 Zweidimensionale Zufallsvariablen

- Ziel der **deskriptiven** (beschreibenden) **Statistik**: das Datenmaterial mit Hilfe tabellarischer und grafischer Repräsentationen sowie geeigneter Kenngrößen übersichtlich aufbereiten.
- Ziel der **induktiven** (schließenden) **Statistik**: Verifikation theoretischer Modelle anhand von Daten und Testen von Hypothesen über unbekannte Parameter auf Basis von Wahrscheinlichkeitsmodellen.

Motivation

Beispiel: Lernerfolg



Übergang von der deskriptiven zur induktiven Statistik:

- **Induktives Schließen**, statistische Inferenz, Repräsentationsschluss: Schluss von einer Teilgesamtheit auf die Grundgesamtheit
- **Induktive Statistik**: Methoden, die es erlauben von den Beobachtungen einer Teilgesamtheit (= *Stichprobe*) auf bestimmte Charakteristika der dazugehörenden *Grundgesamtheit* zu schließen.

- Beim Glücksspiel meinen Sie, dass die „Gegenseite“ mit manipulierten Würfeln spielt, d.h. die Zahl 6 fällt häufiger als andere Zahlen. Wie können Sie Ihre Vermutung überprüfen?
- Vor der Bundestagswahl prognostiziert ein Wahlforschungsinstitut, dass 30% der Stimmen auf die CDU/CSU entfallen. Wie kann eine solche Prognose verlässlich durchgeführt werden?

Ziel: Schlussfolgerungen von einer Stichprobe auf die Grundgesamtheit

Schokoladenwürfel





Quelle: <https://blog.ritter-sport.de/2014/08/05/schatzt-euch-bunt/>

Man beachte: Wahrscheinlichkeitsrechnung

- ist **mehr** als Grundlage der schließenden Statistik
- hat enorme eigenständige ökonomische Bedeutung z.B. in
 - ▶ Mikroökonomik
 - ▶ Investition und Finanzierung
 - ▶ Portfoliotheorie

1. Einleitung
2. Grundlagen
3. Eindimensionale Zufallsvariablen
4. Ausgewählte theoretische Verteilungen
5. Grundzüge der Stichprobentheorie
6. Statistische Schätzverfahren
7. Statistische Testverfahren
8. Zweidimensionale Zufallsvariablen

- 1 Einleitung
- 2 Grundlagen
- 3 Eindimensionale Zufallsvariablen
- 4 Ausgewählte theoretische Verteilungen
- 5 Grundzüge der Stichprobentheorie
- 6 Statistische Schätzverfahren
- 7 Statistische Testverfahren
- 8 Zweidimensionale Zufallsvariablen

Zufallsvorgänge (bzw. stochastische Vorgänge) sind durch zwei wesentliche Eigenschaften gekennzeichnet:

1. Sie besitzen verschiedene, sich gegenseitig ausschließende Ausgänge, die bereits vor Beginn des Vorgangs bekannt sind;
2. es ist nicht vorhersehbar, welcher Ausgang tatsächlich eintreten wird.

Beispiele für Zufallsvorgänge:

- Der Ausgang eines Fußballspiels, der Kurs einer Aktie am nächsten Tag oder die realisierte Augenzahl eines Würfelwurfes.

- Ist ein Zufallsvorgang unverändert beliebig oft wiederholbar, liegt ein **Zufallsexperiment** vor. Die unveränderte Wiederholbarkeit beschreibt man auch als unter gleichen (Rand-)Bedingungen wiederholbar.
- D.h., dass die Randbedingungen wie bei naturwissenschaftlichen Experimenten kontrolliert werden können. Dies stellt sicher, dass die Bedingungen, unter denen das Experiment stattgefunden hat, auch bei weiteren Durchführungen hätten eingehalten werden können.
- Damit gehören alle Zufallsvorgänge, die fiktiv unter gleichen Bedingungen wiederholbar sind, zu den Zufallsexperimenten. Dies erlaubt es, auch Zufallsvorgänge als Zufallsexperimente aufzufassen, deren praktische Wiederholung unter gleichen Bedingungen schwierig wäre.

- Alle Ausgänge ω_i eines Zufallsvorganges bzw. -experimentes fasst man zu einem **Stichprobenraum** (Ergebnis- bzw. Ereignisraum) Ω zusammen. Der Stichprobenraum ist eine Menge, deren Elemente die Ausgänge sind.
- Ω kann endlich oder unendlich viele Ausgänge enthalten. Lassen sich die unendlich vielen Ausgänge mit den natürlichen Zahlen \mathbb{N} abzählen, bezeichnet man Ω als **abzählbar unendlich**. Gelingt dies nicht, heißt Ω **überabzählbar unendlich**.

$$\Omega = \{\omega_1, \dots, \omega_m, \dots\}.$$

Beispiel 2.1: Würfelwurf.

Der Zufallsvorgang *Werfen eines Würfels* hat sechs mögliche Ausgänge.

Ω ist endlich und lässt sich schreiben als:

$$\Omega = \{\omega_1, \omega_2, \dots, \omega_6\} = \{\omega_i, i = 1, \dots, 6\} = \{1, 2, 3, 4, 5, 6\}.$$

Beispiel 2.2: Münzwurf.

Wirft man eine Münze mit Seiten *Zahl* (Z) und *Kopf* (K) bis zum ersten Mal Z erscheint, lauten die möglichen Ausgänge:

$$\begin{aligned}\omega_1 &= Z && \text{(zum ersten Mal } \textit{Zahl} \text{ im ersten Wurf),} \\ \omega_2 &= KZ && \text{(zum ersten Mal } \textit{Zahl} \text{ im zweiten Wurf),} \\ &\vdots \\ \omega_m &= \underbrace{K \dots K}_{m-1} Z && \text{(zum ersten Mal } \textit{Zahl} \text{ im } m\text{-ten Wurf),} \\ &\vdots\end{aligned}$$

Der Stichprobenraum ist hier abzählbar unendlich:

$$\Omega = \{\omega_1, \dots, \omega_m, \dots\}.$$

Beispiel 2.3: Zugverspätung.

Die Verspätung eines Zuges in Minuten sei ein Zufallsvorgang mit Ausgängen im Intervall $[0; 10]$. Bei unendlicher Messgenauigkeit sind überabzählbar viele Verspätungen möglich, da das Intervall $[0; 10]$ Teilmenge der reellen Zahlen \mathbb{R} ist.

Die reellen Zahlen sind, wie auch jede ihrer Teilmengen, mächtiger als \mathbb{N} und daher überabzählbar unendlich.

- Jede Teilmenge von Ω heißt **(Zufalls–)Ereignis**.
- Da eine Menge auch Teilmenge von sich selbst und die leere Menge \emptyset Teilmenge jeder Menge ist, sind Ω und \emptyset selbst Ereignisse von Ω .
- Ein Ereignis $A \subset \Omega$ tritt ein, wenn der Ausgang ω_i des Zufallsvorgangs Element von A ist: $\omega_i \in A$.
- Da ein Zufallsvorgang immer in einem Ausgang $\omega_i \in \Omega$ mündet, ist Ω auch das **sichere Ereignis**.
- Analog hierzu heißt die leere Menge \emptyset das **unmögliche Ereignis**, weil kein $\omega_i \in \Omega$ existiert, das Element der leeren Menge \emptyset ist: $\omega_i \notin \emptyset, i = 1, \dots, m$.

- Teilmengen $\{\omega_i\}$, deren einziges Element ein Ausgang $\omega_i \in \Omega$ ist, heißen **Elementarereignisse**.
- Umfassen Teilmengen mehrere Ausgänge, nennt man sie **zusammengesetzte Ereignisse**.
- Z. B. ist beim *Wurf eines Würfels* der Ausgang: *Augenzahl 3 liegt oben* ein Elementarereignis und wird geschrieben als $\{3\}$; das Ereignis A: *gerade Augenzahl liegt oben* ist ein zusammengesetztes Ereignis, als Menge geschrieben: $\{2, 4, 6\}$.
- A tritt ein, wenn der Würfelwurf die Augenzahl 2, 4 oder 6 ergibt.

- Die insgesamt möglichen Ereignisse eines Zufallsvorgangs findet man, indem alle Teilmengen für Ω gebildet werden.
- Die Zusammenfassung dieser Teilmengen führt bei endlichem oder abzählbar unendlichem Stichprobenraum Ω zur **Potenzmenge** $PM(\Omega)$.
- Die Anzahl der Ereignisse der Potenzmenge ist 2^m , wobei m der Anzahl der Elemente von Ω entspricht.
- Siehe Buch, S. 11.

Beispiel 2.4: Potenzmenge.

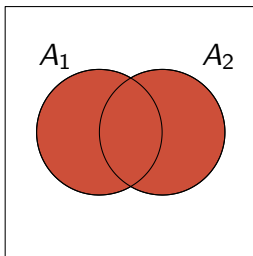
Ein Zufallsvorgang hat den Stichprobenraum $\Omega = \{1, 2, 3\}$; wegen $m = 3$ beträgt die Anzahl der möglichen Ereignisse

$$2^3 = 8.$$

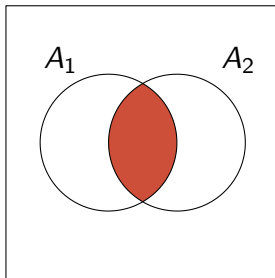
Diese Ereignisse lauten $A_1 = \emptyset$, $A_2 = \{1\}$, $A_3 = \{2\}$, $A_4 = \{3\}$, $A_5 = \{1, 2\}$, $A_6 = \{1, 3\}$, $A_7 = \{2, 3\}$, $A_8 = \{1, 2, 3\} = \Omega$. Die Potenzmenge ist

$$PM(\Omega) = \{A_1, \dots, A_8\}.$$

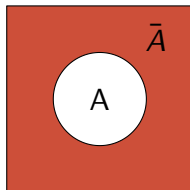
- Zwischen den Ereignissen können bestimmte Beziehungen vorliegen.
- **Beispiel:** Das Ereignis $A_5 = \{1, 2\}$ tritt dann ein, wenn entweder $A_2 = \{1\}$ oder $A_3 = \{2\}$ eintritt. A_5 heißt daher **Vereinigungsereignis**, geschrieben als $A_5 = A_2 \cup A_3$.
- Verallgemeinert erhält man Vereinigungsereignisse V als $V = \bigcup_{j=1}^n A_j$.
- Für $n = 2$ ist V als rote Fläche im **Venn-Diagramm** wiedergegeben, wobei das Rechteck den Stichprobenraum Ω festlegt.



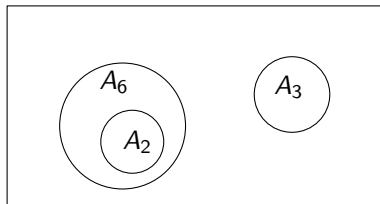
- Ist der Schnitt zweier Ereignisse A_i und A_j nicht leer, gilt $A_i \cap A_j = D \neq \emptyset$, so treten mit D auch die Ereignisse A_i und A_j ein.
- D heißt daher **Durchschnittsereignis**, das allgemein definiert ist als $D = \bigcap_{j=1}^n A_j$. D tritt ein, wenn alle A_j eintreten.



- Tritt A_i genau dann ein, wenn A_j nicht eintritt, so sind die beiden Ereignisse zueinander komplementär.
- A_i heißt **Komplementärereignis** oder kurz Komplement und lässt sich schreiben als $A_i = \bar{A}_j$. Natürlich ist auch A_j Komplementärereignis zu A_i : $A_j = \bar{A}_i$.
- Im Beispiel 2.4 ist das Ereignis $A_2 = \{1\}$ das Komplement zu $A_7 = \{2, 3\}$: $A_2 = \bar{A}_7$. Umgekehrt gilt auch $A_7 = \bar{A}_2$.
- Für A und sein Komplement \bar{A} gelten $A \cup \bar{A} = \Omega$ und $A \cap \bar{A} = \emptyset$.



- A_i und A_j heißen **disjunkt**, wenn ihr Schnitt leer ist: $A_i \cap A_j = \emptyset$.
- Komplementäre Ereignisse sind daher immer auch disjunkt, die Umkehrung gilt aber nicht.
- So sind im Beispiel 2.4 $A_2 = \{1\}$ und $A_3 = \{2\}$ zwar disjunkt, aber nicht komplementär. Denn wenn A_3 nicht eintritt, folgt nicht notwendigerweise das Eintreten von A_2 ; es könnte auch $A_4 = \{3\}$ eintreten.

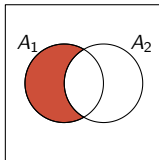


- A_i ist ein **Teilereignis** von A_j , wenn jeder Ausgang eines Zufallsvorgangs, der zu A_i gehört, auch in A_j liegt, A_j aber mindestens einen Ausgang ω_j enthält, der nicht auch in A_i enthalten ist.
- A_i ist eine **echte Teilmenge** von $A_j : A_i \subset A_j$.
- In der Abbildung repräsentieren die Kreise die Ereignisse A_2 , A_3 und A_6 des Beispiels 2.4.

- Das **Differenzereignis** $A_i \setminus A_j$ tritt dann ein, wenn der Ausgang des Zufallsvorgangs in A_i , aber nicht in A_j liegt.
- Man nennt $A_i \setminus A_j$ auch das **relative Komplement** zu A_j bezüglich A_i . Alternativ kann man schreiben $A_i \setminus A_j = A_i \cap \bar{A}_j$. Im Beispiel 2.4 folgt für $A_6 = \{1, 3\}$ und $A_5 = \{1, 2\}$ das Differenzereignis $A_6 \setminus A_5$ als

$$A_6 \setminus A_5 = A_6 \cap \bar{A}_5 = \{1, 3\} \cap \{3\} = \{3\}.$$

Nur $\omega_3 = 3$ führt dazu, dass $A_6 \setminus A_5$ eintritt. Das Beispiel zeigt, dass im Allgemeinen $A_i \setminus A_j \neq A_j \setminus A_i$.



Würfelwurf 1



```
x <- c(1,3,5)
```

```
y <- c(2,3,6)
```

```
union(x,y)
```

```
## [1] 1 3 5 2 6
```

```
intersect(x,y)
```

```
## [1] 3
```

```
setdiff(x,y)
```

```
## [1] 1 5
```

```
setdiff(y,x)
```

```
## [1] 2 6
```

```
setdiff(union(x,y),y) # Komplement von y, wenn Omega = union(x,y)
```

```
## [1] 1 5
```

```
rje::powerSet(x)
```

```
## [[1]]  
## numeric(0)  
##  
## [[2]]  
## [1] 1  
##  
## [[3]]  
## [1] 3  
##  
## [[4]]  
## [1] 1 3  
##  
## [[5]]  
## [1] 5  
##  
## [[6]]  
## [1] 1 5  
##  
## [[7]]  
## [1] 3 5  
##
```

Aufgabe 2.5:

Ein Würfel wird einmal gewürfelt. Es sind folgende Ereignisse definiert:

$$\begin{aligned} A_1 &= \{1, 2\}, A_2 = \{3, 4\}, A_3 = \{2, 4\}, A_4 = \{1, 2, 3, 4\}, \\ A_5 &= \{4\}, A_6 = \{3, 4, 5, 6\}, A_7 = \{5, 6\}, A_8 = \{1, 2, 4, 5\}, \\ A_9 &= \{2, 3, 6\}, A_{10} = \{1, 4, 5\}. \end{aligned}$$

Bestimmen Sie folgende Ereignisse und stellen Sie diese jeweils in einem Venn-Diagramm grafisch dar!

- | | | |
|--------------------|--|-----------------------|
| (a) $A_1 \cup A_2$ | (b) $A_2 \cap A_3$ | (c) \bar{A}_1 |
| (d) $A_1 \cap A_2$ | (e) $A_8 \cap \bar{A}_9 = A_8 \setminus A_9$ | (f) $A_5 \subset A_4$ |

- Jedes zusammengesetzte Ereignis A kann in disjunkte Teilereignisse $A_j \neq \emptyset$ so zerlegt werden, dass gilt: $A = \bigcup_{j=1}^n A_j$. A ergibt sich jetzt als disjunkte Vereinigung.
- In Beispiel 2.4 lässt sich A_8 in $A_2 = \{1\}$ und $A_7 = \{2, 3\}$, aber auch in $A_2 = \{1\}$, $A_3 = \{2\}$ und $A_4 = \{3\}$ zerlegen. Beim zweiten Fall wurde A_8 in Elementarereignisse zerlegt.
- Die Zerlegung von A in Elementarereignisse heißt **kanonische Darstellung**: Jedes Ereignis ergibt sich eindeutig als Vereinigungsereignis von Elementarereignissen:

$$A = \bigcup_{j=1}^n \{\omega_j\}.$$

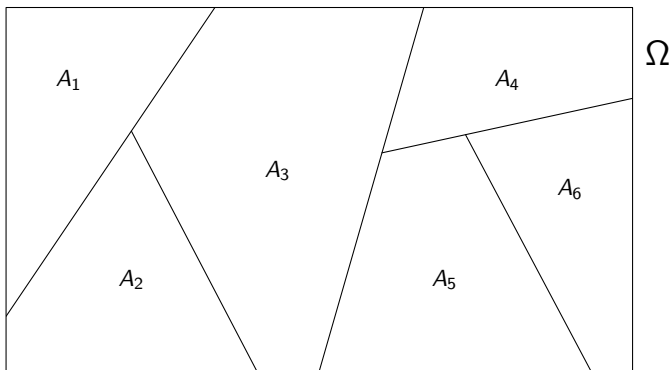
- Da auch Ω zu den zusammengesetzten Ereignissen gehört, kann es auch in (Teil-)Ereignisse A_j zerlegt werden.
- Die Menge $\{A_1, \dots, A_n\}$ bildet ein **vollständiges System von Ereignissen**, wenn für diese Zerlegung gilt:

$$(1) \quad \Omega = A_1 \cup \dots \cup A_n = \bigcup_{j=1}^n A_j,$$

$$(2) \quad A_i \cap A_j = \emptyset \quad \text{für } i \neq j,$$

$$(3) \quad A_j \neq \emptyset \quad \text{für } j = 1, \dots, n,$$

- Ein vollständiges System wird auch **vollständiges Ereignissystem** bzw. **vollständige Zerlegung** genannt.



Aufgabe 2.6:

Ein Würfel wird einmal gewürfelt. Es sind folgende Ereignisse definiert:

$$A_1 = \{2, 4, 6\}, A_2 = \{1, 3, 5\}, A_3 = \{2\}, \\ A_4 = \{4, 6\}, A_5 = \{1, 3\}, A_6 = \{5, 6\}.$$

Welche der folgenden Mengen bildet ein vollständiges Ereignissystem?

- (a) $F_1 = \{A_1, A_2\}$
- (b) $F_2 = \{A_1, A_5, A_6\}$
- (c) $F_3 = \{A_3, A_4, A_5\}$
- (d) $F_4 = \{A_2, A_3, A_4\}$

- Zufallsvorgänge zeichnen sich dadurch aus, dass ungewiss ist, welches ihrer möglichen Ereignisse eintritt.
- Wahrscheinlichkeiten quantifizieren die Chance des Eintretens eines bestimmten Ereignisses und werden mit P symbolisiert.
- Für ein Ereignis $A \subset \Omega$ gibt $P(A)$ jetzt die **Wahrscheinlichkeit** für das Eintreten von A an.
- P ist also eine Funktion, die Ereignissen reelle Zahlen zuordnet, welche wir als Wahrscheinlichkeit bezeichnen.

- Die axiomatische Grundlage für Wahrscheinlichkeiten wurde von **Kolmogoroff** entwickelt.
- Auf der Basis dieser Axiomatik lässt sich die Wahrscheinlichkeitsfunktion wie folgt definieren:

Kolmogoroff-Axiome

(K1) $P(A) \geq 0$ für alle A ,

(K2) $P(\Omega) = 1$,

(K3)

$$P\left(\bigcup_{j=1}^{\infty} A_j\right) = P(A_1) + P(A_2) + \dots = \sum_{j=1}^{\infty} P(A_j)$$

für alle A_i und A_j , die paarweise disjunkt sind: $A_i \cap A_j = \emptyset$, $i \neq j$.

Kolmogoroff-Axiome

- (K1) **Nichtnegativität**: Die Wahrscheinlichkeit P eines Ereignisses A ist stets nichtnegativ.
- (K2) **Normierung** der Wahrscheinlichkeit.
- (K3) **Volladditivität**: Die Wahrscheinlichkeit für die Vereinigung paarweise disjunkter Ereignisse gleicht der Summe der Einzelwahrscheinlichkeiten.

Jedes Ereignis $A \subset \Omega$ wird somit durch P in das geschlossene Intervall $[0, 1] \subset \mathbb{R}$ abgebildet.

Aus diesen Axiomen lassen sich folgende Rechenregeln ableiten:

Rechenregeln

$$(R1) \quad P(A) + P(\bar{A}) = 1,$$

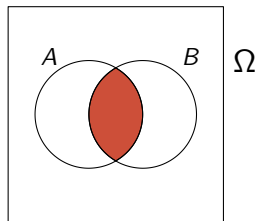
$$(R2) \quad P(A) \leq P(B) \text{ für } A \subset B,$$

$$(R3) \quad P(A_1 \cup A_2 \cup \dots \cup A_n) = \sum_{j=1}^n P(A_j) \text{ (für paarweise disjunkte Ereignisse),}$$

$$(R4) \quad P(A \cup B) = P(A) + P(B) - P(A \cap B) \text{ (für paarweise nicht disjunkte Ereignisse).}$$

Ihre Herleitung findet sich im Buch auf S. 23.

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



- A und B sind nicht disjunkt, da $A \cap B \neq \emptyset$. $A \cap B$ entspricht der roten Fläche.
- $(A \cup B)$ würde mit $P(A) + P(B)$ die Wahrscheinlichkeit für $A \cap B$ doppelt erfassen; folglich muss $P(A \cap B)$ subtrahiert werden.

- Der **Additionssatz für disjunkte Ereignisse**

$$P(A_1 \cup A_2 \cup \dots \cup A_n) = \sum_{j=1}^n P(A_j)$$

liefert eine Vorschrift für die Wahrscheinlichkeit eines Ereignisses $A = \{\omega_1, \omega_2, \dots, \omega_n\}$ eines abzählbar unendlichen Stichprobenraumes Ω .

- **Elementarereignisse** $\{\omega_j\}$, $j = 1, \dots, n$, sind paarweise disjunkt.


Aufgabe 2.7:

Es sind folgende Wahrscheinlichkeiten gegeben:

$$P(A) = 0.4, P(B) = 0.7, P(A \cap B) = 0.25.$$

Bestimmen Sie die Wahrscheinlichkeiten

$$P(A \cup B), P(\bar{B}), P(A \cap \bar{B}), P(A \cup \bar{B})$$

Wir verwenden einen Punkt anstelle eines Kommas als **Dezimaltrennzeichen**, da dies der angloamerikanischen Schreibweise der Programmiersprache  entspricht.

Aufgabe 2.8:

Ein Würfel wird ein Mal gewürfelt. Wie groß ist die Wahrscheinlichkeit,

- a) eine 2 oder eine 3 zu würfeln?
- b) eine 2, eine 3 oder eine 4 zu würfeln?
- c) eine ungerade Zahl oder eine gerade Zahl zu würfeln?
- d) eine Augenzahl kleiner als 4 oder eine gerade Augenzahl zu würfeln?
- e) eine Augenzahl kleiner als 4 und eine gerade Augenzahl zu würfeln?

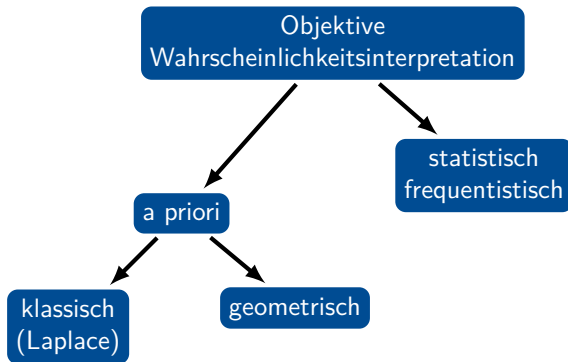
Aufgabe 2.9:

In einer Urne befinden sich 20 rote und 30 grüne Kugeln. 5 rote und 10 grüne Kugeln sind mit einer 1 beschriftet. Wie groß ist die Wahrscheinlichkeit, dass

- a) eine gezogene Kugel rot oder mit einer 1 beschriftet ist?
- b) eine gezogene Kugel nicht mit einer 1 beschriftet ist?
- c) eine rote Kugel gezogen wird und diese nicht mit einer 1 beschriftet ist?
- d) eine Kugel gezogen wird, die rot oder nicht mit einer 1 beschriftet ist?

- Mit den **Axiomen von Kolmogoroff** und den Regeln (R1) bis (R4) sind nur die allgemeinen Eigenschaften der Wahrscheinlichkeiten festgelegt, nicht jedoch welche Werte sie bei bestimmten Ereignissen annehmen.
- Hierzu muss eine dem Zufallsvorgang adäquate **Wahrscheinlichkeitsinterpretation** vorliegen.

Mit dem subjektiven und statistisch/frequentistischen Wahrscheinlichkeitsbegriff befasst sich das Buch auf S. 25.



Aufgabe 2.10:

Student Paul kommt jeden Morgen zwischen 8.00 und 8.30 Uhr an einer Bushaltestelle an, an der zwei Buslinien halten:

Linie A: 8.04, 8.14, 8.24

Linie B: 8.10, 8.20, 8.30.

Wie groß ist die Wahrscheinlichkeit, dass er Linie A nimmt?

- **a-priori:**

- ▶ **Laplace- (bzw. klassische) Wahrscheinlichkeit:** Anwendung bei Zufallsvorgängen mit endlichem Ω und deren Elementarereignisse die gleiche Eintrittswahrscheinlichkeit besitzen (**Laplace-Experimente**).
- ▶ Laplace-Experimente können als die zufällige Entnahme aus einer endlichen Menge von Objekten charakterisiert werden. Wird mehrfach gezogen, muss zwischen *Ziehen mit Zurücklegen* oder *Ziehen ohne Zurücklegen* unterschieden werden.

Laplace-Wahrscheinlichkeit

Die Wahrscheinlichkeit für ein Elementarereignis $\{\omega_i\}$, $i = 1, \dots, m$ beträgt dann $P(\{\omega_i\}) = \frac{1}{m}$.

$P(A)$ erhält man als

$$P(A) = \frac{\text{Anzahl der für } A \text{ günstigen Ausgänge}}{\text{Anzahl der möglichen Ausgänge}}.$$

Die Anzahl der Elemente einer Menge M heißt **Mächtigkeit** $|M|$. Die Laplace-Wahrscheinlichkeit ist dann

$$P(A) = \frac{|A|}{|\Omega|}.$$

Würfelwurf 2



Aufgabe 2.11: Laplace-Experimente.

- (a) In einer Urne befinden sich 20 Kugeln, von denen 8 rot sind. Berechnen Sie die Wahrscheinlichkeit eine rote Kugel bei einer zufälligen Entnahme zu erhalten.
- (b) Wie groß ist die Wahrscheinlichkeit, nach Ziehen einer roten Kugel ohne Zurücklegen im zweiten Zug erneut eine rote Kugel zu ziehen?
- (c) Ein **Laplace-Würfel** (idealer Würfel) wird geworfen. Berechnen Sie die Wahrscheinlichkeit, dass eine Augenzahl größer als 2 oben liegt.

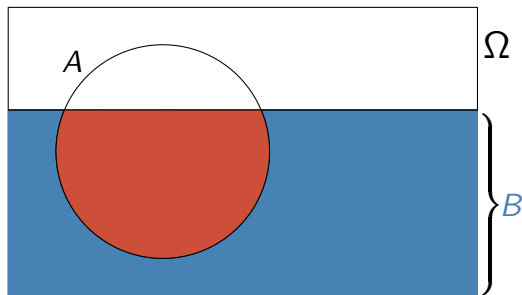
Aufgabe 2.12:

Ein Laplace-Würfel wird einmal geworfen. Wie groß ist die Wahrscheinlichkeit, dass eine Augenzahl kleiner 3 fällt?

Aufgabe 2.13:

Eine Laplace-Münze und ein Laplace-Würfel werden gemeinsam geworfen. Wie groß ist die Wahrscheinlichkeit, dass Kopf und eine Augenzahl größer als 4 fällt?

- Die Berechnung von Wahrscheinlichkeiten erfolgt bislang unter Bezug auf den ganzen Stichprobenraum Ω .
- Es lassen sich aber auch dann Wahrscheinlichkeiten für A berechnen, wenn nicht ganz Ω , sondern nur noch ein Teil B davon relevant ist. Die Abbildung verdeutlicht die Veränderung des Bezugssystems.



- Der Kreis entspricht A , das untere Rechteck B und das rote Kreissegment $A \cap B$.
- $P(A)$, $P(B)$ und $P(A \cap B)$ sind die Wahrscheinlichkeiten für A , B und $A \cap B$, wenn Ω zugrunde liegt.
- Es kann aber auch die Wahrscheinlichkeit für A unter der Bedingung berechnet werden, dass nur noch die Ausgänge in B relevant sind.
- Diese **bedingte Wahrscheinlichkeit** wird mit $P(A|B)$ bezeichnet.
- In der Abbildung ist $P(A)$ das Verhältnis der Kreisfläche zur Fläche des Rechtecks Ω ; die bedingte Wahrscheinlichkeit $P(A|B)$ jedoch das Verhältnis der roten Fläche zur Fläche des Rechtecks B .

Bedingte Wahrscheinlichkeit

Es seien $A, B \subset \Omega$ Ereignisse und $P(B) > 0$. Für die **bedingte Wahrscheinlichkeit** $P(A|B)$ gilt

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

Beispiel 2.14: Idealer Würfel.

Ein idealer Würfel wird geworfen. Also sind $\Omega = \{1, 2, \dots, 6\}$ und $P(\{\omega_i\}) = 1/6$ für $i = 1, \dots, 6$.

A ist das Ereignis, eine 1 zu würfeln; B das Ereignis einer ungeraden Augenzahl. $A \cap B$ tritt bei einer 1 ein. Die Wahrscheinlichkeiten betragen $P(A) = 1/6$, $P(B) = 1/2$ und $P(A \cap B) = 1/6$.

Was ist nun die Wahrscheinlichkeit einer 1 unter der Bedingung, dass eine ungerade Augenzahl eintritt? Der durch die Bedingung gegebene Stichprobenraum lautet $B = \{1, 3, 5\}$, also gilt $P(A|B) = 1/3$.

Denselben Wert erhält man nach der eingeführten Definition:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{1}{6}}{\frac{1}{2}} = \frac{1}{3}.$$

Aufgabe 2.15:

Nach der achtmaligen Durchführung des Zufallsexperiments *zweimaliges Werfen einer Münze* erhält man folgendes Resultat:

Versuch	1	2	3	4	5	6	7	8
1. Wurf	K	Z	Z	Z	K	Z	Z	K
2. Wurf	Z	Z	K	K	K	Z	K	Z

Ereignis A : K beim 1. Wurf, Ereignis B : K beim 2. Wurf

Berechnen Sie folgende Wahrscheinlichkeiten:

- (a) $P(A)$ (b) $P(B)$ (c) $P(A \cap B)$
(d) $P(A|B)$ (e) $P(B|A)$

Aufgabe 2.16:

In einem Dorf leben 1000 Personen. Es ist bekannt, dass 600 Einwohner nach Mallorca in den Urlaub fahren und dass 500 Einwohner die Bild-Zeitung lesen. Zusätzlich weiß man, dass von den 600 Mallorca-Urlaubern 400 die Bild lesen. Wie groß ist die Wahrscheinlichkeit, dass

- (a) eine Person Mallorca-Urlauber ist,
- (b) eine Person Bild-Zeitung-Leser ist,
- (c) eine Person Mallorca-Urlauber und Bild-Zeitung-Leser ist,
- (d) wenn ein Einwohner Bild-Zeitung liest, dieser auch ein Mallorca-Urlauber ist,
- (e) wenn ein Einwohner Mallorca-Urlauber ist, dieser auch Bild-Zeitung liest?

- $P(A|B)$ lässt sich als die Wahrscheinlichkeit von A unter der zusätzlichen Information interpretieren, dass B bereits eingetreten ist.
- Übt diese Information keinen Einfluss auf die Wahrscheinlichkeit von A aus, gilt $P(A|B) = P(A)$.
- A ist dann **unabhängig** von B .
- Dann ist aber auch B unabhängig von A , wie folgende Umformung zeigt:

$$\begin{aligned} P(B|A) &= \frac{P(A \cap B)}{P(A)} = \frac{P(A|B)P(B)}{P(A)} \\ &= \frac{P(A)P(B)}{P(A)} = P(B) \quad (\text{wegen } P(A|B) = P(A)). \end{aligned}$$

Würfelwurf 3



- Diese über die Wahrscheinlichkeiten festgelegte Eigenschaft zweier Ereignisse heißt **paarweise stochastische Unabhängigkeit**.
- Dies liefert eine einfache Prüfung auf stochastische Unabhängigkeit:

Definition 2.17: Paarweise stochastische Unabhängigkeit.

Zwei Ereignisse A und B heißen **paarweise stochastisch unabhängig**, wenn eine der drei folgenden Bedingungen erfüllt ist:

- (1) $P(A \cap B) = P(A)P(B)$,
- (2) $P(A|B) = P(A)$ falls $P(B) > 0$,
- (3) $P(B|A) = P(B)$ falls $P(A) > 0$;

andernfalls bezeichnet man sie als **stochastisch abhängig**.

Beispiel 2.18: Zweimaliges Würfeln.

Ein idealer Würfel wird zweimal geworfen. A sei die *Augenzahl 1 im ersten Wurf*, B sei die *Augenzahl 2 im zweiten Wurf*.

Der Stichprobenraum umfasst 36 geordnete Zahlenpaare, weil zu unterscheiden ist, wann welche Augenzahl eintritt:

$$\Omega = \{(1, 1), \dots, (1, 6), \dots, (6, 1), \dots, (6, 6)\}.$$

Beispiel 2.18: Fortsetzung.

Für die Ereignisse erhält man:

$$A = \{(1, 1), \dots, (1, 6)\}, P(A) = \frac{|A|}{|\Omega|} = \frac{6}{36} = \frac{1}{6},$$

$$B = \{(1, 2), \dots, (6, 2)\}, P(B) = \frac{|B|}{|\Omega|} = \frac{6}{36} = \frac{1}{6},$$

$$A \cap B = \{(1, 2)\}, P(A \cap B) = \frac{|A \cap B|}{|\Omega|} = \frac{1}{36}.$$

Wegen

$$P(A \cap B) = \frac{1}{36} = P(A)P(B) = \frac{1}{6} \cdot \frac{1}{6}$$

folgt paarweise stochastische Unabhängigkeit.

Aufgabe 2.19:

Aus einer Urne mit vier Kugeln mit den Ziffern 1,2,3,4 wird eine entnommen. Für welche Ereignisse liegt paarweise stochastische Unabhängigkeit vor?

- (a) $A = \{1, 2\}$, $B = \{1\}$
- (b) $A = \{1, 2, 3\}$, $B = \{1, 2, 4\}$
- (c) $A = \{1, 2\}$, $B = \{1, 3\}$

- Das Buch auf S. 34 diskutiert die Verallgemeinerung der stochastischen Unabhängigkeit auf mehr als zwei Ereignisse.
- Sind Ereignisse gemeinsam stochastisch unabhängig, dann sind sie auch paarweise stochastisch unabhängig. Die Umkehrung gilt allerdings nicht!

Mit dem Konzept der bedingten Wahrscheinlichkeiten lassen sich nützliche Sätze der Wahrscheinlichkeitsalgebra gewinnen:

Definition 2.20: Multiplikationssatz für 2 Ereignisse.

Es seien $A, B \subset \Omega$ Ereignisse mit $P(A) > 0$ und $P(B) > 0$. Dann gilt

(a) bei stochastischer Abhängigkeit:

$$P(A \cap B) = P(A|B)P(B) = P(B|A)P(A);$$

(b) bei stochastischer Unabhängigkeit:

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

Die Verallgemeinerung der Ergebnisse des Multiplikationssatzes für zwei Ereignisse führt zum Multiplikationssatz für n Ereignisse:

Definition 2.21: Multiplikationssatz für n Ereignisse.

Es seien $A_j \subset \Omega$, $j = 1, \dots, n$ Ereignisse mit $A_1 \cap A_2 \cap \dots \cap A_{n-1} \neq \emptyset$. Dann gilt

(a) bei stochastischer Abhängigkeit:

$$P(A_1 \cap A_2 \cap \dots \cap A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \cdot \dots \\ \cdot P(A_n|A_1 \cap A_2 \cap \dots \cap A_{n-1});$$

(b) bei gemeinsamer stochastischer Unabhängigkeit:

$$P(A_1 \cap A_2 \cap \dots \cap A_n) = P(A_1)P(A_2) \cdot \dots \cdot P(A_n). \quad (2.1)$$

Beispiel 2.22: Geburten.

Ein Ehepaar möchte wissen welche der drei folgenden Anordnungen von vier Kindern am wahrscheinlichsten ist: JJJJ, MMMM oder JMJM.

Wir unterstellen, dass die Wahrscheinlichkeiten für einen *Jungen* (J) und ein *Mädchen* (M) jeweils 0.5 sind. Dann sind alle drei Anordnungsmöglichkeiten gleich wahrscheinlich.

Der Grund dafür ist, dass die Ereignisse gemeinsam stochastisch unabhängig sind:

$$P(JJJJ) = P(MMMM) = P(JMJM) = 0.5 \cdot 0.5 \cdot 0.5 \cdot 0.5 = 0.5^4 = 0.0625.$$

Die Wahrscheinlichkeit für 2 J und 2 M ist allerdings wesentlich höher:

$$P(2J \text{ und } 2M) = \frac{6}{16} = 0.375.$$

Aufgabe 2.23:

Eine Urne enthält 5 weiße und 5 rote Kugeln. Wie groß ist die Wahrscheinlichkeit, dass gleich die ersten drei Kugeln, die gezogen werden, rot sind (ZoZ)?

Aufgabe 2.24:

In einer Urne befinden sich 3 schwarze und 2 weiße Kugeln. Wie groß ist die Wahrscheinlichkeit, nach dreimaligem Ziehen ohne Zurücklegen

- (a) eine weiße Kugel im zweiten Zug,
- (b) im dritten Zug die erste weiße Kugel,
- (c) alle weißen Kugeln
zu ziehen?

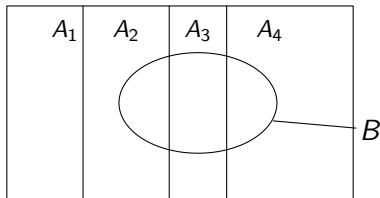
Jede Wahrscheinlichkeit $P(B) > 0$ kann auf bedingte Wahrscheinlichkeiten zurückgeführt werden:

Definition 2.25: Satz von der totalen Wahrscheinlichkeit.

Es sei $\{A_1, \dots, A_n\}$, mit A_j für $j = 1, \dots, n$ ein vollständiges System von Ereignissen. Für das Ereignis B gilt dann

$$\begin{aligned} P(B) &= P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + \dots + P(B|A_n)P(A_n) \\ &= \sum_{j=1}^n P(B|A_j)P(A_j). \end{aligned} \tag{2.2}$$

- Die folgende Abbildung zeigt eine solche Zerlegung:



- $P(B)$ ist gleich dem gewogenen arithmetischen Mittel der bedingten Wahrscheinlichkeiten $P(B|A_j)$ mit Gewichten $P(A_j)$:

$$P(B) = \sum_{j=1}^n P(B|A_j)P(A_j).$$

Aufgabe 2.26:

In Urne 1 befinden sich 14 weiße und 6 schwarze Kugeln, in Urne 2 befinden sich 2 weiße und 8 schwarze Kugeln und in Urne 3 befinden sich 3 weiße und 7 schwarze Kugeln.

Um aus einer Urne Kugeln entnehmen zu dürfen, muss man würfeln. Bei einer 1, 2 oder 3 darf man eine Kugel aus Urne 1 ziehen, bei einer 4 oder 5 aus Urne 2 und bei einer 6 aus Urne 3.

Wie groß ist die Wahrscheinlichkeit eine weiße Kugel zu ziehen?

Aufgabe 2.27:

In einer Stadt sind 60% der Einwohner Frauen und 40% Männer. Es ist bekannt, dass 30% der Frauen und 50% der Männer rauchen.

Wie groß ist die Wahrscheinlichkeit, dass ein Raucher männlich ist?

- Mit dem **Satz für die bedingte Wahrscheinlichkeit** und dem **Multiplikationssatz für 2 Ereignisse** lässt sich eine weitere Regel gewinnen, vorausgesetzt $P(B) > 0$:

$$\begin{aligned} P(A_i|B) &= \frac{P(A_i \cap B)}{P(B)} \quad (\text{nach Satz für bedingte Wahrscheinlichkeit}) \\ &= \frac{P(B|A_i)P(A_i)}{P(B)} \quad (\text{nach Multiplikationssatz}) \end{aligned}$$

- Ersetzt man im letzten Bruch $P(B)$ durch den Satz der totalen Wahrscheinlichkeit (2.25), erhält man das **Bayes'sche Theorem**:

Definition 2.28: Satz von Bayes.

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_{j=1}^n P(B|A_j)P(A_j)} \quad \text{für alle } i = 1, \dots, n.$$

- Das Theorem von Bayes gibt an, wie die Wahrscheinlichkeit für ein Ereignis A nach Durchführen eines Experimentes oder einer Beobachtung neu berechnet werden kann.
- Interpretiere hierzu das vollständige Ereignissystem $A_j, j = 1, \dots, n$ als Menge sich ausschließender Hypothesen oder disjunkter Zustände. Die Wahrscheinlichkeiten $P(A_j)$ sind aufgrund von Ausgangsinformationen bestimmt; sie sind daher **a-priori-Wahrscheinlichkeiten** für jedes A_j .
- Mit dem Bayes'schen Theorem erhält man jetzt die **a-posteriori-Wahrscheinlichkeiten** für A_j , wenn die Beobachtung B bzw. das experimentelle Ergebnis B vorliegt. Das Eintreten des Ereignisses B ist eine zusätzliche Information.

Beispiel 2.29: Bayes'sches Theorem.

Ein Test erkennt eine Krankheit mit einer Wahrscheinlichkeit von 95%, wenn diese tatsächlich vorliegt. Es gibt aber auch *falsche positive* Befunde mit einer Wahrscheinlichkeit von 1%. 0.1% der Bevölkerung hat die Krankheit. Was ist die Wahrscheinlichkeit, dass eine Person die Krankheit hat, wenn der Test positiv ausfällt?

$$P(K|+) = \frac{P(+|K) P(K)}{P(+)} = \frac{0.95 \cdot 0.001}{0.95 \cdot 0.001 + 0.01 \cdot 0.999} = 0.0868,$$

wobei $\Omega = K \cup \bar{K}$ mit $K \cap \bar{K} = \emptyset$ und daher:

$$\begin{aligned} P(+) &= P(+ \cap K) + P(+ \cap \bar{K}) \\ &= P(+|K) P(K) + P(+|\bar{K}) P(\bar{K}). \end{aligned}$$

Beispiel 2.29: Fortsetzung.

Der Satz von Bayes findet in sehr vielen unterschiedlichen Disziplinen Anwendung, z.B. wie [hier](#) zu sehen auch in der Justiz.

Oder auch in Situationen des alltäglichen Lebens wie [hier](#).

Auch Spamfilter funktionieren nach diesem Prinzip.

Aufgabe 2.30:

Das Angelrevier eines Hobbyanglers besteht aus den drei Seen S1, S2 und S3, wobei er keinen der drei Seen bevorzugt.

Die Wahrscheinlichkeit dafür, dass er innerhalb einer Stunde einen Fisch fängt, beträgt für den See S1 60%, für den See S2 50% und für den See S3 80%.

Nach einer Stunde Angelzeit kann sich der Hobbyangler über den Fang eines Fisches freuen. Wie groß ist die Wahrscheinlichkeit, dass der Fisch in dem See S2 geangelt wurde?

Aufgabe 2.31:

Ein Autohersteller bezieht Airbags von 2 Zulieferern. Beim Zulieferer 1, der insgesamt 65% aller Airbags liefert, sind 98% der Airbags fehlerfrei. 95% der von Zulieferer 2 gelieferten Airbags sind nicht defekt.

- (a) Wenn der Hersteller ein fehlerhaftes Teil bekommt, wie groß ist die Wahrscheinlichkeit, dass es von Zulieferer 1 kommt?
- (b) Wie groß ist die Wahrscheinlichkeit, dass ein fehlerhafter Airbag von Zulieferer 2 kommt?
- (c) Wie groß ist die Wahrscheinlichkeit, dass ein geliefertes Airbag defekt ist?

- Zur Berechnung von Wahrscheinlichkeiten nach dem klassischen Konzept müssen die „günstigen“ Ausgänge zu den möglichen Ausgängen ins Verhältnis gesetzt werden.
- Oft ist aber gerade die Ermittlung der günstigen Ausgänge nicht einfach.
- Formeln für ihre Berechnung sind daher nützlich. Der Zweig der Mathematik, der sich hiermit beschäftigt, ist die **Kombinatorik**.
- Sie untersucht die verschiedenen Möglichkeiten der Anordnung oder Auswahl von Objekten, im Folgenden „Elemente“.

- Die Entwicklung der Kombinatorikformeln lässt sich anschaulich an einer Gesamtheit mit N nummerierten Kugeln $1, 2, \dots, N$ zeigen.
- Jede Zusammenstellung aller N Kugeln (Zahlen) in irgendeiner Anordnung ist eine **Permutation**. Die Anzahl unterscheidbarer Permutationen erhält man durch folgende Überlegung.

- Um die erste Stelle der Anordnung zu besetzen, stehen N Kugeln zur Auswahl. Danach bleiben zur Besetzung der zweiten Stelle $N - 1$, für die dritte Stelle $N - 2$ Kugeln und schließlich für die letzte (N -te) Stelle nur noch eine Kugel übrig.
- Die Gesamtzahl unterscheidbarer Permutationen ist das Produkt aus all diesen Besetzungsmöglichkeiten, also

$$N(N - 1)(N - 2) \cdot \dots \cdot 1.$$

- Ein solches Produkt heißt **N -Fakultät**, geschrieben als $N!$, wobei $N \in \mathbb{N}$.
- Für $N = 4$ erhält man $4! = 4 \cdot 3 \cdot 2 \cdot 1 = 24$, für $N = 1$: $1! = 1$.
- Damit die Rekursionsformel $N! = N(N - 1)!$ uneingeschränkt auf \mathbb{N} gilt, definiere zudem $0! = 1$.
- Bezeichnet man die Permutationen mit P , gibt der folgende Satz die Anzahl der Anordnungen von N verschiedenen Elementen wieder:

Definition 2.32: Permutation ohne Wiederholung.

$$P(N) = N! = N(N - 1)(N - 2) \cdot \dots \cdot 1.$$

Kombinatorik



Beispiel 2.33: Kugeln.

Es sollen drei Kugeln mit den Ziffern 1, 2 und 3 unterschiedlich angeordnet werden. Die möglichen Permutationen lauten:

1	2	3	1	3	2	3	1	2
2	1	3	2	3	1	3	2	1

Ihre Anzahl erhält man nach der Formel für Permutationen ohne Wiederholung als

$$3! = 6.$$

```
### Beispiel: Kugeln (1/3)
```

```
prod(1:3)
```

```
## [1] 6
```

```
factorial(3) # kürzer
```

```
## [1] 6
```

- Wie zuvor: Urne mit N nummerierten Kugeln $1, 2, \dots, N$
- Neue Situation: Es sollen n Elemente aus der Urne mit N Kugeln gezogen werden ($n < N$). Dabei spielt die Reihenfolge der Entnahme keine Rolle.
- Da die Reihenfolge unerheblich ist, lassen sich die N Elemente in n Elemente, die gezogen werden und in $N - n$ ausgeschlossene Elemente unterteilen.
- Die N Elemente unterscheiden sich jetzt nur noch hinsichtlich dieser beiden Merkmale.
- Die Formel für die Anzahl an Möglichkeiten, n Kugeln aus N zu ziehen mit $n < N$ kann über Permutationen entwickelt werden.

Beispiel 2.34: Lottozahlen (6 aus 49).

Bei der Lottoziehung wird $n = 6$ mal aus einer Lostrommel mit $N = 49$ nummerierten Kugeln gezogen. Für die Wahrscheinlichkeit beim Lotto zu gewinnen, d.h. alle 6 gezogenen Zahlen richtig zu tippen, muss die Anzahl der günstigen durch die Anzahl der möglichen Ausgänge dividiert werden.

Die Anzahl der möglichen Ausgänge können wir wie folgt herleiten:

Es gibt insgesamt $N = 49$ Möglichkeiten, die erste Kugel zu ziehen.

Danach gibt es noch $N - 1 = 48$ Möglichkeiten, die zweite Kugel zu ziehen, usw.

Schließlich gibt es noch $N - n + 1 = 44$ Möglichkeiten, die 6. Kugel zu ziehen.

Insgesamt erhalten wir

$$49 \cdot 48 \cdot 47 \cdot 46 \cdot 45 \cdot 44 = N \cdot (N - 1) \cdot \dots \cdot (N - n + 1)$$

Möglichkeiten, 6 Kugeln aus 49 Kugeln zu ziehen.

Beispiel 2.34: Fortsetzung.

$$49 \cdot 48 \cdot 47 \cdot 46 \cdot 45 \cdot 44 = N \cdot (N - 1) \cdot \dots \cdot (N - n + 1)$$

kann auch über die Fakultät dargestellt werden

$$49 \cdot 48 \cdot 47 \cdot 46 \cdot 45 \cdot 44 = \frac{49!}{43!}$$

Nun sind aber jeweils genau $6!$ der Möglichkeiten, 6 aus 49 Kugeln zu ziehen, Permutationen voneinander. Durch Division der $6!$ Permutationen der selben gezogenen Kugeln erhalten wir

$$\frac{49!}{43!6!} = 13.983.816$$

Möglichkeiten 6 Kugeln aus 49 zu ziehen, ohne dabei auf die Reihenfolge der Entnahme zu achten.

Beispiel 2.34: Fortsetzung.

Da nur ein günstiger Ausgang existiert, beträgt die gesuchte Wahrscheinlichkeit:

$$P(G) = \frac{1}{13.983.816} = 0.0000000715 \hat{=} 0.00000715\%.$$

Aber das heißt nicht, dass es unmöglich ist. Siehe z.B. [hier](#)

Allgemein erhält man

$$\begin{aligned}\frac{N!}{n!(N-n)!} &= \frac{N(N-1) \cdot \dots \cdot (N-n+1)(N-n)(N-n-1) \cdot \dots \cdot 1}{n(n-1) \cdot \dots \cdot 1 \cdot (N-n)(N-n-1) \cdot \dots \cdot 1} \\ &= \frac{N(N-1) \cdot \dots \cdot (N-n+1)}{n(n-1) \cdot \dots \cdot 1}\end{aligned}$$

Dies ist der **Binomialkoeffizient** $\binom{N}{n}$, gelesen „ N über n “

Definition 2.35: Binomialkoeffizient.

$$\frac{N!}{n!(N-n)!} = \binom{N}{n}. \quad (2.3)$$

Beispiel 2.36: Texas Hold'em.

Die Wahrscheinlichkeit bei der Poker-Variante Texas Hold'em eine Starthand mit zwei Assen zu erhalten, kann ebenfalls mit der Formel für die Kombination ohne Wiederholung berechnet werden. Anzahl der günstigen Ausgänge:

$$\binom{4}{2} = \frac{4!}{2!(4-2)!} = 6 :$$

(♠, ♣), (♣, ♦), usw. Anzahl der möglichen Ausgänge (Starthände):

$$\binom{52}{2} = \frac{52!}{2!(52-2)!} = 1326.$$

Wahrscheinlichkeit:

$$P(\text{zwei Asse}) = \frac{6}{1326} = 0.0045 \hat{=} 0.45\%.$$

```
## # Beispiel: Lottozahlen (6 aus 49)
# Anzahl möglicher Ergebnisse
factorial(49) / (factorial(6) * factorial(43))
```

```
## [1] 13983816
```

```
choose(49, 6) # kürzer
```

```
## [1] 13983816
```

```
1/choose(49,6) #WSK für 6 richtige im Lotto
```

```
## [1] 7.151e-08
```

```
### Beispiel Texas Hold'em
```

```
(hands <- choose(52,2))
```

```
## [1] 1326
```

```
(aces <- choose(4,2))
```

```
## [1] 6
```

```
aces/hands #WSK für 2 Asse als Starthand
```

```
## [1] 0.004525
```

Aufgabe 2.37:

Eine Maschine kann 10 verschiedene Werkstücke produzieren. Wie viele verschiedene Fertigungsreihenfolgen dieser 10 Werkstücke gibt es?

Aufgabe 2.38:

Eine Urne enthält 5 Bälle. Wie viele Anordnungsmöglichkeiten gibt es, wenn die Bälle mit den Ziffern 1 bis 5 durchnummeriert sind?

Aufgabe 2.39:

Wie groß ist die Wahrscheinlichkeit bei einem Kartenspiel 13 Karten zu erhalten, von denen 7 Karten Pik sind? (Hinweis: ein Kartenspiel besteht aus 52 Karten: 13 Kreuz, 13 Pik, 13 Herz, 13 Karo)

- 1 Einleitung
- 2 Grundlagen
- 3 Eindimensionale Zufallsvariablen**
- 4 Ausgewählte theoretische Verteilungen
- 5 Grundzüge der Stichprobentheorie
- 6 Statistische Schätzverfahren
- 7 Statistische Testverfahren
- 8 Zweidimensionale Zufallsvariablen

- Die in Ω aufgeführten Ausgänge ω_i müssen keine **Zahlen** sein, sondern sind oft **artmäßig** angegeben.
- Beim *Münzwurf* ist der artmäßig festgelegte Stichprobenraum

$$\Omega = \{\omega_1 : \textit{Kopf liegt oben}, \omega_2 : \textit{Zahl liegt oben}\}.$$

- Beim Würfelbeispiel ist die Überführung der Ausgänge in Zahlen einfach. Setzt man $\omega_i = i$, $i = 1, \dots, 6$, enthält der Stichprobenraum nur Zahlen:

$$\Omega = \{1, \dots, 6\}.$$

- Der Münzwurf liefert keine natürliche Zahlenvorgabe. Die Zuordnung könnte lauten: *Kopf* $\rightarrow 0$ und *Zahl* $\rightarrow 1$, aber auch: *Kopf* $\rightarrow 12$ und *Zahl* $\rightarrow 27$.

Definition 3.1: Eindimensionale Zufallsvariable.

Eine Funktion X , die jedem Ausgang $\omega \in \Omega$ eine reelle Zahl $x \in \mathbb{R}$ zuordnet,

$$X : \Omega \longrightarrow \mathbb{R},$$

heißt **eindimensionale Zufallsvariable** oder kurz auch **Zufallsvariable**, wenn $\{\omega \in \Omega | X(\omega) \leq x\}$ für alle $x \in \mathbb{R}$ ein Ereignis ist.

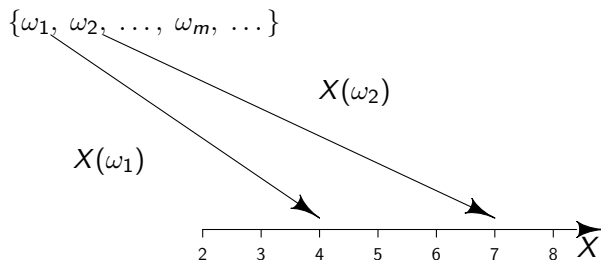
- $x = X(\omega)$ ist **Wert** bzw. **Realisation** von X ; die Menge

$$\{x | x = X(\omega) \in \mathbb{R}, \omega \in \Omega\} = X(\Omega)$$

heißt **Wertebereich**. Er gibt an, welche Werte die Zufallsvariable X bei gegebenem Ω annehmen kann.

- Durch die Bedingung in der vorangeführten Definition lassen sich alle Ereignisse A als Intervalle der reellen Zahlen oder als reelle Zahlen selbst angeben, die jetzt die Ereignisse für X darstellen.
- Umgekehrt folgt, dass für jedes Intervall $X \leq x$ ein Ereignis existiert. X bildet somit Ω und jedes Ereignis auf die reellen Zahlen ab.

- Die nachfolgende Grafik veranschaulicht diesen Zusammenhang:



- Für

$$A = \{\omega \in \Omega | X(\omega) \leq x\}$$

existiert die **Wahrscheinlichkeit** $P(A)$, die jetzt als die Wahrscheinlichkeit für das Ereignis $\{X \leq x\}$ interpretiert werden kann:

$$P(A) = P(\{X \leq x\}) = P(X \leq x).$$

- X überträgt somit auch die Wahrscheinlichkeiten von den Ereignissen eines Zufallsexperimentes auf die reellen Zahlen.
- Spezifiziert man so für alle Ereignisse des neuen Stichprobenraumes $X(\Omega)$ die Wahrscheinlichkeiten, erhält man die **Wahrscheinlichkeitsverteilung** von X .

Beispiel 3.2: Münzwurf.

Den Stichprobenraum für das *2-malige Werfen einer Münze* erhält man als

$$\Omega = \{(K, K), (K, Z), (Z, K), (Z, Z)\}, \quad Z: \text{Zahl}, K: \text{Kopf}.$$

X bildet die Ausgänge ω_i , $i = 1, \dots, 4$ wie folgt in die reellen Zahlen ab

$$X = \begin{cases} 0 & (K, K) \\ 1 & (K, Z) \quad \text{oder} \quad (Z, K) \\ 2 & (Z, Z). \end{cases}$$

Beispiel 3.3: Münzwurf.

Als Wertebereich erhält man $X(\Omega) = \{0, 1, 2\}$. Für z.B. $X(\omega) = 0$ lautet das zugehörige Ereignis $A_1 = \{(K, K)\}$. $\{X \leq 1\}$ entspricht dem Ereignis $A_2 = \{(K, K), (K, Z), (Z, K)\}$.

Ist die Münze fair, sind alle Elementarereignisse gleich wahrscheinlich:

$$P(\{K, K\}) = P(\{K, Z\}) = P(\{Z, K\}) = P(\{Z, Z\}) = \frac{1}{4}.$$

$P(\{X = 0\})$ und $P(\{X \leq 1\})$ ermittelt man dann über $P(A_1)$ und $P(A_2)$:

$$P(X = 0) = P(A_1) = \frac{1}{4} \quad \text{und} \quad P(X \leq 1) = P(A_2) = \frac{3}{4}.$$

- Ist Ω endlich oder abzählbar unendlich, liegt eine **diskrete** Zufallsvariable vor.
- Bei einem überabzählbar unendlichen Stichprobenraum kann X jeden beliebigen Zahlenwert eines vorgegebenen Intervalls der reellen Zahlen annehmen, wobei auch $(-\infty, \infty)$ zulässig ist.
- Die Zufallsvariable X nennen wir dann **stetig**.

Zufallsvariablen 1



Beispiel 3.4: diskrete / stetige Zufallsvariable.

- (a) Ist X die Anzahl der Würfe einer Münze, bis zum ersten Mal *Zahl* oben liegt, sind Ω und Wertebereich $X(\Omega)$ abzählbar unendlich.

X ist daher eine **diskrete** Zufallsvariable mit $X(\omega_i) = x_i = i$ für $i = 1, 2, \dots, m, m+1, \dots$

- (b) Bei *Verspätung eines Zuges* X kann bei unendlich genauer Zeitmessung jede Verspätung eines Intervalls $[x_1, x_2] \subset \mathbb{R}$ eintreten.

X hat einen überabzählbar unendlichen Wertebereich und ist daher **stetig**.

- Jede auf $X(\Omega)$ definierte reellwertige Funktion g überführt X in eine neue Zufallsvariable $Y = g(X)$.
- Auch Y liegt der Stichprobenraum Ω zugrunde, denn Y stellt eine geschachtelte Funktion dar:

$$Y = g[X(\Omega)].$$

- X sei im Beispiel 3.4 (b) in Minuten gemessen. Für die Verspätung in Sekunden resultiert eine Zufallsvariable $Y = 60X$. Ist $X = 3$ (Minuten), folgt $Y = 180$ (Sekunden). Die Wahrscheinlichkeit für eine Verspätung zwischen 2 und 3 Minuten ist dieselbe wie für eine Verspätung zwischen 120 und 180 Sekunden:

$$P(2 \leq X \leq 3) = P(120 \leq Y \leq 180).$$

- Man überträgt die Wahrscheinlichkeiten für Ereignisse aus Ω jetzt auf die Ereignisse von Y .

- Um z.B. die Wahrscheinlichkeit für Ereignisse der Art

$$\{X \leq x\}, \{x_1 < X \leq x_2\}, \{X = x_1\} \text{ oder } \{X \neq x_1\}$$

zu berechnen, ist die **Verteilungsfunktion** hilfreich.

Definition 3.5: Verteilungsfunktion.

Eine Funktion F mit

$$F(x) = P(X \leq x), \quad x \in \mathbb{R}, \quad (3.1)$$

heißt **Verteilungsfunktion** von X .

- Aufgrund der **Kolmogoroff-Axiome** hat $F(x)$ folgende Eigenschaften:

(1) $F(x)$ ist **monoton wachsend** mit Werten des Intervalls $[0,1]$: Gilt $x_2 > x_1$, so folgt daraus $F(x_2) \geq F(x_1)$.

(2)

$$\lim_{x \rightarrow -\infty} F(x) = F(-\infty) = 0$$

und

$$\lim_{x \rightarrow \infty} F(x) = F(\infty) = 1.$$

(3) $F(x)$ ist an jeder Stelle x **rechtsseitig stetig**, d.h.

$$\lim_{h \rightarrow 0^+} F(x + h) = F(x).$$

Die Notation $h \rightarrow 0^+$ besagt, dass h mit einem Wert größer 0 startet und bei der Grenzwertberechnung sich daher von rechts der 0 nähert.

(4) Für jedes Ereignis $\{x_1 < X \leq x_2\}$ gilt

$$P(x_1 < X \leq x_2) = P(X \leq x_2) - P(X \leq x_1) = F(x_2) - F(x_1).$$

- Diese vier allgemeinen Eigenschaften gelten für diskrete und stetige Zufallsvariablen.
- Da aber die weitere mathematische Behandlung eine Fallunterscheidung verlangt, erfolgt eine getrennte Darstellung.

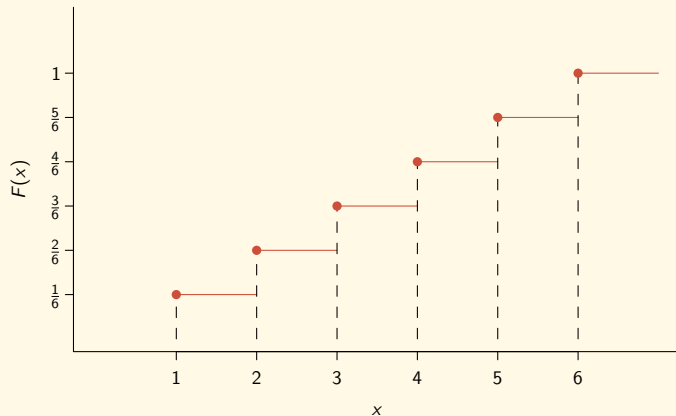
Beispiel 3.6: Fairer Würfel.

Der Wertebereich ist $X(\Omega) = \{1, 2, 3, 4, 5, 6\}$ mit $P(X = x_i) = \frac{1}{6}$ für $i = 1, \dots, 6$.

Die Verteilungsfunktion $F(x)$ ist dann

$$F(x) = \begin{cases} 0 & x < 1 \\ \frac{1}{6} & 1 \leq x < 2 \\ \frac{2}{6} & 2 \leq x < 3 \\ \vdots & \\ \frac{5}{6} & 5 \leq x < 6 \\ 1 & x \geq 6. \end{cases}$$

Beispiel 3.6: Fortsetzung.



- Bei diskreten Zufallsvariablen ist $F(x)$ eine **Treppenfunktion**.
- Die Stufenhöhen entsprechen den Wahrscheinlichkeiten für $X = x_i$.
- Die **Wahrscheinlichkeitsfunktion** $f(x)$ einer diskreten Zufallsvariablen ordnet den $x_i = X(\omega_i)$ Wahrscheinlichkeiten zu.
- $f(x)$ heißt manchmal auch **Wahrscheinlichkeitsmassefunktion**.

Definition 3.7: Wahrscheinlichkeitsfunktion.

Die Funktion

$$f : X(\Omega) \longrightarrow \mathbb{R}^+$$

mit

$$f(x) = \begin{cases} P(X = x_i) = p_i & i = 1, 2, \dots \\ 0 & \text{sonst.} \end{cases}$$

heißt **Wahrscheinlichkeitsfunktion** von X und ordnet jeder reellen Zahl x_i die Wahrscheinlichkeit $P(X = x_i)$ zu.

Für diese gilt

$$f(x_i) \geq 0 \quad \text{und} \quad \sum_i f(x_i) = 1.$$

Beispiel 3.7: Fortsetzung.

$$f(x) = \begin{cases} P(X = x_i) = p_i = \frac{1}{6} & i = 1, \dots, 6 \\ 0 & \text{sonst.} \end{cases}$$

$$\begin{aligned} P(x_1 < X \leq x_2) &= \sum_{x_1 < x_i \leq x_2} f(x_i) = \sum_{x_1 < x_i \leq x_2} P(X = x_i) \\ &= \sum_{i: x_1 < x_i \leq x_2} p_i. \end{aligned}$$

- Die Intervallgrenzen x_1 und x_2 sind beliebig; sie müssen nicht Elemente von $X(\Omega)$ sein. Für bspw. $x_1 = 1.5$ und $x_2 = 4.8$ folgt für den **fairen Würfel**

$$\begin{aligned} P(1.5 < X \leq 4.8) &= P(X = 2) + P(X = 3) + P(X = 4) \\ &= p_1 + p_2 + p_3 = \frac{1}{2}. \end{aligned}$$

Aufgabe 3.8:

Gegeben ist das Zufallsexperiment *dreimaliger Münzwurf einer unfairen Münze* mit $P(K) = \frac{1}{3}$ und $P(Z) = \frac{2}{3}$.

- (a) Stellen Sie den Stichprobenraum Ω und den Wertebereich $X(\Omega)$ für X : *Anzahl Kopf* dar!
- (b) Entwickeln Sie die Wahrscheinlichkeitsfunktion und stellen Sie diese grafisch dar!
- (c) Entwickeln Sie die Verteilungsfunktion und stellen Sie diese auch grafisch dar!

Aufgabe 3.9:

Für Zufallsvariable X : *Anzahl der Autoverkäufe pro Tag* gilt:

X	0	1	2	3	4	5
$f(x)$	0.18	0.39	0.24	0.14	0.04	0.01
$F(x)$	0.18	0.57	0.81	0.95	0.99	1

Wie groß ist die Wahrscheinlichkeit, dass

- (a) am Tag ein Auto verkauft wird?
- (b) am Tag höchstens 2 Autos verkauft werden?
- (c) am Tag mehr als 2, aber höchstens 4 Autos verkauft werden?
- (d) am Tag mindestens 2, aber höchstens 4 Autos verkauft werden?

- Eine **stetige Zufallsvariable** kann jeden Wert eines Intervalls der reellen Zahlen annehmen.
- $P(X \leq x)$ resultiert hier wegen der überabzählbar unendlich vielen Realisationen von X durch **Integration** einer geeigneten Funktion $f(x)$.
- Da $-\infty < X < \infty$ das sichere Ereignis ist, muss gelten:

$$f(x) \geq 0 \quad \text{und} \quad P(-\infty < X < \infty) = \int_{-\infty}^{\infty} f(x) dx = 1.$$

Derartige Funktionen heißen **Dichtefunktionen**.

Definition 3.10: Dichtefunktion.

Eine Funktion $f : X(\Omega) \rightarrow \mathbb{R}^+$ heißt **Wahrscheinlichkeitsdichte** oder **Dichtefunktion**, wenn gilt

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(u) du \quad \text{für jedes } x \in \mathbb{R}.$$

- Für alle Werte von x , in denen $F(x)$ differenzierbar ist, erhält man

$$\frac{dF(x)}{dx} = f(x);$$

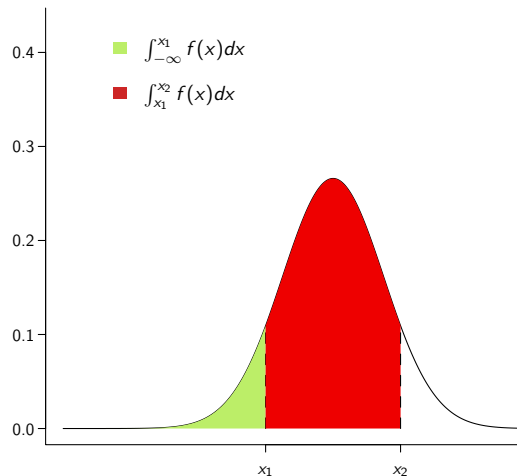
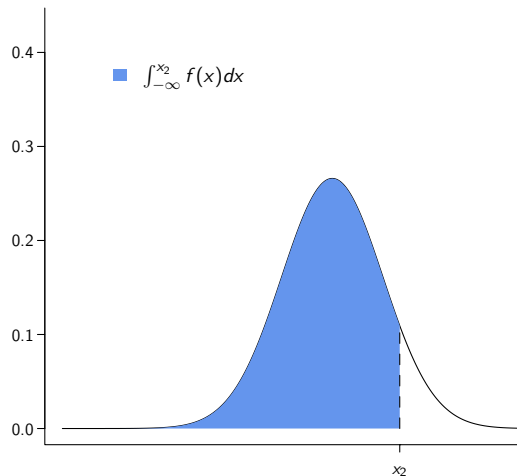
d.h. die Dichtefunktion ist die erste Ableitung der Verteilungsfunktion.

- Die Wahrscheinlichkeit für ein Ereignis $(x_1 < X \leq x_2)$ ist dann

$$\begin{aligned} P(x_1 < X \leq x_2) &= F(x_2) - F(x_1) = \int_{-\infty}^{x_2} f(x) dx - \int_{-\infty}^{x_1} f(x) dx \\ &= \int_{x_1}^{x_2} f(x) dx. \end{aligned}$$

Wahrscheinlichkeitsverteilungen von Zufallsvariablen

Wahrscheinlichkeitsdichte



- Für einen einzelnen Wert gilt also im stetigen Fall, dass

$$P(X = x_1) = \int_{x_1}^{x_1} f(x) dx = 0$$

für beliebiges x_1 .

- Das heißt aber nicht, dass diese Ereignisse unmöglich sind! Vielmehr zeigt es, dass $f(x)$ im stetigen Fall nicht die Wahrscheinlichkeit für $X = x$ angibt.
- Daher ist auch möglich, dass für einige Werte $f(x) > 1$ gilt.
- Da im stetigen Fall immer gilt $P(X = x) = 0$, stimmen folgende Wahrscheinlichkeiten überein:

$$\begin{aligned} P(x_1 \leq X \leq x_2) &= P(x_1 < X \leq x_2) = P(x_1 \leq X < x_2) \\ &= P(x_1 < X < x_2). \end{aligned}$$

Beispiel 3.11: Berechnung von Wahrscheinlichkeiten bei stetigen ZV.

Eine stetige Zufallsvariable X hat die Verteilungsfunktion

$$F(x) = \begin{cases} 0 & x \leq 0 \\ \frac{1}{8}x^3 & 0 < x < 2 \\ 1 & x \geq 2. \end{cases}$$

Da $F(x)$ über den drei Teilintervallen differenzierbar ist, ist die Dichtefunktion die Ableitung von $F(x)$:

$$f(x) = \begin{cases} \frac{3}{8}x^2 & 0 < x \leq 2 \\ 0 & \text{sonst.} \end{cases}$$

Zufallsvariablen 2



Beispiel 3.12: Berechnung von Wahrscheinlichkeiten bei stetigen ZV.

$P(1 < X \leq \frac{3}{2})$ lässt sich über $F(x)$ und $f(x)$ berechnen:

$$P\left(1 < X \leq \frac{3}{2}\right) = F\left(\frac{3}{2}\right) - F(1) = \frac{27}{64} - \frac{1}{8} = \frac{19}{64};$$

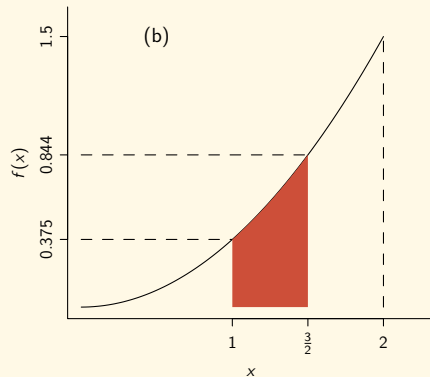
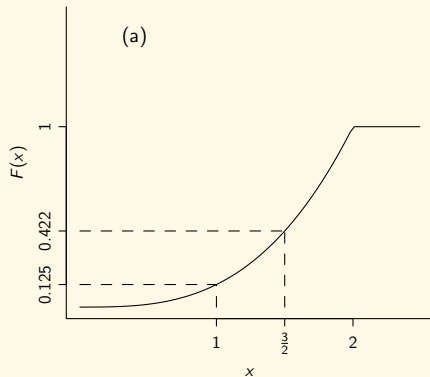
aber auch

$$\int_1^{\frac{3}{2}} \frac{3}{8}x^2 dx = \frac{1}{8}x^3 \Big|_1^{\frac{3}{2}} = \frac{27}{64} - \frac{1}{8} = \frac{19}{64}.$$

Die folgenden Grafiken zeigen beide Funktionen.

Abbildung (a) gibt die Wahrscheinlichkeit als Differenz der beiden Funktionswerte, Abbildung (b) als Flächeninhalt wieder.

Beispiel 3.12: Fortsetzung.



```
# Verteilungsfunktion definieren
```

```
myCdf <- function(x) {  
  if(x < 0){  
    cdf <- 0  
  } else if(x >= 0 & x <= 2){  
    cdf <- 0.125*x^3  
  } else {  
    cdf <- 1  
  }  
  cdf  
}
```

```
# Dichtefunktion definieren
```

```
myPdf <- function(x) ifelse(x > 0 & x <= 2, .375 * x^2, 0)
```

```
myCdf(1.5) - myCdf(1)
```

```
## [1] 0.2969
```

```
integrate(f = myPdf, lower = 1, upper = 1.5)
```

```
## 0.2969 with absolute error < 3.3e-15
```

Aufgabe 3.13:

Für die Zufallsvariable X gilt

$$f(x) = \begin{cases} \frac{1}{2}x & 0 \leq x \leq 2 \\ 0 & \text{sonst.} \end{cases}$$

- (a) Zeigen Sie, dass $f(x)$ eine Dichtefunktion ist und stellen Sie diese grafisch dar!
- (b) Berechnen Sie die Verteilungsfunktion und stellen Sie diese grafisch dar!
- (c) Bestimmen Sie die folgenden Wahrscheinlichkeiten:

(i) $P(X = 1)$ (ii) $P(X \leq 1)$

(iii) $P(\frac{1}{2} \leq X \leq \frac{3}{2})$ (iv) $P(\frac{1}{2} < X \leq \frac{3}{2})$

Aufgabe 3.14:

Die Zufallsvariable X hat die folgende Verteilungsfunktion

$$F(x) = \begin{cases} 0 & x \leq 0 \\ \frac{1}{27}(x-3)^3 + 1 & 0 < x < 3 \\ 1 & x \geq 3. \end{cases}$$

Bestimmen Sie die Dichtefunktion!

- Wie in der deskriptiven Statistik lassen sich **Parameter** zur Charakterisierung von jetzt Wahrscheinlichkeitsverteilungen entwickeln.
- Die wichtigsten Maßzahlen zur Beschreibung von Verteilungen sind **Lage- und Streuungsparameter**.
- Die zentrale Stellung des arithmetischen Mittels in der deskriptiven Statistik nimmt in der Wahrscheinlichkeitstheorie der **Erwartungswert** ein.
- Der mit E symbolisierte Erwartungswert soll zunächst an einem Beispiel erläutert werden.

Beispiel 3.15: Glücksspiel.

Für einen Einsatz von 2€ darf man einen fairen Würfel einmal werfen. Die Hälfte der geworfenen Augenzahl wird (in €) ausgezahlt.

Welchen Gewinn kann ein Spieler erwarten?

- Die Auszahlungen transformieren die diskrete Zufallsvariable $X = \{1, 2, 3, 4, 5, 6\}$ in eine neue Zufallsvariable mit Wertebereich

$$\left\{ \frac{1}{2}, 1, \frac{3}{2}, 2, \frac{5}{2}, 3 \right\}.$$

- Berücksichtigt man die Einsatzkosten von 2€, erhält man als mögliche Gewinne

$$\left\{ -\frac{3}{2}, -1, -\frac{1}{2}, 0, \frac{1}{2}, 1 \right\}.$$

Beispiel 3.16: Glücksspiel.

- Diese Werte sind die Realisationen der neuen Zufallsvariablen

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

- Jeder Ausgang $y_i \in Y$ tritt ein mit Wahrscheinlichkeit

$$P(Y = y_i) = P(X = x_i) = \frac{1}{6}.$$

- Der zu erwartende Gewinn ergibt sich intuitiv, indem jeder mögliche Ausgang mit seiner Eintrittswahrscheinlichkeit gewichtet wird:

$$E(Y) = \left(-\frac{3}{2}\right) \cdot \frac{1}{6} + (-1) \cdot \frac{1}{6} + \left(-\frac{1}{2}\right) \cdot \frac{1}{6} + 0 \cdot \frac{1}{6} + \frac{1}{2} \cdot \frac{1}{6} + 1 \cdot \frac{1}{6} = -0.25.$$

Sei $g(x)$ eine auf $X(\Omega)$ definierte reellwertige Funktion.

Definition 3.17: Erwartungswert einer diskreten Zufallsvariable.

Sei X eine **diskrete** Zufallsvariable mit der Wahrscheinlichkeitsfunktion $f(x_i)$.

Falls $\sum |g(x_i)|f(x_i) < \infty$, ist der **Erwartungswert** von $g(X)$ gegeben als

$$E[g(X)] = \sum_{i=1}^m g(x_i)f(x_i).$$

Definition 3.18: Erwartungswert einer stetigen Zufallsvariable.

Sei X eine **stetige** Zufallsvariable mit Dichtefunktion $f(x)$.

Falls $\int_{-\infty}^{\infty} |g(x)|f(x) dx < \infty$, so ist der **Erwartungswert** von $g(X)$ gegeben als

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x) dx.$$

- Das Integral in der vorangegangenen Definition bezieht sich auf den größtmöglichen **Wertebereich** von X .
- Falls die Dichtefunktion nur über einem Intervall $a \leq X \leq b$ Werte größer als null annimmt, erfolgt die Integration über diesem Intervall.
- Für $g(X) = X^\alpha$, $\alpha \in \mathbb{N}$ heißt der Erwartungswert **Anfangsmoment der Ordnung α** :

$$E(X^\alpha) = \begin{cases} \sum_{i=1}^m x_i^\alpha f(x_i) & X: \text{diskret} \\ \int_{-\infty}^{\infty} x^\alpha f(x) dx & X: \text{stetig.} \end{cases} \quad (3.2)$$

Definition 3.19: Erwartungswert einer Zufallsvariablen ($\alpha = 1$).

Der Erwartungswert einer Zufallsvariablen X ist definiert als

$$E(X) = \mu = \begin{cases} \sum_{i=1}^m x_i f(x_i) & X: \text{diskret} \\ \int_{-\infty}^{\infty} x f(x) dx & X: \text{stetig.} \end{cases}$$

Beispiel 3.20: Erwartungswert - diskreter Fall.

Der Erwartungswert beim Werfen eines fairen Würfels beträgt

$$E(X) = \sum_{i=1}^m x_i f(x_i) = \sum_{i=1}^m x_i p_i = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + \dots + 6 \cdot \frac{1}{6} = 3.5.$$

Man sieht an diesem Ergebnis, dass der Erwartungswert diskreter Zufallsvariablen keine Realisation von X sein muss.

Beispiel 3.21: Roulette.

Beim Roulette kann man auf Zahlen $(0, 1, 2, \dots, 36)$ oder Farben (rot und schwarz) setzen. Wird die Farbe richtig vorhergesagt, ist der Gewinn gleich der Höhe des Einsatzes. Jeweils 18 Zahlen sind rot sowie schwarz (die Null hat keine Farbe).

Setzt man also 10€ auf rot, beträgt der erwartete Gewinn

$$E(X) = 10 \cdot \frac{18}{37} + (-10) \cdot \frac{19}{37} = -0.2703.$$

Beispiel 3.22: Erwartungswert - stetiger Fall.

Als Erwartungswert der stetigen Zufallsvariablen mit $f(x) = \frac{3}{8}x^2$, $0 < x \leq 2$ erhält man

$$E(X) = \int_0^2 xf(x) dx = \int_0^2 x \frac{3}{8}x^2 dx = \int_0^2 \frac{3}{8}x^3 dx = \frac{3}{32}x^4 \Big|_0^2 = \frac{3}{2}. \quad (3.3)$$

- Für den **Erwartungswertoperator** E gelten unabhängig davon, ob X diskret oder stetig ist, folgende Regeln für alle $a, b \in \mathbb{R}$:

$$(a) \quad E(a) = a, \quad (3.4)$$

$$(b) \quad E(a + bX) = a + b E(X), \quad (3.5)$$

$$(c) \quad E \left[\sum_{j=1}^n c_j g_j(X) \right] = \sum_{j=1}^n c_j E[g_j(X)] = c_1 E[g_1(X)] + \dots + c_n E[g_n(X)],$$

falls g_j eine stetige Funktion für alle $j = 1, \dots, n$ ist. Für deren Herleitungen siehe Buch S. 69 ff.

- Der Erwartungswert einer diskreten Zufallsvariablen stimmt formal mit dem **arithmetischen Mittel**

$$\bar{x} = \sum_{i=1}^m x_i h_i$$

überein, wobei h_i die relative Häufigkeit der i -ten Merkmalsausprägung im Datensatz bezeichnet.

- Trotz der formalen Übereinstimmung besteht zwischen beiden Maßzahlen ein erheblicher Unterschied, dessen Beachtung für ein genaues Verständnis beider Konzepte bedeutsam ist.
- Der Erwartungswert $E(X)$ charakterisiert die Verteilung einer Zufallsvariablen gemäß der statistischen Wahrscheinlichkeitsdefinition.
- Dieser Wert bedarf keiner tatsächlichen Durchführungen und ist von diesen unabhängig.

- Das arithmetische Mittel hingegen kennzeichnet den durchschnittlichen Wert eines Datensatzes, d.h. es müssen Beobachtungen vorliegen.
- Wird ein Experiment n -mal durchgeführt, so liegen die **Realisationen** x_1, \dots, x_n von X vor, für die das arithmetische Mittel

$$\bar{x} = \frac{1}{n} \sum_{j=1}^n x_j$$

berechnet werden kann.

- Das arithmetische Mittel kennzeichnet jetzt den Datensatz, nicht jedoch $E(X)$.
- Erhöht man n , wird sich \bar{x} $E(X)$ annähern. Dazu später mehr.

- **Varianz** bzw. **Standardabweichung** quantifizieren die **Streuung** einer Verteilung.
- Auch die Varianz ergibt sich für eine spezielle Funktion $g(X)$ aus den Definitionen für den Erwartungswert. Für $g(X) = (X - \mu)^\alpha$, $\alpha \in \mathbb{N}$, heißt der Erwartungswert

$$E[(X - \mu)^\alpha] = \begin{cases} \sum_{i=1}^m (x_i - \mu)^\alpha f(x_i) & X: \text{diskret} \\ \int_{-\infty}^{\infty} (x - \mu)^\alpha f(x) dx & X: \text{stetig} \end{cases}$$

Zentralmoment der Ordnung α .

- Für $\alpha = 2$ resultiert die Varianz als Zentralmoment zweiter Ordnung:

$$\text{Var}(X) = E[(X - \mu)^2] = \sigma^2.$$

Definition 3.23: Standardabweichung.

Die **Standardabweichung** σ ist definiert als $\sigma = \sqrt{\sigma^2}$.

Definition 3.24: Varianz: Spezieller Verschiebungssatz.

Die Varianz kann durch Anfangsmomente dargestellt werden:

$$\sigma^2 = E[(X - \mu)^2] = E(X^2) - 2\mu E(X) + E(\mu^2).$$

Die rechte Seite der Gleichung lässt sich zusammenfassen zu

$$\sigma^2 = E(X^2) - [E(X)]^2 = E(X^2) - \mu^2. \quad (3.6)$$

$E(X^2)$ ist das Anfangsmoment zweiter Ordnung, $[E(X)]^2$ das Quadrat des Anfangsmoments erster Ordnung.

Beispiel 3.25: Varianz - diskreter Fall.

Die Varianz beim Werfen eines fairen Würfels beträgt

$$\sigma^2 = E(X^2) - [E(X)]^2 = \sum_{i=1}^m x_i^2 f(x_i) - \mu^2 = 1^2 \cdot \frac{1}{6} + \dots + 6^2 \cdot \frac{1}{6} - 3.5^2 = 2.9167.$$

Standardabweichung: $\sigma = \sqrt{\sigma^2} = 1.7078$.

Beispiel 3.26: Roulette.

Setzt man also 10€ auf rot, beträgt die Varianz

$$\sigma^2 = E(X^2) - [E(X)]^2 = 10^2 \cdot \frac{18}{37} + (-10)^2 \cdot \frac{19}{37} - (-0.2703)^2 = 99.927.$$

Standardabweichung: $\sigma = \sqrt{\sigma^2} = 9.9963$.

Beispiel 3.27: Varianz - stetiger Fall.

Als Varianz der stetigen Zufallsvariablen mit $f(x) = \frac{3}{8}x^2$, $0 < x \leq 2$ erhält man

$$\begin{aligned}\sigma^2 = E(X^2) - [E(X)]^2 &= \int_0^2 x^2 f(x) dx - \left[\int_0^2 x f(x) dx \right]^2 \\ &= \int_0^2 x^2 \frac{3}{8} x^2 dx - \left[\int_0^2 x \frac{3}{8} x^2 dx \right]^2\end{aligned}\quad (3.7)$$

$$\stackrel{(3.3)}{=} \int_0^2 \frac{3}{8} x^4 dx - \left(\frac{3}{2} \right)^2\quad (3.8)$$

$$= \frac{3}{40} x^5 \Big|_0^2 - \frac{9}{4} = \frac{12}{5} - \frac{9}{4} = \frac{3}{20}.\quad (3.9)$$

Für die Varianz gelten die folgenden Regeln:

$$\text{Var}(a) = 0 \quad (3.10)$$

$$\text{Var}(a + bX) = b^2 \text{Var}(X) \quad (3.11)$$

für Konstanten $a, b \in \mathbb{R}$.

Ergebnis (3.11) weisen wir in (3.16) nach.

Zufallsvariablen 3



Aufgabe 3.28:

Berechnen Sie (a) den Erwartungswert und (b) die Varianz bzw. die Standardabweichung für die diskrete Zufallsvariable X : *Anzahl der Autoverkäufe pro Tag*

X	0	1	2	3	4	5
$f(x)$	0.18	0.39	0.24	0.14	0.04	0.01
$F(x)$	0.18	0.57	0.81	0.95	0.99	1

Aufgabe 3.29:

Berechnen Sie (a) den Erwartungswert und (b) die Varianz bzw. die Standardabweichung für die stetige Variable X mit der Dichtefunktion

$$f(x) = \begin{cases} -\frac{1}{72}x^2 + \frac{1}{12}x + \frac{1}{12} & 0 \leq x \leq 6 \\ 0 & \text{sonst.} \end{cases}$$

Aufgabe 3.30:

Für die diskrete Zufallsvariable X ist folgende Wahrscheinlichkeitsfunktion gegeben

$$f(x) = \begin{cases} 0.1 & x = -1 \\ 0.3 & x = 0 \\ 0.6 & x = 1 \\ 0 & \text{sonst.} \end{cases}$$

- (i) Berechnen Sie (a) den Erwartungswert und (b) die Varianz bzw. die Standardabweichung für X !
- (ii) Berechnen Sie (a) den Erwartungswert und (b) die Varianz bzw. die Standardabweichung für $Y = 2X^2 + 1$!

Mehrere Zufallsvariablen, die entweder alle diskret oder alle stetig sind, können zu einer neuen Zufallsvariablen verknüpft werden.

Definition 3.31: Linearkombination.

Gegeben seien die Zufallsvariablen X_1, X_2, \dots, X_n und Konstanten

$$\lambda_1, \lambda_2, \dots, \lambda_n.$$

Die Zufallsvariable

$$Y = \lambda_1 X_1 + \lambda_2 X_2 + \dots + \lambda_n X_n = \sum_{j=1}^n \lambda_j X_j$$

heißt **Linearkombination** der Zufallsvariablen $X_j, j = 1, \dots, n$.

Besonders wichtige Linearkombinationen sind die folgenden:

- (1) Sind alle $\lambda_j = 1$, erhält man die **Summe** der n Zufallsvariablen X_j ,

$$\tilde{S} = X_1 + X_2 + \dots + X_n = \sum_{j=1}^n X_j.$$

- (2) Für $\lambda_j = 1/n$ resultiert das **arithmetische Mittel** der X_j ,

$$\bar{X} = \frac{1}{n}X_1 + \frac{1}{n}X_2 + \dots + \frac{1}{n}X_n = \frac{1}{n} \sum_{j=1}^n X_j = \frac{1}{n}\tilde{S}.$$

- (3) $\lambda_j = 0$ für $j = 3, \dots, n$ und $X_2 = 1$ liefert eine **Lineartransformation**

$$Y = \lambda_1 X_1 + \lambda_2,$$

da Y eine lineare Funktion von X_1 ist.

- Bei der Varianzberechnung muss unterschieden werden, ob die Zufallsvariablen X_j paarweise stochastisch unabhängig sind oder nicht.
- Es sei hier zunächst **paarweise stochastische Unabhängigkeit** angenommen.

Fall (1):

$$\begin{aligned} E(\tilde{S}) &= E(X_1 + \dots + X_n) = E(X_1) + \dots + E(X_n) \\ &= \sum_{j=1}^n E(X_j) = \sum_{j=1}^n \mu_j =: \mu_{\tilde{S}}. \end{aligned} \tag{3.12}$$

Die vorletzte Umformung zeigt, dass die Operatoren E und \sum in ihrer Anwendung vertauscht werden können.

$$\begin{aligned} \text{Var}(\tilde{S}) &= \text{Var}(X_1 + \dots + X_n) = \sigma_1^2 + \dots + \sigma_n^2 \\ &= \sum_{j=1}^n \sigma_j^2 =: \sigma_{\tilde{S}}^2. \end{aligned} \tag{3.13}$$

Diese Berechnung heißt **Additionssatz für Varianzen (paarweise) stochastisch unabhängiger Zufallsvariablen**.

Fall (2):

$$E(\bar{X}) = E\left(\frac{1}{n} \sum_{j=1}^n X_j\right) \stackrel{(3.5)}{=} \frac{1}{n} E(\tilde{S}) = \frac{1}{n} \sum_{j=1}^n \mu_j =: \mu_{\bar{X}} \quad (3.14)$$

sowie

$$\begin{aligned} \text{Var}(\bar{X}) &= \text{Var}\left(\frac{1}{n} \tilde{S}\right) \\ &\stackrel{(3.11)}{=} \frac{1}{n^2} \text{Var}(\tilde{S}) \\ &\stackrel{(3.13)}{=} \frac{1}{n^2} \sum_{j=1}^n \sigma_j^2 \\ &=: \sigma_{\bar{X}}^2. \end{aligned}$$

Fall (3):

$$E(Y) = E(\lambda_1 X_1 + \lambda_2) \stackrel{(3.5)}{=} \lambda_1 E(X_1) + \lambda_2 = \lambda_1 \mu_1 + \lambda_2 = \mu_Y \quad (3.15)$$

sowie

$$\begin{aligned} \text{Var}(Y) &\stackrel{(3.6)}{=} E[(Y - \mu_Y)^2] \\ &\stackrel{(3.15)}{=} E[(\lambda_1 X_1 + \lambda_2 - \lambda_1 \mu_1 - \lambda_2)^2] \\ &= E\{[\lambda_1(X_1 - \mu_1)]^2\} \\ &\stackrel{(3.5)}{=} \lambda_1^2 E[(X_1 - \mu_1)^2] \\ &= \lambda_1^2 \text{Var}(X_1) = \lambda_1^2 \sigma_1^2 = \sigma_Y^2. \end{aligned} \quad (3.16)$$

Damit ist (3.11) gezeigt.

- Für die Lineartransformation

$$Y = \lambda_1 X_1 + \lambda_2 \quad \text{mit} \quad \lambda_1 = \frac{1}{\sigma_1}, \lambda_2 = -\frac{\mu_1}{\sigma_1} \quad (3.17)$$

folgt

$$E(Y) = \frac{1}{\sigma_1} \mu_1 - \frac{\mu_1}{\sigma_1} = 0$$

und

$$\text{Var}(Y) = \frac{1}{\sigma_1^2} \sigma_1^2 = 1.$$

Y heißt dann **standardisierte** Zufallsvariable.

- Alle Zufallsvariablen mit endlichem Erwartungswert und einer endlichen Varianz größer als null lassen sich so standardisieren.
- Haben die Zufallsvariablen $X_j, j = 1, 2, \dots, n$ im Fall (1) alle dieselbe Verteilung und folgt diese dann auch für die Verteilung ihrer Summe $\tilde{S} = \sum_{j=1}^n X_j$, liegt die sogenannte **Reproduktionseigenschaft** bzw. **Reproduktivität** vor.

Aufgabe 3.32:

Ein Investor legt ein Viertel seines Vermögens in einem Immobilienfonds (IF) und den Rest in einem Aktienfonds (AF) an. Für diese gilt

$$\begin{aligned} E(\text{IF}) &= 0.28 & \sigma_{\text{IF}} &= 0.2 \\ E(\text{AF}) &= 0.12 & \sigma_{\text{AF}} &= 0.06. \end{aligned}$$

- (a) Wie groß ist die erwartete Gesamtrendite, wenn die Renditen beider Investitionsprojekte unabhängig voneinander sind?
- (b) Wie groß ist die Varianz der Gesamtrendite, wenn die Renditen beider Investitionsprojekte unabhängig voneinander sind?

- Sind nur μ und $\sigma^2 < \infty$ von X bekannt, nicht jedoch ihre konkrete Verteilungsfunktion, lassen sich die Wahrscheinlichkeiten für komplementäre Ereignisse der Art

$$\{|X - \mu| < c\} = \{\mu - c < X < \mu + c\}$$

und

$$\{|X - \mu| \geq c\} = \{X \leq \mu - c, X \geq \mu + c\}$$

für jede beliebige Konstante $c > 0$ **abschätzen**.

- So gewinnt man zusätzlich zu μ und σ^2 weitere Informationen über Lage und Streuung der Verteilung von X .
- Die Abschätzung geschieht mit der **Tschebyscheff'schen Ungleichung**, die in zwei äquivalenten Versionen angegeben werden kann.

- Es lässt sich zeigen, dass für jedes X mit $\sigma^2 < \infty$ gilt

$$P(|X - \mu| \geq c) \leq \frac{\sigma^2}{c^2} \text{ bzw. } P(|X - \mu| < c) > 1 - \frac{\sigma^2}{c^2} \quad (3.18)$$

für alle $c > 0$.

- Setzt man $c = k\sigma$, folgt die **Tschebyscheff'sche Ungleichung**:

$$(a) P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2} \quad \text{bzw.} \quad (b) P(|X - \mu| < k\sigma) > 1 - \frac{1}{k^2}. \quad (3.19)$$

- Nur für $c > \sigma$ bzw. $k > 1$ liegen informative Abschätzungen vor.
- Die Abschätzung der Wahrscheinlichkeit liefert eine untere bzw. obere Grenze für die exakte Wahrscheinlichkeit.

Der Approximationsfehler kann jedoch beträchtlich sein.

Beispiel 3.33:

Die stetige Zufallsvariable X mit der Dichtefunktion $f(x) = \frac{3}{8}x^2$, $0 < x \leq 2$, sonst null, hat einen Erwartungswert von $\mu \stackrel{(3.3)}{=} \frac{3}{2}$ und eine Varianz von $\sigma^2 \stackrel{(3.9)}{=} \frac{3}{20}$ bzw. $\sigma \approx 0.3873$.

Die exakte Wahrscheinlichkeit für das Ereignis $1 \leq X \leq 2$ beträgt

$$P(1 \leq X \leq 2) = \int_1^2 \frac{3}{8}x^2 dx = F(2) - F(1) = \frac{7}{8} = 0.875.$$

Beispiel 3.33: Fortsetzung.

Zur Abschätzung müssen c und k bestimmt werden. Aus

$$1 \leq X \leq 2 \Leftrightarrow \left| X - \frac{3}{2} \right| \leq \frac{1}{2},$$

folgt $c = \frac{1}{2}$. k ergibt sich wegen $k\sigma = c$ durch Lösen der Gleichung

$$0.3873k = \frac{1}{2}$$

als $k = 1.291$.

Damit kann die Abschätzung berechnet werden:

$$P\left(\left| X - \frac{3}{2} \right| \leq \frac{1}{2}\right) > 1 - \frac{1}{(1.291)^2} \approx 0.4.$$

Aufgabe 3.34:

Für den *Fernsehkonsument pro Tag in den USA* ist bekannt: $\mu = 4.4\text{h}$ und $\sigma = 2.1\text{h}$.

Wie groß ist die Wahrscheinlichkeit, dass eine Person zwischen 0.2 und 8.6h fernsieht?

- 1 Einleitung
- 2 Grundlagen
- 3 Eindimensionale Zufallsvariablen
- 4 Ausgewählte theoretische Verteilungen**
- 5 Grundzüge der Stichprobentheorie
- 6 Statistische Schätzverfahren
- 7 Statistische Testverfahren
- 8 Zweidimensionale Zufallsvariablen

- **Zufallsvariablen** und **Wahrscheinlichkeitsverteilungen** abstrahieren von tatsächlichen, in der Realität ablaufenden Zufallsvorgängen.
- Eine wesentliche Aufgabe der Statistik liegt darin, aus **theoretischen Verteilungen** die zu finden, die einen realen Zufallsvorgang am besten erfasst.
- Da viele ökonomische Zufallsvorgänge auf gleichen Voraussetzungen basieren, reichen einige ausgewählte Verteilungsfunktionen zur Lösung der meisten praktischen Probleme aus.

Bei diskreten Zufallsvariablen lassen sich theoretische Verteilungsfunktionen oft über ein **Bernoulli-Experiment** entwickeln. Dies liegt vor, wenn ein Zufallsvorgang

1. nur in zwei Ausgänge A und \bar{A} mündet, deren
 2. Wahrscheinlichkeiten über Durchführungen konstant bleiben und
 3. die einzelnen Durchführungen unabhängig voneinander sind.
- Obwohl die dritte Bedingung suggeriert, dass bei einem Bernoulli-Experiment mehrere Durchführungen vorliegen müssen, bezeichnet man so auch die einmalige Durchführung.
 - Bei Wiederholungen gibt man deren Anzahl an und nennt sie auch **Bernoulli-Kette der Länge n** .

- Die Ableitung theoretischer stetiger Verteilungen aus einfachen Annahmen eines Zufallsexperiments wie im diskreten Fall gelingt nur selten.
- Jedoch lassen sich aus diskreten Verteilungen nach Grenzübergang für die Anzahl der Durchführungen **stetige Verteilungen** gewinnen.
- Auch eignen sich in vielen empirischen Situationen bestimmte theoretische stetige Verteilungen als **Approximation**.

- Wenn X die Werte $x_1 = 1$ und $x_2 = 0$ annimmt, liegt eine **Nulleins-** oder **Bernoulli-Verteilung** vor.
- Die **Bernoulliverteilung** ist der wichtigste Spezialfall der Zweipunktverteilung (vgl. Buch auf S. 81).
- Die letzte Bezeichnung unterstreicht, dass X auf einem Bernoulli-Experiment basiert. Die Wahrscheinlichkeitsfunktion ist

$$\begin{aligned} f(x) &= \begin{cases} p & x = 1 \\ 1 - p & x = 0 \\ 0 & \text{sonst,} \end{cases} \\ &= p^x(1 - p)^{1-x} \quad \text{für } x = 0, 1, \quad \text{sonst null.} \end{aligned}$$

Erwartungswert und Varianz lauten jetzt

$$\mu = \sum_{i=1}^2 x_i p_i = 1 \cdot p + 0 \cdot (1 - p) = p$$

sowie

$$\begin{aligned}\sigma^2 &\stackrel{(3.6)}{=} E(X^2) - [E(X)]^2 \\ &\stackrel{(3.2)}{=} \sum_{i=1}^2 x_i^2 p_i - p^2 \\ &= 1^2 p + 0^2 (1 - p) - p^2 \\ &= p - p^2 \\ &= p(1 - p).\end{aligned}\tag{4.1}$$

```
# Stichprobe der Größe 1  
rbinom(n = 1, size = 1, prob = 0.5)
```

```
## [1] 0
```

```
# Stichprobe der Größe 100000  
x <- rbinom(n = 100000, size = 1, prob = 0.5)  
mean(x) # Mittelwert, nicht Erwartungswert!
```

```
## [1] 0.4992
```

```
var(x)
```

```
## [1] 0.25
```

Siehe Buch, S. 82.

Viele praktische Zufallsvorgänge besitzen die Struktur eines Bernoulli-Experiments, d.h. sie münden in nur zwei Ausgänge A und \bar{A} und sind beliebig oft unter (mehr oder weniger) gleichen Bedingungen wiederholbar. Beispiele:

- (a) Glücksspiel mit dem Ausgang A als Gewinn,
- (b) Geburten, wobei A die Geburt eines Mädchens anzeigt,
- (c) Überprüfung einer Anzeigekampagne mit A : Kenntnis der Aktion,
- (d) Qualitätskontrolle bei der Serienproduktion mit A : *defektes Produkt*.

- Diese Vorgänge können als *Ziehen mit Zurücklegen aus einer Urne* idealisiert werden, wobei die Kugeln gemäß A oder \bar{A} unterteilt sind.
- Die dafür definierte Zufallsvariable X_j ist dann pro Durchführung j , $j = 1, \dots, n$, bernoulliverteilt.
- Hierbei bildet X_j A in die 1 und das Komplementärereignis \bar{A} in die 0 ab:

$$X_j(\Omega) = \{0, 1\} \quad \text{mit} \quad X_j(A) = 1 \quad \text{und} \quad X_j(\bar{A}) = 0.$$

- Wegen der beliebigen Wiederholbarkeit gilt für jedes j

$$P(X_j = 0) = 1 - p = q \quad \text{und} \quad P(X_j = 1) = p.$$

- Bei n -maliger Durchführung eines Bernoulli-Experiments stellt $X = X_1 + X_2 + \dots + X_n$ die Anzahl der Eintritte von A in einer **Bernoulli-Kette** der Länge n dar.
- Der Wertebereich für X entspricht den möglichen Summenwerten:

$$X = \{x | x = 0, 1, 2, \dots, n\}.$$

- Das Ereignis $X = x$ tritt ein, wenn x der Zufallsvariablen X_1, \dots, X_n eins und die übrigen $n - x$ Variablen null sind, z.B. die Folge

$$\underbrace{1\ 1\ \dots\ 1}_{x\text{-mal}}\ \underbrace{0\ 0\ \dots\ 0}_{(n-x)\text{-mal}}$$

- Da die Durchführungen und damit die X_j unabhängig sind, ist die Wahrscheinlichkeit für das Eintreten dieser Kette

$$p^x(1-p)^{n-x}.$$

- Das Ereignis $X = x$ tritt aber auch bei jeder Permutation dieser Kette ein, also insgesamt $\binom{n}{x}$ -mal (s. auch (2.3)).
- Die Wahrscheinlichkeitsfunktion für X lautet daher

$$f(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & x = 0, 1, \dots, n \\ 0 & \text{sonst.} \end{cases}$$

- Eine solche Zufallsvariable heißt **binomialverteilt**, symbolisiert als $B(n, p)$.
- Die Binomialverteilung hängt also von n und p ab. Hier ein Video zur Geschichte der Binomialverteilung.

Die Verteilungsfunktion (3.1), $F(x)$, ist

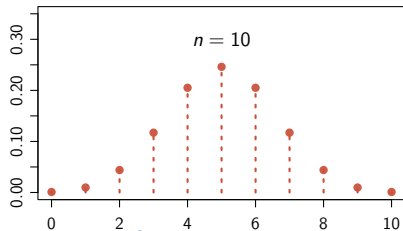
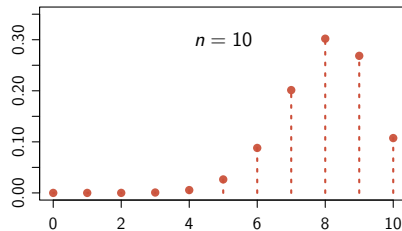
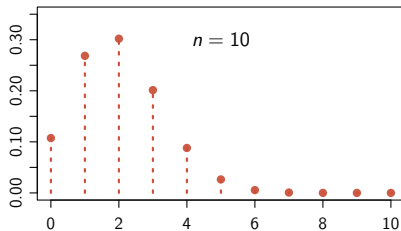
$$F(x) = \begin{cases} 0 & x < 0 \\ \sum_{k=0}^{\lfloor x \rfloor} \binom{n}{k} p^k (1-p)^{n-k} & 0 \leq x < n, \quad k = 0, 1, \dots, \lfloor x \rfloor \\ 1 & x \geq n. \end{cases}$$

Hier stellt $\lfloor x \rfloor$ den ganzzahligen Teil von x dar, z.B. $\lfloor 3.81 \rfloor = \text{int}(3.81) = 3$.

- Erst die Festlegung von n und p spezifiziert eine konkrete Verteilung.
- Für $n = 1$ resultiert als Spezialfall die Zweipunktverteilung.
- Die nachfolgende Abbildung zeigt den Effekt, den p bei konstantem n (hier $n = 10$) auf $f(x)$ ausübt.
 - ▶ Für $0 < p < 0.5$ ist $f(x)$ linkssteil.
 - ▶ Für $0.5 < p < 1$ ist $f(x)$ rechtssteil.
 - ▶ Für $p = 0.5$ ist $f(x)$ symmetrisch.

Binomialverteilung

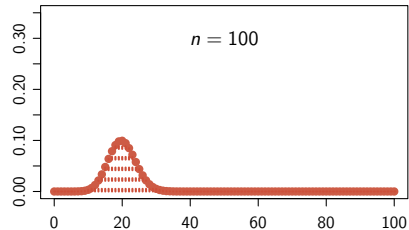
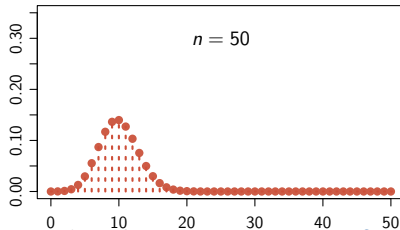
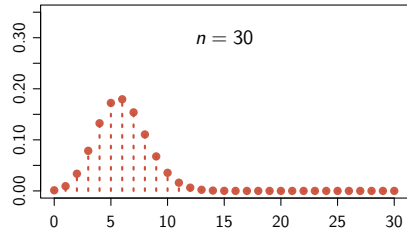
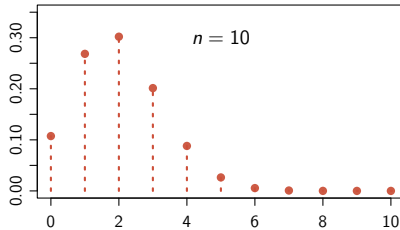
Variation von p bei Konstanz von n



- Den Effekt von n bei konstantem p verdeutlicht die folgende Abbildung.
- In allen vier Fällen ist $p = 0.2$, die Anzahl der Durchführungen ist $n = 10$, $n = 30$, $n = 50$ und $n = 100$.
- Die Symmetrie wird mit zunehmendem n ausgeprägter.

Binomialverteilung

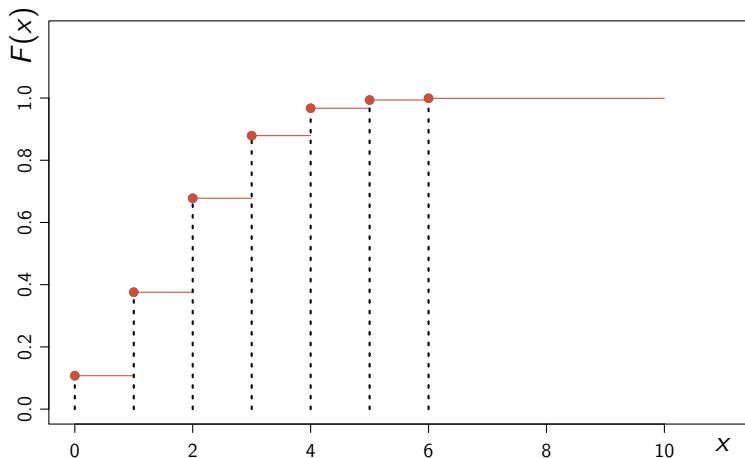
Wahrscheinlichkeitsfunktion



Binomialverteilung

Verteilungsfunktion für $p = 0.2$ und $n = 10$

$F(x)$ macht bei $x = 2$, wo $f(x)$ ein Maximum besitzt, den größten Sprung.



- Da X als $X = \sum_{j=1}^n X_j$ definiert ist und jedes X_j einer Bernoulli-Verteilung mit dem Erwartungswert $E(X_j) = p$ folgt, gilt

$$\mu = E\left(\sum_{j=1}^n X_j\right) \stackrel{(3.12)}{=} \sum_{j=1}^n E(X_j) = \sum_{j=1}^n p = np. \quad (4.2)$$

- Die Varianz von X lässt sich mit dem **Additionssatz für Varianzen stochastisch unabhängiger Zufallsvariablen** (3.13) ermitteln.
- Da jedes X_j , $j = 1, \dots, n$ dieselbe Varianz $\text{Var}(X_j) = pq$ besitzt, folgt für die Summe X

$$\sigma^2 = \text{Var}\left(\sum_{j=1}^n X_j\right) = \sum_{j=1}^n \text{Var}(X_j) = \sum_{j=1}^n pq \stackrel{(4.1)}{=} npq. \quad (4.3)$$

- Der Einfluss der **Scharparameter** n und p kann an den Gleichungen

$$\mu = np \quad \text{und} \quad \sigma^2 = npq$$

erkannt werden.

- Erhöht man n bei konstantem p , steigen μ und σ^2 ; für die drei Verteilungen in der Abbildung erhält man $\mu_1 = 2$, $\sigma_1^2 = 1.6$, $\mu_2 = 6$, $\sigma_2^2 = 4.8$ und $\mu_3 = 10$, $\sigma_3^2 = 8$.
- Eine größeres p bei konstantem n lässt μ wachsen, während σ^2 nur bis $p = 0.5$ steigt, dann wieder fällt.

Verteilungen 1



Beispiel 4.1: Glücksrad.

Bei einem Glücksspiel betrage die Wahrscheinlichkeit zu gewinnen $p = 0.2$.

Die Binomialverteilung $B(5, 0.2)$ liefert die Wahrscheinlichkeit, dass bei 5 Teilnahmen mindestens einmal gewonnen wird, wenn die Spiele unabhängig sind. $P(X \geq 1)$ berechnet sich nach dem Additionssatz für disjunkte Ereignisse als

$$\begin{aligned} P(X \geq 1) &= f(x = 1) + \dots + f(x = 5) \\ &= \binom{5}{1} \cdot 0.2 \cdot 0.8^4 + \dots + \binom{5}{5} \cdot 0.2^5 \cdot 0.8^0 \\ &= 0.4096 + 0.2048 + 0.0512 + 0.0064 + 0.0003 = 0.6723. \end{aligned}$$

Beispiel 4.2: Glücksrad.

Schneller findet man die gesuchte Wahrscheinlichkeit über das Komplementärereignis – kein Gewinn –, das zu $X = 0$ führt. Aus

$$P(X = 0) + P(X \geq 1) = 1$$

folgt

$$P(X \geq 1) = 1 - P(X = 0) = 1 - \binom{5}{0} \cdot 0.2^0 \cdot 0.8^5 = 1 - 0.3277 = 0.6723.$$

$F(x)$ liefert die gesuchte Wahrscheinlichkeit $P(X \leq x)$ direkt.

```
1 - pbinom(q = 0, size = 5, prob = 0.2)
```

```
## [1] 0.6723
```

```
pbinom(q = 0, size = 5, prob = 0.2, lower.tail = F)
```

```
## [1] 0.6723
```

```
x <- rbinom(n = 10000, size = 5, prob = 0.2)
head(x, n = 20)
```

```
## [1] 2 3 1 0 1 1 1 1 0 1 2 2 1 1 2 1 0 0 0 0
```

```
mean(x) # Stichprobenmittelwert
```

```
## [1] 1.007
```

```
var(x) # Stichprobenvarianz
```

```
## [1] 0.7963
```

```
dbinom(x = 0:5, size = 5, prob = 0.2) # Wahrscheinlichkeitsfunktion
```

```
## [1] 0.32768 0.40960 0.20480 0.05120 0.00640 0.00032
```

```
pbinom(q = 0:5, size = 5, prob = 0.2) # Verteilungsfunktion
```

```
## [1] 0.3277 0.7373 0.9421 0.9933 0.9997 1.0000
```

```
sum(x == 2) / 10000 # relative Häufigkeit der Realisationen gleich zwei
```

```
## [1] 0.2074
```


Beispiel 4.3: Produktion.

Bei einer Fertigung seien 5% der Produkte fehlerhaft. Zur Qualitätsprüfung werden 5 Produkte entnommen. Gesucht sind die Wahrscheinlichkeiten für genau ein bzw. genau zwei defekte Produkte.

$$\begin{aligned}P(X = 1) = f(x = 1) &= \binom{5}{1} \cdot 0.05^1 \cdot (1 - 0.05)^{5-1} \\&= \frac{5!}{1!(5-1)!} \cdot 0.05 \cdot 0.95^4 = 0.2036.\end{aligned}$$

$$P(X = 2) = f(x = 2) = \binom{5}{2} \cdot 0.05^2 \cdot (1 - 0.05)^{5-2} = 0.0214.$$

```
dbinom(x = 1:2, size = 5, prob = 0.05)
```

```
## [1] 0.20363 0.02143
```

- Die Summe $\tilde{S} = X + Y$ zweier unabhängiger und binomialverteilter Zufallsvariablen X und Y mit demselben p ist ebenfalls binomialverteilt.
- Aus $X: B(m, p)$ und $Y: B(n, p)$ folgt also $\tilde{S}: B(m + n, p)$ – die Binomialverteilung ist **reproduktiv**.
- Denn: Da X die Summe von m unabhängigen, zweipunktverteilten Zufallsvariablen und Y die Summe von n unabhängigen, zweipunktverteilten Zufallsvariablen ist, ist \tilde{S} die Summe von $m + n$ unabhängigen zweipunktverteilten Zufallsvariablen.
- Die Wahrscheinlichkeitsfunktion $f(s)$ für $s = 0, 1, 2, \dots, m + n$ ist daher analog zu $f(x)$

$$f(s) = \binom{m+n}{s} p^s q^{m+n-s},$$

null sonst.

- Die Binomialverteilung liefert die Wahrscheinlichkeit für x -mal *Erfolg* in einer Bernoulli-Kette mit Länge n .
- Manchmal interessiert aber die Wahrscheinlichkeit, dass eine Bernoulli-Kette die Länge n erreicht, bevor zum ersten Mal *Erfolg* eintritt.
- X ist nun definiert als Anzahl der Fehlversuche bis zum ersten Erfolg.
- Bei $X = 0$ tritt der Erfolg bereits bei der ersten Durchführung ein, bei $X = x$ bei der $(x + 1)$ -ten Wiederholung.

- Die Wahrscheinlichkeiten für diese Ereignisse liefert der Multiplikationssatz für unabhängige Ereignisse (2.1).
- Ist die Erfolgswahrscheinlichkeit wieder p (und für *Misserfolg* $q = 1 - p$), ergibt sich

$$P(X = 0) = p, P(X = 1) = pq, P(X = 2) = pq^2, \text{ usw.}$$

- Verallgemeinert erhält man die Wahrscheinlichkeitsfunktion $f(x)$ als

$$f(x) = \begin{cases} pq^x & x = 0, 1, 2, \dots \\ 0 & \text{sonst.} \end{cases}$$

- Eine solche Zufallsvariable heißt **geometrisch verteilt**; die Verteilung hängt nur von p ab. Kurzform: $G(p)$.
- Für $f(x)$ gilt mit der Summenformel für eine geometrische Reihe

$$\begin{aligned}\sum_{x=0}^{\infty} f(x) &= p + pq + pq^2 + \dots \\ &= p(1 + q + q^2 + \dots) \\ &= p \sum_{x=0}^{\infty} q^x = p \frac{1}{1-q} = p \frac{1}{p} = 1.\end{aligned}$$

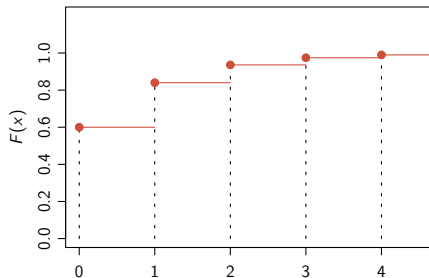
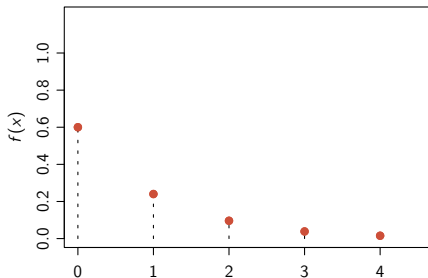
- Für $x = 0, 1, 2, \dots$ lautet die Verteilungsfunktion

$$F(x) = \sum_{k=0}^x pq^k.$$

Das Buch auf S. 94-95 zeigt, dass $F(x)$ geschrieben werden kann als

$$F(x) = \begin{cases} 0 & x < 0 \\ 1 - q^{[x]+1} & x \geq 0 \\ 1 & x = \infty. \end{cases}$$

Beispiele für $p = 0.6$:



- Je größer p , desto schneller strebt $f(x)$ gegen null und $F(x)$ gegen eins.
- q^{x+1} ist die Komplementärwahrscheinlichkeit zu $F(x)$.
- Da $F(x) = P(X \leq x)$, ist $P(X > x) = q^{x+1}$.
- Damit $X > x$ überhaupt eintreten kann, müssen zunächst genau $x + 1$ Misserfolge vorliegen; die Anzahl nachfolgender Fehlversuche ist dann unerheblich.

Beispiel 4.4: Investition.

Bei einem Investitionsobjekt beträgt die Wahrscheinlichkeit, Gewinn pro Periode zu erzielen: $p = 0.7$.

Wenn die Periodengewinne unabhängig sind, liefert die geometrische Verteilung die Wahrscheinlichkeit für den ersten Gewinn in der vierten Periode.

X nimmt den Wert $x = 3$ an, $F(x)$ ist an der Stelle $x = 3$

$$F(3) = 1 - 0.3^4 = 0.9919.$$

$F(3) = P(X \leq 3)$ ist die Wahrscheinlichkeit dafür, dass es höchstens drei Perioden dauert, bis in der vierten Periode der erste Gewinn realisiert wird. Man bezeichnet $G(p)$ wegen dieser Interpretation auch als **Wartezeitverteilung**.

Beispiel 4.5: Investition.

Die Wahrscheinlichkeit der geometrischen Verteilung $f(x = 3) = 0.7 \cdot 0.3^3 = 0.0189$ darf nicht mit der Wahrscheinlichkeit der Binomialverteilung verwechselt werden, die man für $n = 4$ und $x = 1$ (Erfolg) erhält. Diese beträgt

$$f(x = 1) = \binom{4}{1} 0.7 \cdot 0.3^3 = 0.0756.$$

Sie gibt die Wahrscheinlichkeit einer Bernoulli-Kette der Länge $n = 4$ mit einem *Erfolg* an. Da dieser nicht erst bei der vierten Durchführung eintreten muss, ist die Wahrscheinlichkeit notwendigerweise größer als bei der geometrischen Verteilung.

Beispiel 4.5: Fortsetzung.

```
# Verteilungsfunktion der geometrischen Verteilung
```

```
pgeom(q = 3, prob = 0.7)
```

```
## [1] 0.9919
```

```
# Alternativ Berechnung über die Dichtefunktion
```

```
sum(dgeom(x = 0:3, prob = 0.7))
```

```
## [1] 0.9919
```

```
# Zufallszahlen aus der Verteilung ziehen
```

```
rgeom(n = 100, prob = 0.7)
```

```
## [1] 0 0 0 0 0 0 0 0 1 0 0 1 1 0 0 0 0 0 1 0 1 1 1 0 0 0 0 4 1 0 0 0 0 1 1 0 0
```

```
## [38] 0 1 2 0 2 0 0 1 2 0 0 0 0 2 0 0 0 0 2 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 2 1 0 0 0
```

```
## [75] 0 1 1 0 0 0 2 0 0 2 0 0 1 0 0 0 0 0 0 0 0 0 2 0 0 0 0
```

```
# 3 tritt sehr selten auf, 4 so gut wie nie
```

Erwartungswert und Varianz der geometrischen Verteilung lauten

$$\mu = \frac{q}{p}$$

bzw.

$$\sigma^2 = \frac{q}{p^2}.$$

Eine ausführliche Herleitung der beiden Parameter findet sich im Anhang.

Verteilungen 2



- Die Problemstellung der geometrischen Verteilung lässt sich verallgemeinern: Wie oft ist ein Bernoulli-Experiment durchzuführen, bis genau r Erfolge eingetreten sind?
- X ist jetzt definiert als die Anzahl der Misserfolge bis zum r -ten Erfolg.
- Damit r Erfolge eintreten können, muss das Bernoulli-Experiment mindestens r -mal durchgeführt werden.
- $X = 0$ bedeutet, dass alle r Durchführungen Erfolge sind; $X = 1$ bedeutet, dass sich der r -te Erfolg in der $(r + 1)$ -ten Durchführung einstellt.
- Allgemein bedeutet $X = x$, dass der r -te Erfolg bei der $(x + r)$ -ten Ausführung vorliegt.
- Bevor der r -te Erfolg eintritt, müssen in $x + r - 1$ Durchführungen genau $r - 1$ Erfolge realisiert worden sein; in der letzten, der $(x + r)$ -ten Wiederholung muss *Erfolg* eintreten.

- Die Wahrscheinlichkeit für $r - 1$ Erfolge bei $x + r - 1$ Durchführungen erhält man mit der Binomialverteilung (4.2) als

$$\binom{x + r - 1}{r - 1} p^{r-1} q^x.$$

- Die Erfolgswahrscheinlichkeit im $(x + r)$ -ten Durchgang beträgt wie bei allen anderen Durchführungen p .
- Die Wahrscheinlichkeit, dass der r -te Erfolg genau bei der $(r + x)$ -ten Durchführung eintritt, ist wegen Unabhängigkeit der Experimente das Produkt beider Wahrscheinlichkeiten.
- Damit lautet

$$f(x) = \begin{cases} \binom{x + r - 1}{r - 1} p^r q^x & x = 0, 1, 2, \dots \\ 0 & \text{sonst.} \end{cases}$$

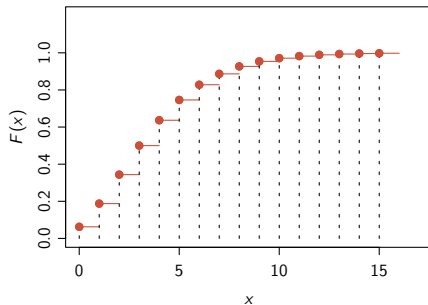
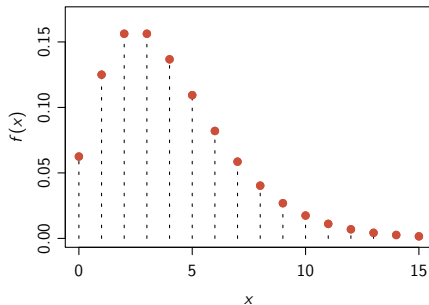
- Eine solche Zufallsvariable heißt **negativ binomialverteilt** oder auch **Pascal-verteilt**.
- Die negative Binomialverteilung hängt von den Parametern r und p ab; symbolisiere sie daher mit $NB(r, p)$.
- Die Verteilungsfunktion lautet

$$F(x) = \begin{cases} 0 & x < 0 \\ \sum_{k=0}^{[x]} \binom{k+r-1}{r-1} p^r q^k & 0 \leq x, \quad k = 0, 1, \dots, [x]. \end{cases}$$

Negative Binomialverteilung

Wahrscheinlichkeits- und Verteilungsfunktion

In der Abbildung sind Wahrscheinlichkeits- und Verteilungsfunktion der negativen Binomialverteilung für $r = 4$ und $p = 0.5$ dargestellt.



- Die Herleitung des Erwartungswertes und der Varianz finden sich im Buch auf S. 96-97.
- Erwartungswert und Varianz für X betragen

$$\begin{aligned}\mu &= \frac{rq}{p} \quad \text{und} \\ \sigma^2 &= \frac{rq}{p^2}.\end{aligned}$$

Beispiel 4.6: Glücksspiel.

Bei einem Glücksspiel hat derjenige gewonnen, der 4-mal die 6 mit der geringsten Anzahl an Würfeln erzielt. Die Wahrscheinlichkeit, dass ein Teilnehmer dieses Ziel mit sechs Würfeln erreicht, lässt sich mit der negativen Binomialverteilung berechnen.

In diesem Fall muss beim sechsten Wurf zum vierten Mal die 6 eintreten. Die Parameter der negativen Binomialverteilung lauten $p = \frac{1}{6}$ und $r = 4$; X nimmt den Wert zwei an. Es ergibt sich

$$f(2) = \binom{5}{3} p^4 q^2 = 0.0054.$$

Beispiel 4.6: Fortsetzung.

Der Erwartungswert des Spiels beträgt dann

$$\mu = \frac{4 \cdot 5/6}{1/6} = 20,$$

d.h. der vierte Erfolg kann nach 20 Misserfolgen erwartet werden bzw. es muss bis zum vierten Erfolg durchschnittlich 24-mal gespielt werden.

Die Varianz folgt als

$$\sigma^2 = \frac{4 \cdot 5/6}{1/36} = 120,$$

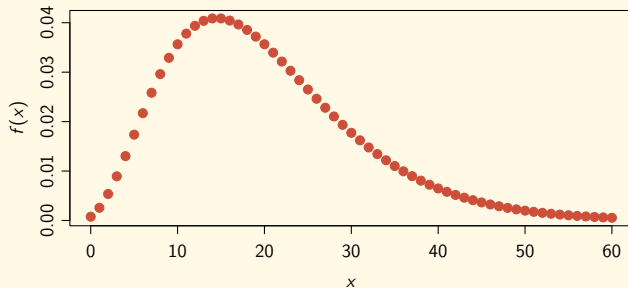
oder $\sigma = 10.9545$. Je größer σ^2 , desto geringer ist die Information, die der Erwartungswert, als mittlere Wartezeit interpretiert, liefert.

Beispiel 4.6: Fortsetzung.

```
dnbinom(x = 2, size = 4, prob = 1/6)
```

```
## [1] 0.005358
```

```
plot(x = 0:60, y = dnbinom(x = 0:60, size = 4, prob = 1/6), pch = 19,  
     col = "tomato3", ylab = "$f(x)$", xlab = "$x$")
```



- Den bisher behandelten Verteilungen liegt *Ziehen mit Zurücklegen* zugrunde.
- Die **hypergeometrische Verteilung** ergibt sich, wenn die Zufallsvorgänge durch *Ziehen ohne Zurücklegen* charakterisiert werden können, bei denen wieder nur das Eintreten einer Eigenschaft A relevant ist.
- Ausgangspunkt ist daher eine endliche Grundgesamtheit vom Umfang N , bei der $M < N$ Elemente eine Eigenschaft A besitzen, die übrigen $N - M$ Elemente nicht.
- Die Zufallsvariablen X_j , $j = 1, \dots, n$, sind eins, wenn das gezogene Element bei der j -ten Entnahme A aufweist; sie sind null wenn nicht.
- Zieht man n -mal ohne Zurücklegen, gibt $X = \sum_{j=1}^n X_j$ an, wie viele der Entnahmen die Eigenschaft A aufweisen.

- Um Wahrscheinlichkeiten zu bestimmen, muss zunächst die Anzahl der Möglichkeiten, n Elemente aus N auszuwählen, berechnet werden.
- Es gibt $\binom{N}{n}$ Möglichkeiten, siehe (2.3).
- Wir nehmen an, dass sie alle die gleiche Ziehungswahrscheinlichkeit besitzen.
- Für $P(X = x)$ benötigt man die Anzahl der Möglichkeiten, die x Elemente mit A enthalten.
- Es gibt $\binom{M}{x}$ Möglichkeiten, x aus M Elementen ohne Zurücklegen auszuwählen, siehe wiederum (2.3).

- Zu jeder dieser Möglichkeiten existieren $\binom{N-M}{n-x}$ Möglichkeiten, $n-x$ Elemente aus $N-M$ Elementen auszuwählen, die nicht A aufweisen.
- Insgesamt gibt es somit $\binom{M}{x} \binom{N-M}{n-x}$ Möglichkeiten, die x Elemente mit Eigenschaft A enthalten.
- Das Verhältnis dieser Anzahl zu den insgesamt möglichen Auswahlen ist die Wahrscheinlichkeit für $X = x$.

- Die Wahrscheinlichkeitsfunktion resultiert also als

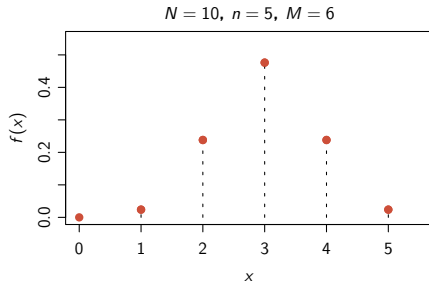
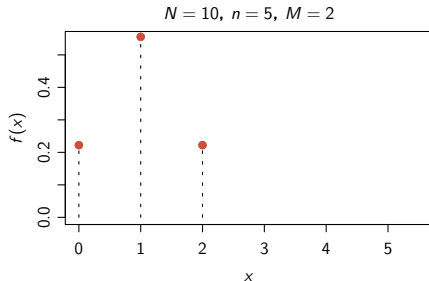
$$f(x) = \begin{cases} \frac{\binom{M}{x} \binom{N-M}{n-x}}{\binom{N}{n}} & \max\{0, n + M - N\} \leq x \leq \min\{n, M\} \\ 0 & \text{sonst.} \end{cases}$$

- Die Verteilungsfunktion folgt hieraus durch Summation über k :

$$F(x) = \begin{cases} 0 & x < \max\{0, n + M - N\} = a \\ \sum_{k=a}^{[x]} \frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}} & a \leq x \leq \min\{n, M\}, \quad k = a, \dots, [x] \\ 1 & x > \min\{n, M\} \end{cases}$$

- Die Beschränkungen für den Wertebereich von X werden im Buch auf S. 109 hergeleitet.

- $f(x)$ und $F(x)$ zeigen, dass die hypergeometrische Verteilung die Parameter N , M und n besitzt; kürze sie daher mit $H(N, M, n)$ ab.
- In der folgenden Abbildung unterscheiden sich die $f(x)$ nur in M :



- Mit zunehmendem M wandert $f(x)$ nach rechts.

- Erwartungswert und Varianz der hypergeometrischen Verteilung sind (Herleitung siehe Anhang):

$$\mu = np \quad \text{bzw.} \quad \sigma^2 = np(1-p) \frac{N-n}{N-1}.$$

- Binomial- und hypergeometrische Verteilung besitzen denselben Erwartungswert, obwohl sich die zugrunde liegenden Zufallsvorgänge unterscheiden (vgl. (4.2)).
- Für den Erwartungswert ist es daher unbedeutend, ob bei dem Urnenmodell das entnommene Element zurückgelegt wird oder nicht.
- Die **Varianz** der hypergeometrischen Verteilung stimmt bis auf den Faktor $\frac{N-n}{N-1}$ auch mit der für die Binomialverteilung überein (vgl. (4.3)).

- Da sich beim *Ziehen mit Zurücklegen* die Grundgesamtheit nicht erschöpft, nennt man $\frac{N-n}{N-1}$ den **Korrekturfaktor für endliche Gesamtheiten**.

- Bei $N = n$ ist, logischerweise, $\sigma^2 = 0$ (warum?).

- Für hinreichend große N gilt

$$\frac{N-n}{N-1} \approx \frac{N-n}{N} = 1 - \frac{n}{N}.$$

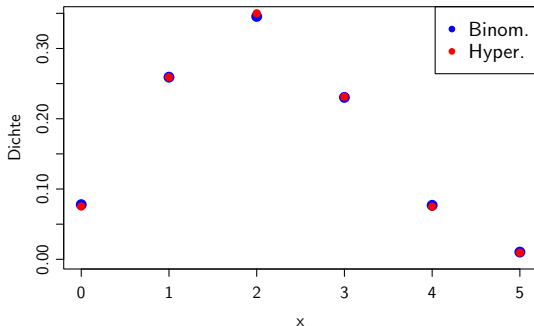
- Der Korrekturfaktor ist bei kleinem Auswahlsatz n/N fast eins.
- Dann bleibt die Wahrscheinlichkeit, Elemente mit A zu erhalten, auch beim *Ziehen ohne Zurücklegen* von Entnahme zu Entnahme fast gleich.
- Die Binomialverteilung kann als Approximation für die hypergeometrische Verteilung dienen, wenn der **Auswahlsatz** n/N kleiner als 5% ist.

Hypergeometrische Verteilung

Vergleich Binomial- und hypergeometrische Verteilung

```
N <- 200; n <- 5  
x <- 0:N; M <- 80  
p <- M/N
```

```
plot(x, dbinom(x,n,p), xlim = c(0,n), cex = 1.3, col = "blue", lwd = 2, ylab = "Dichte")
```



Beispiel 4.7: Lotto, 6 aus 49 (bis 2013).

Man kann sechs der Zahlen 1 bis 49 auf $\binom{N}{n} = \binom{49}{6} = 13\,983\,816$ unterschiedliche Arten ankreuzen. Durch Ziehen ohne Zurücklegen werden sechs Gewinnzahlen ermittelt.

Damit sind die 49 Zahlen in eine Teilmenge mit den $M = 6$ Gewinnzahlen und eine andere Teilmenge der $N - M = 43$ restlichen Zahlen aufgeteilt.

Die Wahrscheinlichkeit, dass in einer Kombination $x = 0, 1, 2, \dots, 6$ Gewinnzahlen vorkommen, erhält man aus

$$f(x) = \frac{\binom{6}{x} \binom{43}{6-x}}{\binom{49}{6}}.$$

Beispiel 4.7: Fortsetzung.

Die Wahrscheinlichkeiten für $x = 0$, $x = 3$ und $x = 5$ betragen gerundet

$$f(x = 0) = 0.4360, f(x = 3) = 0.0177 \text{ und } f(x = 5) = 0.00002.$$

Nach der Ziehung der sechs Gewinnzahlen wird eine Zusatzzahl gezogen. Die Wahrscheinlichkeit für das Ereignis *drei Gewinnzahlen (GZ) und die Zusatzzahl (ZZ)* erhalten wir aus

$$\begin{aligned} P(\text{GZ} = x \text{ und ZZ}) &= P(\text{ZZ} | \text{GZ} = x) P(\text{GZ} = x) \\ &= \frac{6 - x}{43} f(x). \end{aligned}$$

Beispiel 4.7: Fortsetzung.

WSK für 0, 1, 2, ..., 6 Gewinnzahlen

```
round(dhyper(x = 0:6, m = 6, n = 43, k = 6), 5)
```

```
## [1] 0.43596 0.41302 0.13238 0.01765 0.00097 0.00002 0.00000
```

3 Gewinnzahlen + Zusatzzahl

m = 6 -> 6 mögliche Treffer, m = 3 -> nur noch 3 mögliche Treffer

```
dhyper(x = 3, m = 6, n = 43, k = 6) * dhyper(x = 1, m = 3, n = 40, k = 1)
```

```
## [1] 0.001231
```

Man bemerke (siehe ?dhyper), dass $m = M$, $n = N - M$ und $k = n$.

Beachte außerdem im letzten Schritt, dass $M = 6 - x$, $x = 1$, $n = 1$ und $N = 43$, so dass

$$\frac{\binom{6-x}{1} \binom{43-(6-x)}{0}}{\binom{43}{1}} = \frac{6-x}{43}$$

Aufgabe 4.8:

Geben Sie für X , das aus dem Zufallsexperiment *einmaliger Würfelwurf* resultiert, die Wahrscheinlichkeits- und die Verteilungsfunktion an!

Aufgabe 4.9:

Aus einer Urne mit zwei weißen Kugeln und einer schwarzen Kugel wird einmal gezogen. X nimmt 1 an, wenn *weiß* gezogen wird und 0, wenn *schwarz* gezogen wird. Geben Sie für X die Wahrscheinlichkeits- und die Verteilungsfunktion an!

Aufgabe 4.10:

In einer Urne befinden sich 30 grüne und 60 schwarze Kugeln. Es werden 4 Kugeln mit Zurücklegen aus der Urne entnommen.

- (a) Wie groß ist die Wahrscheinlichkeit 2 grüne Kugeln zu ziehen?
- (b) Wie groß ist die Wahrscheinlichkeit höchstens 2 grüne Kugeln zu ziehen?
- (c) Berechnen Sie den Erwartungswert und die Varianz für X !

Aufgabe 4.11:

In einer Urne befinden sich 10 Kugeln von denen 6 schwarz und die restlichen weiß sind. Wenn 5 Kugeln aus der Urne ohne Zurücklegen gezogen werden, wie groß ist die Wahrscheinlichkeit, dass 3 davon schwarz sind?

Aufgabe 4.12:

Zufallsexperiment: Einmaliger Würfelwurf.

- (a) Wie groß ist die Wahrscheinlichkeit, dass im 8. Wurf das 3. Mal die 6 gefallen ist?
- (b) Wie groß ist die Wahrscheinlichkeit, dass spätestens im 8. Wurf die dritte 6 fällt?
- (c) Wie oft muss durchschnittlich gewürfelt werden bis die dritte 6 fällt?

Aufgabe 4.13:

Bei der Untersuchung von Investitionstätigkeiten von Bankangestellten hat sich herausgestellt, dass pro Jahr 15 Transaktionen an der Börse getätigt werden.

- (a) Wie groß ist die Wahrscheinlichkeit, dass in einem Jahr eine Transaktion durchgeführt wird?
- (b) Wie groß ist die Wahrscheinlichkeit, dass in einem Monat eine Transaktion durchgeführt wird?

Aufgabe 4.14:

Ein Basketballspieler hat während seiner Karriere 84% seiner Freiwürfe verwandelt. Wie groß ist die Wahrscheinlichkeit, dass er in einem Spiel 6 von 16 Freiwürfen nicht verwandelt?

Aufgabe 4.15:

In einer Gruppe von 10 Personen präferieren 6 Personen Cola und die restlichen Personen Fanta. Wenn 3 Personen aus dieser Gruppe nach ihrem Lieblingsgetränk gefragt werden, wie groß ist die Wahrscheinlichkeit, dass die Mehrheit Fanta bevorzugt?

Aufgabe 4.16:

Ein Student kommt nach einer Studentenparty angetrunken nach Hause. Er hat einen Schlüsselbund mit 8 Schlüsseln und sucht nach dem Haustürschlüssel.

- (a) Wie groß ist die Wahrscheinlichkeit, dass der 6. Schlüssel, den er ausprobiert in das Schloss passt?
- (b) Wie groß ist die Wahrscheinlichkeit, dass spätestens der 6. Schlüssel passt?
- (c) Wie viele Schlüssel muss er durchschnittlich ausprobieren bis er den richtigen Schlüssel gefunden hat?

Aufgabe 4.17:

Ein Spieler setzt beim Roulettespiel bei jedem der 200 durchgeführten Spiele auf die 17. Die Wahrscheinlichkeit einmal zu gewinnen ist $P(\text{Zahl } 17) = \frac{1}{37} = 0.027$.

Wie groß ist die Wahrscheinlichkeit, dass er genau acht Mal gewinnt?

- Die wohl wichtigste Verteilung der Statistik ist die **Normalverteilung**. Sie wurde von de Moivre, Laplace und Gauß entwickelt.
- Die Normalverteilung beschreibt bestimmte konkrete empirische Verteilungen recht gut.
- Für die Induktive Statistik aber noch bedeutsamer ist, dass bei hinreichend großem Stichprobenumfang auch viele in der Schätz- und Testtheorie auftretenden theoretischen Verteilungen durch die Normalverteilung gut approximiert werden.

- Die Dichte der Normalverteilung lautet

$$f(x) \stackrel{\text{Def. 3.10}}{=} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \quad -\infty < x < \infty,$$

mit Erwartungswert μ und Varianz σ^2 als Scharparameter.

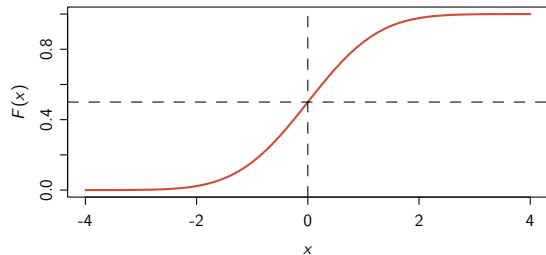
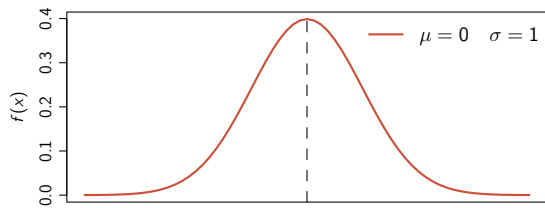
- Daraus erhält man die Verteilungsfunktion als

$$F(x) \stackrel{\text{Def. 3.5}}{=} \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^x e^{-\frac{1}{2}\left(\frac{u-\mu}{\sigma}\right)^2} du, \quad -\infty < x < \infty.$$

- Die Normalverteilung wird abgekürzt mit $N(\mu, \sigma^2)$.

Normalverteilung

Dichte- und Verteilungsfunktion für $\mu = 0$ und $\sigma^2 = 1$



Aufgabe 4.18:

Zeigen Sie, dass die Dichtefunktion der Normalverteilung an der Stelle $x = \mu$ ein Maximum besitzt!

- Die Dichtefunktion ist symmetrisch um μ ; an den Stellen $\mu + x$ und $\mu - x$ nimmt $f(x)$ für alle $x \in \mathbb{R}$ dieselben Werte an.
- Die Dichtefunktion fällt rechts und links von μ streng monoton und nähert sich asymptotisch der Abszisse für $x \rightarrow \pm\infty$.
- Wegen der **Symmetrie** gilt Median = Modus = Erwartungswert.
- Also hat die Verteilungsfunktion bei $x = \mu$ einen Wendepunkt.

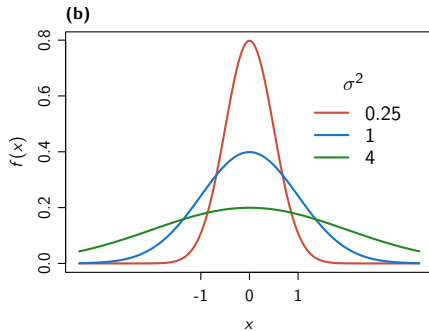
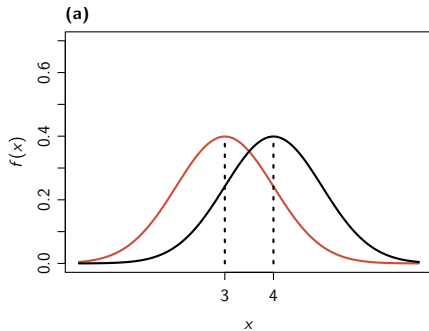
- Wird μ variiert, verschiebt sich die Dichtefunktion auf der x -Achse, ohne dass sich ihre Form ändert.
- Die Auswirkung einer Variation von σ illustriert der Funktionswert $f(x)$ an der Stelle $x = \mu$:

$$f(\mu) = \frac{1}{\sigma\sqrt{2\pi}}.$$

- Nimmt σ zu, wird $f(\mu)$ kleiner. Daher verläuft die Dichtefunktion mit zunehmendem σ flacher.
- Die nächsten Abbildungen zeigen $f(x)$
 - (a) für $\mu = 3$ und $\mu = 4$ bei konstantem σ , sowie
 - (b) für $\sigma^2 = 0.25$, $\sigma^2 = 1$ und $\sigma^2 = 4$ bei konstantem $\mu = 0$.

Normalverteilung

Variation von μ und σ



- Die Normalverteilung besitzt die **Reproduktivitätseigenschaft**, d.h. Linearkombinationen von normalverteilten X_1, \dots, X_n sind auch normalverteilt.
- Zwei wichtige Spezialfälle:
 - (a) Die **Summe** \tilde{S} der X_j , $\tilde{S} = \sum_{j=1}^n X_j$ und
 - (b) das **arithmetische Mittel** $\bar{X} = \frac{1}{n} \sum_{j=1}^n X_j$ sind normalverteilt.
- Im folgenden Kapitel mehr zu den dann resultierenden Parametern der Normalverteilung von \tilde{S} und \bar{X} .

- Ist $X : N(\mu_X, \sigma_X^2)$, dann ist auch $Y = a + bX$ normalverteilt mit Parametern

$$\mu_Y \stackrel{(3.15)}{=} a + b\mu_X \quad \text{und} \quad \sigma_Y^2 \stackrel{(3.16)}{=} b^2\sigma_X^2.$$

- Die Transformation

$$Y = a + bX \quad \text{mit} \quad a = -\mu_X/\sigma_X \quad \text{und} \quad b = 1/\sigma_X \quad (4.4)$$

führt zu einer standardisierten Zufallsvariablen (s. (3.17)), mit $\mu_Y = 0$ und $\sigma_Y^2 = 1$.

- Ist X normalverteilt, so ist die standardisierte Zufallsvariable Z $N(0, 1)$ -verteilt, d.h. **standardnormalverteilt**.
- Zusammengefasst:

$$Z = \frac{X - \mu}{\sigma}, \quad \text{mit } X : N(\mu, \sigma^2) \text{ und } Z : N(0, 1).$$

- Für die Wahrscheinlichkeiten gilt somit

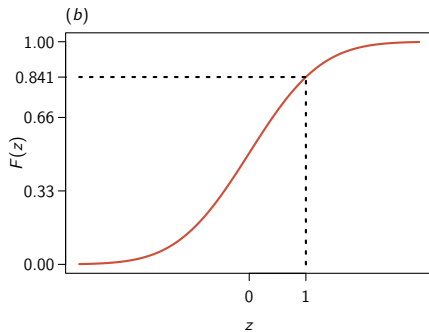
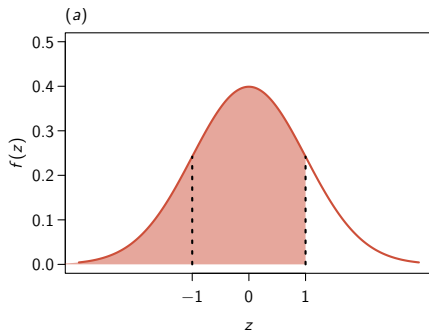
$$P(X \leq a) = P\left(Z \leq \frac{a - \mu}{\sigma}\right). \quad (4.5)$$

- Dichte- und Verteilungsfunktion von Z :

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}, \quad F(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-\frac{u^2}{2}} du.$$

- Die Dichtefunktion hat das Maximum an der Stelle $z = \mu = 0$ und Wendepunkte bei $z = \pm\sigma = \pm 1$; der Wendepunkt der Verteilungsfunktion $F(z)$ liegt bei $z = \mu = 0$.

- Die folgenden Graphen zeigen Dichte- und Verteilungsfunktion einer standardnormalverteilten Zufallsvariable:



Verteilungen 3



- **Symmetrie** der Normalverteilung um Null impliziert, dass

$$P(Z \leq -z_1) = P(Z \geq z_1) = 1 - P(Z \leq z_1)$$

für jedes $z_1 \in \mathbb{R}$.

- Die rötliche Fläche in (a) hat dieselbe Größe wie die Ordinate $F(z_1)$ in (b): $P(Z \leq z_1)$.
- Die Wahrscheinlichkeit, dass Z Werte im Intervall $[z_1, z_2]$ annimmt, berechnet sich (wie üblich) als

$$P(z_1 \leq Z \leq z_2) = P(Z \leq z_2) - P(Z < z_1) = F(z_2) - F(z_1).$$

- Früher hat man $F(z)$ in folgender Tabelle abgelesen:

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817

Beispiel 4.19: Intervallwahrscheinlichkeit.

Es ist $P(Z \leq 1) = F(1) = 0.8413$. Ferner ist

$$P(Z \leq -1) = P(Z \geq 1) = 1 - P(Z \leq 1) = 1 - F(1) = 0.1587.$$

Damit ergibt sich die Wahrscheinlichkeit für Werte im zentralen Schwankungsintervall $-1 \leq Z \leq 1$ als

$$P(-1 \leq Z \leq 1) = P(Z \leq 1) - P(Z \leq -1) \approx 0.8413 - 0.1587 = 0.6826.$$

Beispiel 4.19: Fortsetzung.

```
pnorm(c(-1, 1), mean = 0, sd = 1) # oder pnorm(c(-1, 1))
```

```
## [1] 0.1587 0.8413
```

```
pnorm(1) - pnorm(-1)
```

```
## [1] 0.6827
```

```
# Symmetrie: Flächeninhalt von -Inf bis 0 entspricht dem von 0 bis Inf
```

```
pnorm(0)
```

```
## [1] 0.5
```

Beispiel 4.20: Schwankungsintervall.

X ist normalverteilt $N(5, 16)$. Standardisieren liefert

$$Z = \frac{X - \mu}{\sigma} = \frac{X - 5}{4}.$$

Dann gilt bspw.

$$\begin{aligned} P(X \geq 8) &\stackrel{(4.5)}{=} P\left(Z \geq \frac{8 - \mu}{\sigma}\right) = P(Z \geq 0.75) \\ &= 1 - P(Z < 0.75) = 1 - F(0.75). \end{aligned}$$

Für $z = 0.75$ erhält man $F(0.75) = P(Z \leq 0.75) = 0.7734$. Die gesuchte Wahrscheinlichkeit beträgt daher

$$P(X \geq 8) = 1 - 0.7734 = 0.2266.$$

Beispiel 4.20: Fortsetzung.

```
1 - pnorm(8, mean = 5, sd = 4)
```

```
## [1] 0.2266
```

```
# Oder kürzer:
```

```
1 - pnorm(0.75)
```

```
## [1] 0.2266
```

Für etwa $P(3 \leq X \leq 8)$ muss noch die untere Wahrscheinlichkeit abgezogen werden.

```
pnorm(0.75) - pnorm(-0.5) # wegen Standardisierung (3-5)/4=-0.5
```

```
## [1] 0.4648
```

```
pnorm(8, mean = 5, sd = 4) - pnorm(3, mean = 5, sd = 4) # einfacher
```

```
## [1] 0.4648
```

Aufgabe 4.21:

Berechnen Sie für $Z : N(0, 1)$ folgende Wahrscheinlichkeiten:

- (a) $P(Z \leq 3)$
- (b) $P(Z < 3)$
- (c) $P(Z \leq 2.34)$
- (d) $P(Z \geq 1.2)$

Aufgabe 4.22:

Berechnen Sie für $X : N(8, 25)$ folgende Wahrscheinlichkeiten:

- (a) $P(X \leq 10)$
- (b) $P(X \geq 20)$

Aufgabe 4.23:

In einem Produktionsprozess ist das Füllvolumen von Bierflaschen normalverteilt mit einem Erwartungswert von 500 ml und einer Varianz von 25 ml^2 .

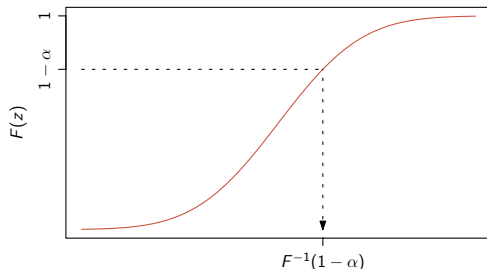
- (a) Welche Mindestfüllmenge kann die Firma garantieren, wenn die Rücklaufquote 1% betragen soll?
- (b) Wie groß ist die Mindestfüllmenge, wenn man sie mit der Ungleichung von Tschebyscheff abschätzt?

- Oft benötigen wir die Werte x , die für vorgegebene Wahrscheinlichkeiten die Gleichung

$$F(x) = P(X \leq x) = 1 - \alpha$$

erfüllen. Diese nennen wir **$1 - \alpha$ -Quantile** einer Verteilung. Links von x liegen also $1 - \alpha$ der Wahrscheinlichkeitsmasse. Ist F invertierbar, so gilt

$$x = F^{-1}(1 - \alpha)$$



- Quantile der Normalverteilung etwa berechnet man in R bequem mit der Funktion `qnorm(p, mean, sd)`.

```
# Quantile der Standardnormalverteilung
```

```
x <- c(0, 0.25, 0.5, 0.75, 1)
```

```
qnorm(p = x, mean = 0, sd = 1)
```

```
## [1]      -Inf -0.6745  0.0000  0.6745      Inf
```

Verteilungen 4



- Ausgangspunkt sind n Zufallsvariablen Z_1, \dots, Z_n , die unabhängig und $N(0, 1)$ -verteilt sind. Die neue Zufallsvariable

$$X = Z_1^2 + Z_2^2 + \dots + Z_n^2$$

folgt einer χ^2 -**Verteilung**.

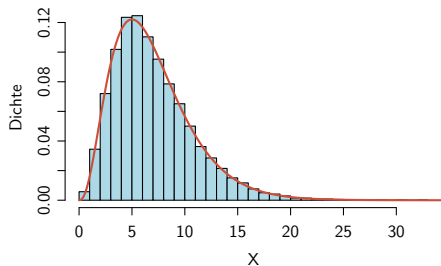
- Die Anzahl der unabhängigen Zufallsvariablen n in dieser Summe wird als **Freiheitsgrad** bezeichnet.
- Erwartungswert und Varianz der χ^2 -Verteilung sind

$$\mu = n \quad \text{und} \quad \sigma^2 = 2n.$$

- Die Freiheitsgrade n sind also Scharparameter der Verteilung. Daher schreibt man die χ^2 -Verteilung kompakt als $\chi^2(n)$.
- Weitere Details finden sich im Buch auf S. 135-138.

```
Wiederholungen <- 20000
FG <- 7 # Freiheitsgrade der chi-Quadrat-Verteilung
Z <- replicate(Wiederholungen, rnorm(FG)) # 20000 Spalten à 7 N(0,1)-ZVen
X <- colSums(Z^2) # Spaltensummen der Quadrate

# Histogramm der quadrierten Spaltensummen:
hist(X, freq = F, col="lightblue", breaks = 40, ylab="Dichte", main="")
grid <- seq(0, 60, 0.03) # x-Achse definieren
# theoretische Dichte hinzufügen:
lines(grid, dchisq(x = grid, df = FG), type = 'l', lwd = 2, col="tomato3")
```



- Bildet man aus unabhängigen Zufallsvariablen $Z : N(0, 1)$ und $Y : \chi^2(n)$ den Quotienten

$$X = \frac{Z}{\sqrt{Y/n}}, \quad (4.6)$$

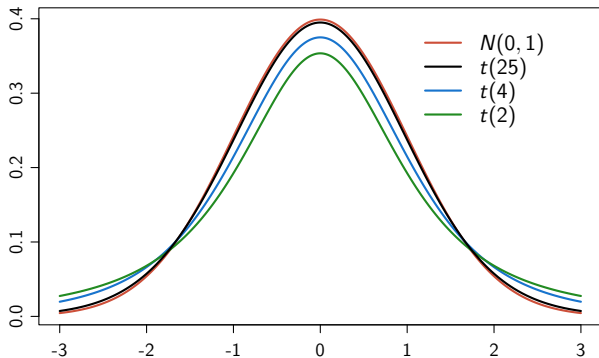
dann ist X **t-verteilt** mit n Freiheitsgraden.

- Die t -Verteilung hängt auch von den Freiheitsgraden n ab; kürze sie daher mit $t(n)$ ab.
- Der **Erwartungswert** ist für $n > 1$ definiert und ist dann immer null. Die **Varianz** ist

$$\sigma^2 = \frac{n}{n-2} \quad \text{für } n > 2.$$

- Die t -Verteilung ist symmetrisch um $\mu = 0$, d.h. $F(-x) = 1 - F(x)$.

```
grid <- seq(-3, 3, .01)
plot(x = grid, y = dnorm(grid), type = "l",
     col = "tomato3", xlab = "", ylab = "", lwd = 2)
lines(x = grid, y = dt(grid, df = 4), col = "dodgerblue3", lwd = 2)
lines(x = grid, y = dt(grid, df = 2), col = "forestgreen", lwd = 2)
```



- Die Abbildung zeigt $f(x)$ der t -Verteilung für $n = \{2, 4, 25\}$ und die Dichte der $N(0, 1)$ -Verteilung.
- Für kleines n verläuft die t -Verteilung flacher als die $N(0, 1)$ -Verteilung.
- Dieser Unterschied verringert sich aber mit wachsendem n , so dass es ab $n \geq 30$ zulässig ist, anstelle der t - die $N(0, 1)$ -Verteilung zu verwenden.

```
x <- seq(-5,5)
round(dnorm(x) - dt(x, 30), 3)

## [1] 0.000 0.000 -0.002 -0.003 0.004 0.003 0.004 -0.003 -0.002 0.000
## [11] 0.000
```

Aufgabe 4.24:

Finden Sie das folgende Quantil der t -Verteilung:

$$t(0.95; 30).$$

- Definitionsgemäß erhält man μ der geometrischen Verteilung als

$$\mu = \sum_{x=0}^{\infty} x p q^x = p q \sum_{x=1}^{\infty} x q^{x-1}.$$

- Um zu ermitteln, ob die Summe existiert und welchen Wert sie annimmt, wird der Ausdruck xq^{x-1} als Ableitung der Funktion q^x nach q geschrieben

$$\frac{d}{dq}(q^x) = xq^{x-1}.$$

- Bei endlichen Summen können die beiden Operationen \sum und $\frac{d}{dq}$ vertauscht werden:

$$\sum_{x=1}^{\infty} x q^{x-1} = \sum_{x=1}^{\infty} \frac{d}{dq}(q^x) = \frac{d}{dq} \left(\sum_{x=1}^{\infty} q^x \right).$$

- Wegen $|q| < 1$ existiert $\sum_{x=1}^{\infty} q^x$; aus der Summenformel für unendliche geometrische Reihen folgt

$$\sum_{x=1}^{\infty} q^x = \frac{q}{1-q}.$$

Die Ableitung dieser Summe lautet

$$\frac{d}{dq} \left(\frac{q}{1-q} \right) = \frac{1}{(1-q)^2}.$$

Damit gilt

$$\sum_{x=1}^{\infty} xq^{x-1} = \frac{1}{(1-q)^2} \text{ für } |q| < 1.$$

- Durch Einsetzen von $\sum_{x=1}^{\infty} xq^{x-1} = \frac{1}{(1-q)^2}$ in $\mu = \sum_{x=0}^{\infty} xpq^x = pq \sum_{x=1}^{\infty} xq^{x-1}$ Daher folgt für den

Erwartungswert

$$\mu = \sum_{x=0}^{\infty} xpq^x = pq \sum_{x=1}^{\infty} xq^{x-1} = \frac{pq}{(1-q)^2} = \frac{q}{p}.$$

- Zur Berechnung der **Varianz** folgt man dem Ansatz

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

- Da μ bereits vorliegt, fehlt nur noch $E(X^2)$:

$$E(X^2) = \sum_{x=0}^{\infty} x^2 p q^x = p q \sum_{x=1}^{\infty} x^2 q^{x-1}.$$

- Die letzte Summe lässt sich wieder als Ableitung nach q darstellen

$$\sum_{x=1}^{\infty} x^2 q^{x-1} = \frac{d}{dq} \left(\sum_{x=1}^{\infty} x q^x \right) = \frac{d}{dq} \left(q \sum_{x=1}^{\infty} x q^{x-1} \right).$$

- Ersetzt man jetzt die rechte Summe durch Gleichung

$$\sum_{x=1}^{\infty} xq^{x-1} = \frac{1}{(1-q)^2},$$

folgt

$$\sum_{x=1}^{\infty} x^2 q^{x-1} = \frac{d}{dq} \left[\frac{q}{(1-q)^2} \right] = \frac{1+q}{(1-q)^3} \text{ für } |q| < 1.$$

- Also ist

$$E(X^2) = pq \frac{1+q}{(1-q)^3} = \frac{q(1+q)}{(1-q)^2}.$$

- Die Varianz der geometrischen Verteilung folgt jetzt als

$$\sigma^2 = E(X^2) - \mu^2 = \frac{q(1+q)}{(1-q)^2} - \frac{q^2}{p^2} = \frac{q}{(1-q)^2} = \frac{q}{p^2}.$$

- Für Erwartungswert und Varianz einer hypergeometrischen Verteilung ist folgende Umformung des Binomialkoeffizienten hilfreich. Es gilt

$$\binom{M}{x} = \frac{M!}{x!(M-x)!} = \frac{M(M-1)!}{x(x-1)!(M-x)!} = \frac{M}{x} \binom{M-1}{x-1}.$$

- Der Erwartungswert kann dann wie folgt aus seiner Definitionsgleichung ermittelt werden

$$\mu = \sum_{x=0}^n x \frac{\binom{M}{x} \binom{N-M}{n-x}}{\binom{N}{n}} = \sum_{x=1}^n x \frac{\frac{M}{x} \binom{M-1}{x-1} \binom{N-M}{n-x}}{\frac{N}{n} \binom{N-1}{n-1}} = n \frac{M}{N} \sum_{x=1}^n \frac{\binom{M-1}{x-1} \binom{N-M}{n-x}}{\binom{N-1}{n-1}}.$$

- M/N stellt die Wahrscheinlichkeit dar, dass beim ersten Zug ein Element mit der Eigenschaft A eintritt.
- Diese Wahrscheinlichkeit beträgt p . Setzt man $z = x - 1$, kann die letzte Summe geschrieben werden

$$\sum_{x=1}^n \frac{\binom{M-1}{x-1} \binom{N-M}{n-x}}{\binom{N-1}{n-1}} = \sum_{z=0}^{n-1} \frac{\binom{M-1}{z} \binom{N-M}{n-1-z}}{\binom{N-1}{n-1}}.$$

- Der letzte Bruch ist aber die Wahrscheinlichkeitsfunktion einer hypergeometrisch verteilten Zufallsvariablen Z mit den Parametern $N - 1$, $M - 1$ und $n - 1$: $f(z|N - 1, M - 1, n - 1)$.
- Da ihre Summe über $z = 0, \dots, n - 1$ gleich eins ist, ist der Erwartungswert μ einer hypergeometrisch verteilten Zufallsvariablen

$$\mu = np.$$

- Forme zur Ermittlung der **Varianz** den Verschiebungssatz um:

$$\sigma^2 = E(X^2) - [E(X)]^2 = E[X(X-1)] + E(X) - [E(X)]^2.$$

- Beachtet man, dass außerdem gilt

$$\binom{M}{x} = \frac{M(M-1)}{x(x-1)} \binom{M-2}{x-2} \quad \text{und} \quad \binom{N}{n} = \frac{N(N-1)}{n(n-1)} \binom{N-2}{n-2},$$

ergibt sich

$$E[X(X-1)] = \sum_{x=0}^n x(x-1) \frac{\binom{M}{x} \binom{N-M}{n-x}}{\binom{N}{n}} = \frac{M(M-1)}{\frac{N(N-1)}{n(n-1)}} \sum_{x=2}^n \frac{\binom{M-2}{x-2} \binom{N-M}{n-x}}{\binom{N-2}{n-2}}.$$

- Setzt man jetzt $z = x - 2$ und stellt den Bruch vor dem letzten Summenzeichen um, erhält man

$$E[X(X-1)] = \frac{p(M-1)n(n-1)}{(N-1)} \sum_{z=0}^{n-2} \frac{\binom{M-2}{z} \binom{N-M}{n-2-z}}{\binom{N-2}{n-2}}, \quad \text{mit } p = \frac{M}{N}.$$

- Der letzte Bruch ist analog zu oben die Wahrscheinlichkeitsfunktion einer hypergeometrisch verteilten Zufallsvariablen mit $N - 2$, $M - 2$ und $n - 2$. Da auch ihre Summe eins ist, resultiert

$$E[X(X - 1)] = \frac{p(M - 1)n(n - 1)}{(N - 1)} = \frac{np(pN - 1)(n - 1)}{N - 1}.$$

- Es sind nun alle Terme bestimmt, so dass

$$\sigma^2 = \frac{np(pN - 1)(n - 1)}{N - 1} + np - n^2 p^2,$$

oder nach einigen Umformungen

$$\sigma^2 = np(1 - p) \frac{N - n}{N - 1}.$$

- 1 Einleitung
- 2 Grundlagen
- 3 Eindimensionale Zufallsvariablen
- 4 Ausgewählte theoretische Verteilungen
- 5 Grundzüge der Stichprobentheorie**
- 6 Statistische Schätzverfahren
- 7 Statistische Testverfahren
- 8 Zweidimensionale Zufallsvariablen

- Enthält eine Grundgesamtheit zu viele Elemente, sind **Stichproben** notwendig.
- Eine Grundgesamtheit kann endlich oder unendlich viele Elemente enthalten:
 - ▶ Theoretische Grundgesamtheiten sind oft (überabzählbar) unendlich, wie z.B. bei stetigen Zufallsvariablen.
 - ▶ Reale Grundgesamtheiten sind meistens sehr groß, aber häufig endlich (z.B. deutsche Bevölkerung).

- Bei unbekannter Verteilung in der Grundgesamtheit muss jede statistische Information hierüber durch Stichproben erhoben werden.
- Das **Ziehen einer Stichprobe** besteht darin, n Elemente aus der Grundgesamtheit auszuwählen.
- So entsteht eine Stichprobe mit dem Umfang oder der Länge n .
- Um die Wahrscheinlichkeitstheorie für die Stichprobenanalyse nutzbar zu machen, muss das **Auswahlverfahren** randomisiert sein.
- Bei zufälliger Auswahl liegen **Zufallsstichproben** vor.

Stichproben 1



- Jede zufällige Ziehung aus der Grundgesamtheit kann als eigenständige Zufallsvariable X_1, \dots, X_n aufgefasst werden.
- Ihre Verteilung hängt von der Verteilung von X in der Grundgesamtheit ab.
- Die X_j bezeichnet man als **Stichprobenvariablen**.
- Wird das Element nach Ziehung wieder in die „Urne“ zurückgelegt, spricht man von **Zufallsstichproben mit Zurücklegen**.
- Geschieht dies nicht, liegen **Zufallsstichproben ohne Zurücklegen** vor (von denen wir im Folgenden abstrahieren wollen).
- Je größer die Grundgesamtheit, desto bedeutungsloser die Unterscheidung ZmZ/ZoZ .

- Eine konkrete Zufallsstichprobe, auch als **Stichprobenrealisation** bzw. **Stichprobenergebnis** bezeichnet, sind also die Zahlen, die sich aus den Realisationen der X_j ergeben,

$$X_1 = x_1, \quad X_2 = x_2, \quad \dots, X_n = x_n$$

oder als n -Tupel geschrieben: (x_1, \dots, x_n) .

- Alle möglichen Stichproben stellen die Ausgänge und diese zusammen den **Stichprobenraum** des Zufallsvorgangs „Ziehen einer Stichprobe“ dar.

Anmerkung 5.1: Uneingeschränkte Zufallsstichprobe.

Jedes Element der Grundgesamtheit besitzt dieselbe Wahrscheinlichkeit, in eine Zufallsstichprobe mit dem Umfang n zu gelangen.

- ZmZ und ZoZ liefern immer **uneingeschränkte Zufallsstichproben** und die X_j sind identisch verteilt wie X .
- Bei ZmZ sind die Stichprobenvariablen zusätzlich stochastisch unabhängig. Man erhält dann **unabhängige Zufallsstichproben**.
- Uneingeschränkte und unabhängige Zufallsstichproben heißen **einfache Zufallsstichproben**.

- Eine Funktion $T = T(X_1, \dots, X_n)$, $n \geq 1$, die nur von Stichprobenvariablen, nicht aber von unbekannten Parametern abhängt, heißt **Stichprobenfunktion** bzw. **Statistik**.
- Stichprobenfunktionen sollen Stichprobenergebnisse in aussagekräftige, einfache statistische Größen überführen.
- So stellen

$$T = \frac{1}{n} \sum_{j=1}^n X_j \quad \text{oder} \quad T = X_1 - 5$$

Stichprobenfunktionen bzw. Statistiken dar, nicht aber

$$Z = \frac{X - \mu}{\sigma},$$

wenn μ und σ unbekannt sind.

- Da die Argumente von T Zufallsvariablen sind, ist T selbst eine Zufallsvariable, deren Verteilung **Stichprobenverteilung** heißt.
- Die Varianz der Stichprobenverteilung nennt man **Fehlervarianz**, die positive Wurzel hieraus **Standardfehler**.
- Eine Realisation von T aufgrund einer konkreten Stichprobe wird mit

$$t = T(x_1, \dots, x_n)$$

bezeichnet.

- Im Folgenden betrachten wir einige spezielle Stichprobenverteilungen.

- Ziehen aller möglicher Stichproben aus einer Grundgesamtheit und jeweiliges Berechnen der Stichprobenfunktion arithmetisches Mittel, $\bar{X} = \frac{1}{n} \sum_{j=1}^n X_j$, liefert dessen **Stichprobenverteilung**.
- Wir betrachten zunächst Erwartungswert und Varianz der Verteilung.
- Der Erwartungswert von \bar{X} ist

$$E(\bar{X}) = E\left(\frac{1}{n} \sum_{j=1}^n X_j\right) = \frac{1}{n} \sum_{j=1}^n E(X_j) = \frac{n}{n} \mu = \mu, \quad (5.1)$$

eine Vereinfachung von (3.14), da nun wegen der identischen Verteilung alle Erwartungswerte gleich sind.

- Bei einfachen Stichproben erhält man die Fehlervarianz als

$$\begin{aligned}\text{Var}(\bar{X}) &= \sigma_{\bar{X}}^2 = \text{Var}\left(\frac{1}{n} \sum_{j=1}^n X_j\right) \stackrel{(3.11)}{=} \frac{1}{n^2} \sum_{j=1}^n \text{Var}(X_j) \\ &= \frac{1}{n^2} \sum_{j=1}^n \sigma^2 = \frac{1}{n} \sigma^2.\end{aligned}\tag{5.2}$$

- Um die Verteilung von \bar{X} zu bestimmen, muss unterschieden werden, ob die Verteilung für X in der Grundgesamtheit bekannt ist oder nicht.
- Ist X **normalverteilt** mit μ und σ^2 , dann ist aufgrund der Reproduktivitätseigenschaft der Normalverteilung auch \bar{X} normalverteilt mit (bei Unabhängigkeit) den oben abgeleiteten Parametern

$$\mu_{\bar{X}} = \mu \quad \text{und} \quad \sigma_{\bar{X}}^2 = \frac{\sigma^2}{n}.$$

Beispiel 5.2: Körpergröße deutscher Männer.

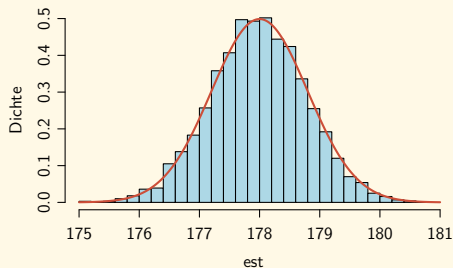
Die Körpergröße deutscher Männer sei in der Grundgesamtheit normalverteilt mit $\mu = 178$ cm und $\sigma^2 = 64$ cm². Das Stichprobenmittel \bar{X} ist dann normalverteilt mit $\mu_{\bar{X}} = 178$ und $\sigma_{\bar{X}}^2 = 64/n$. Wird eine Stichprobe mit dem Umfang $n = 100$ gezogen, ist $\sigma_{\bar{X}}^2 = 64/100 = 0.64$.

Die Wahrscheinlichkeit, dass eine Stichprobe einen Mittelwert von $\bar{x} = 180$ oder größer liefert, ist mit `1-pnorm(180, 178, sqrt(0.64)) = 0.0062`

$$P(\bar{X} \geq 180) = 0.0062.$$

Beispiel 5.2: Fortsetzung.

```
mu    <- 178; sigma <- 8; N <- 100 # EW, Standardabw., Stichprobengröße
grid  <- seq(mu - 3, mu + 3, .01)
# ziehe 5000 Stichproben der Größe 100 und berechne deren Stichprobenmittel:
est   <- colMeans(replicate(5000, rnorm(N, mu, sigma)))
hist(est, freq = F, col = "lightblue", breaks = 40, ylab = "Dichte", main = "")
lines(x = grid, y = dnorm(grid, mu, sigma/sqrt(N)), col = "tomato3", lwd = 2)
```



Aufgabe 5.3:

Die vom Unternehmen A hergestellten Glühlampen besitzen eine normalverteilte Lebensdauer mit einem Erwartungswert von 800 Stunden und einer Standardabweichung von 40 Stunden.

Eine Geschäftsfiliale bekommt eine Lieferung von 100 Glühlampen. Wie groß ist die Wahrscheinlichkeit, dass die durchschnittliche Lebensdauer weniger als 790 Stunden beträgt?

Was passiert mit dieser Verteilung, wenn n sehr groß wird? Betrachte hierzu \bar{X} als Folge von Zufallsvariablen und schreibe deswegen $\bar{X}(n)$.

Definition 5.4: Schwaches Gesetz der großen Zahlen.

Das **schwache Gesetz der großen Zahlen** lautet

$$\lim_{n \rightarrow \infty} P[|\bar{X}(n) - \mu| > \varepsilon] = 0 \quad \text{bzw.} \quad \lim_{n \rightarrow \infty} P[|\bar{X}(n) - \mu| < \varepsilon] = 1.$$

Mehr hierzu im Buch auf S. 171-176.

Der **zentrale Grenzwertsatz von Lindeberg-Levy** liefert bei **unbekannter Verteilung** von X eine Approximation der Verteilung von \bar{X} .

Definition 5.5: Zentraler Grenzwertsatz von Lindeberg-Levy.

Die Verteilung der standardisierten Summe und des standardisierten Durchschnitts, also

$$\frac{\tilde{S}(n) - n\mu}{\sqrt{n}\sigma} \quad \text{bzw.} \quad \frac{\bar{X}(n) - \mu}{\sigma/\sqrt{n}} \rightarrow_D N(0, 1), \quad (5.3)$$

d.h., konvergiert für $n \rightarrow \infty$ gegen die **Standardnormalverteilung** $N(0, 1)$.

Dieser basiert auf der Annahme, dass X_1, \dots, X_n unabhängig identisch verteilt sind mit $E(X_j) = \mu$ und $\text{Var}(X_j) = \sigma^2 > 0$ für alle j . Mehr nicht!

Aus dem zentralen Grenzwertsatz von **Lindeberg-Levy** folgt daher für ein großes n , dass

- (a) die Summe $\tilde{S}(n)$ asymptotisch normalverteilt $N(n\mu, n\sigma^2)$ und
- (b) der Durchschnitt $\bar{X}(n)$ asymptotisch normalverteilt $N(\mu, \sigma^2/n)$

ist. Siehe c1t.R.

Aufgabe 5.6:

Die maximale Traglast eines der beiden Besucheraufzüge im Reichstagsgebäude in Berlin beträgt 4000 kg. Nach Herstellerangaben entspricht dies einer Kapazität von 45 Personen. Das Gewicht der Personen, die den Aufzug benutzen, beträgt im Durchschnitt 88 kg bei einer Varianz von 16 kg^2 .

Wie groß ist die Wahrscheinlichkeit, dass bei 45 Personen die maximale Traglast überschritten wird?

Aufgabe 5.7:

2500 Unternehmensberater wurden bzgl. ihres Jahreseinkommens befragt. Es ergab sich ein Durchschnittseinkommen von 51.800 Euro mit einer Standardabweichung von 4000 Euro.

Wie groß ist die Wahrscheinlichkeit, dass bei einer Stichprobe im Umfang von $n = 30$ der Mittelwert um höchstens 500 Euro vom Erwartungswert abweicht?

Siehe Buch ab S. 192.

- Aus (4.4) folgt, dass der standardisierte Mittelwert

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1), \quad (5.4)$$

also $N(0, 1)$ -verteilt ist, wenn X normalverteilt ist.

- Dieses Ergebnis setzt aber die Kenntnis der Varianz σ^2 von X in der Grundgesamtheit voraus. Ist σ^2 **unbekannt**, ist Z keine Stichprobenfunktion mehr.
- Ersetzt man σ^2 durch die **Stichprobenvarianz**

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2,$$

geht Z über in

$$Z(t) = \frac{\bar{X} - \mu}{S/\sqrt{n}}.$$

- $Z(t)$ ist jetzt aber nicht mehr standardnormalverteilt.
- $Z(t)$ ist bei bekanntem μ eine Stichprobenfunktion.
- Um das Verteilungsgesetz von $Z(t)$ zu entwickeln, erweitere

$$Z(t) = \frac{\bar{X} - \mu}{S/\sqrt{n}} = \frac{\overbrace{\frac{\bar{X} - \mu}{\sigma}}^{N(0,1), \text{ s. (5.4)}}}{\sqrt{\underbrace{\frac{(n-1)S^2}{\sigma^2}}_{\chi^2}} / n-1}.$$

- Der Zähler Z in $Z(t)$ ist standardnormalverteilt.
- $(n-1)S^2/\sigma^2$ ist bei normalverteiltem X χ^2 -verteilt mit $n-1$ Freiheitsgraden.

- Wir erinnern uns (vgl. (4.6)), dass der Quotient aus einer $N(0, 1)$ - und der Wurzel einer χ^2 -verteilten Zufallsvariablen geteilt durch ihre Freiheitsgrade **t-verteilt** ist, sofern die ZVen **unabhängig** sind.
- Da \bar{X} und S^2 auf derselben Stichprobe beruhen, hängt S^2 von \bar{X}^2 ab. Die beiden Zufallsvariablen \bar{X} und S^2 sind daher i.A. stochastisch abhängig.
- Eine Ausnahme liegt vor, wenn X in der Grundgesamtheit **normalverteilt** ist: Dann sind \bar{X} und S^2 unabhängig.
- Ist also X in der Grundgesamtheit normalverteilt, dann folgt $Z(t)$ einer t -Verteilung mit $n - 1$ Freiheitsgraden.

Beispiel 5.8:

Aus einer normalverteilten Grundgesamtheit mit $\mu = 10$ und unbekannter Varianz wird eine Stichprobe mit $n = 9$ gezogen.

Die Stichprobenvarianz sei $s^2 = 9$. Die Variable

$$Z(t) = \frac{\bar{X} - 10}{3/\sqrt{9}} = \bar{X} - 10$$

ist t -verteilt mit $n - 1 = 8$ Freiheitsgraden.

Stichproben 2



Beispiel 5.9:

Das Stichprobenmittel für $n = 9$ betrage $\bar{x} = 12$, so dass $Z(t) = 2$. Die Wahrscheinlichkeit dafür, dass \bar{X} kleiner gleich 12 ist, beträgt

$$P(\bar{X} \leq 12) = P[Z(t) \leq 2].$$

$P[Z(t) \leq 2]$ ermitteln wir in  durch $\text{pt}(2, 8) = 0.9597$.

- Für ein großes n ist die zuvor vorgestellte Vorgehensweise auch bei **unbekannter Verteilung** von X und unbekannter Varianz gültig.
- Aufgrund des Gesetzes der großen Zahlen lässt sich die Varianz für großes n gut über die Stichprobenvarianz approximieren, also $S^2 \approx \sigma^2$.
- Also gilt

$$Z(t) = \frac{\bar{X} - \mu}{S/\sqrt{n}} \approx \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1),$$

d.h. wir verwenden den Zentralen Grenzwertsatz.

- 1 Einleitung
- 2 Grundlagen
- 3 Eindimensionale Zufallsvariablen
- 4 Ausgewählte theoretische Verteilungen
- 5 Grundzüge der Stichprobentheorie
- 6 Statistische Schätzverfahren**
- 7 Statistische Testverfahren
- 8 Zweidimensionale Zufallsvariablen

- Bei vielen empirischen Untersuchungen sind die Parameter in der Grundgesamtheit unbekannt.
- Die **statistische Schätztheorie** befasst sich mit Möglichkeiten, diese Parameter aus Stichproben zu schätzen.
- Die Quantifizierung der Sicherheit solcher Schätzungen nutzt die Ergebnisse der Stichprobentheorie.
- Die Schätzung eines unbekannten Parameters θ einer Grundgesamtheit wird mit einer Stichprobenfunktion durchgeführt, die jetzt **Schätzfunktion** oder **Schätzer** heißt.
- Schätzer haben also auch eine Verteilung.

- Die Schätzfunktion lautet

$$\hat{\theta} = T_{\theta}(X_1, \dots, X_n).$$

$\hat{\theta}$ ist dann eine Zufallsvariable mit Schätzwerten, die unterschiedlich stark von θ abweichen.

- Einsetzen von Stichprobenrealisationen in den Schätzer liefert einen konkreten Schätzwert. Dieser Schätzwert heißt **Punktschätzung**.
- Eine **Intervallschätzung** liefert eine Intervallvorschrift, die mit vorgegebener Wahrscheinlichkeit den unbekannten Parameter θ enthält.

- Wir suchen Schätzer, die bestimmte Qualitätskriterien erfüllen.
- Die Abweichung $\hat{\theta} - \theta$ heißt **Schätzfehler**.
- Eine wünschenswerte Eigenschaft ist, dass der Erwartungswert des Schätzfehlers für jeden Stichprobenumfang n null ist.
- Formal:

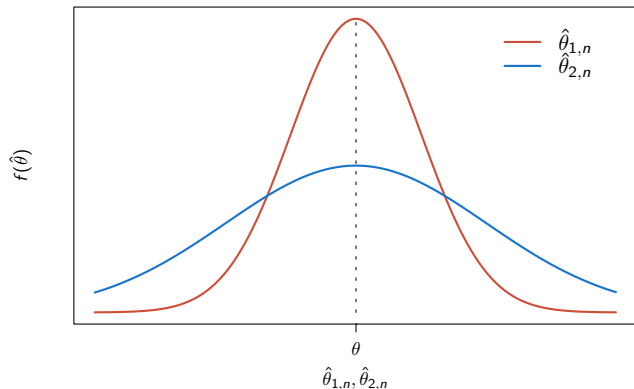
$$E(\hat{\theta}_n - \theta) = 0 \quad \text{oder} \quad E(\hat{\theta}_n) = \theta \quad \text{für alle } n.$$

- Ein Schätzer $\hat{\theta}_n$, die dieses Kriterium erfüllt, heißt **erwartungstreu**, **unverzerrt** oder **unbiased**.
- Andernfalls **nicht erwartungstreu** bzw. **verzerrt**.
- Es gibt oft mehrere erwartungstreue Schätzer.
- Aus der Klasse erwartungstreuer Schätzer liegt es nahe, den mit der **kleinsten Varianz** zu bestimmen.

Schätzfrage



- Hier sind die Stichprobenverteilungen für zwei erwartungstreue Schätzer $\hat{\theta}_{1,n}$ und $\hat{\theta}_{2,n}$ dargestellt. $\hat{\theta}_{1,n}$ ist effizienter als $\hat{\theta}_{2,n}$, da $\text{Var}(\hat{\theta}_{2,n}) > \text{Var}(\hat{\theta}_{1,n})$.



- Den Vorteil einer erwartungstreuen und effizienten Schätzung kann man analytisch zeigen. Die Tschebyscheff-Ungleichung (3.18) liefert

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

- Setzt man $X = \hat{\theta}$, $\mu = E(\hat{\theta}) = \theta$ und $\sigma^2 = \text{Var}(\hat{\theta})$ folgt

$$P(|\hat{\theta} - \theta| \geq \varepsilon) \leq \frac{\text{Var}(\hat{\theta})}{\varepsilon^2}.$$

- Je kleiner $\text{Var}(\hat{\theta})$, desto geringer wird die Wahrscheinlichkeit, dass der absolute Schätzfehler $|\hat{\theta} - \theta|$ größer als ε ist.
- Daher ist $\hat{\theta}_{1,n}$ dem Schätzer $\hat{\theta}_{2,n}$ vorzuziehen.

- Eine weitere wünschenswerte Eigenschaft von erwartungstreuen Schätzern ist die Konvergenz der Varianz mit zunehmendem n gegen null.
- Für $n \rightarrow \infty$ folgt

$$\lim_{n \rightarrow \infty} P(|\hat{\theta}_n - \theta| \geq \varepsilon) = 0 \quad \text{für} \quad \lim_{n \rightarrow \infty} \text{Var}(\hat{\theta}_n) = 0.$$

- Schätzfunktionen mit der Eigenschaft

$$\lim_{n \rightarrow \infty} P(|\hat{\theta}_n - \theta| < \varepsilon) = 1$$

heißen (schwach) **konsistent**.

- Diese Eigenschaft entspricht formal dem Gesetz der großen Zahlen (siehe 5.4).

- Bei verzerrten Schätzern ist der Erwartungswert des Schätzfehlers nicht null:

$$E(\hat{\theta}_n) \neq \theta.$$

- Die Differenz $E(\hat{\theta}_n) - \theta$ misst die **Verzerrung**.
- Bei einigen verzerrten Schätzern stellt sich für $n \rightarrow \infty$ Erwartungstreue ein:

$$\lim_{n \rightarrow \infty} E(\hat{\theta}_n - \theta) = 0 \quad \text{bzw.} \quad \lim_{n \rightarrow \infty} E(\hat{\theta}_n) = \theta.$$

- Diesen Fall nennt man **asymptotische Erwartungstreue**.
- Auch asymptotisch erwartungstreue Schätzer können konsistent sein.
- Betrachte hierzu den **mittleren quadratischen Fehler** (MSE)

$$\text{MSE}(\hat{\theta}_n) = E[(\hat{\theta}_n - \theta)^2].$$

- Das Buch auf S. 214 zeigt, dass der mittlere quadratische Fehler die Summe von Varianz $\text{Var}(\hat{\theta}_n)$ und quadrierter Verzerrung $[E(\hat{\theta}_n) - \theta]^2$ ist:

$$\text{MSE}(\hat{\theta}_n) = \text{Var}(\hat{\theta}_n) + [E(\hat{\theta}_n) - \theta]^2.$$

- Der verzerrte Schätzer ist konsistent, wenn gilt

$$\lim_{n \rightarrow \infty} \text{MSE}(\hat{\theta}_n) = \lim_{n \rightarrow \infty} \text{Var}(\hat{\theta}_n) + \lim_{n \rightarrow \infty} [E(\hat{\theta}_n) - \theta]^2 = 0.$$

- Da weder die Varianz noch die quadrierte Verzerrung negativ werden können, müssen beide Grenzwerte null sein.
- Damit ist eine hinreichende Bedingung für **Konsistenz** gefunden.

$$(a) \lim_{n \rightarrow \infty} [E(\hat{\theta}_n) - \theta]^2 = 0 \quad \text{und} \quad (b) \lim_{n \rightarrow \infty} \text{Var}(\hat{\theta}_n) = 0$$

Beispiel 6.1: \bar{X} als Schätzer für μ .

\bar{X} ist ein erwartungstreuer Schätzer für μ . Der $\text{MSE}(\bar{X})$ ist daher gleich der Varianz,

$$\begin{aligned}\text{MSE}(\bar{X}) &= \text{Var}(\bar{X}) \\ &\stackrel{(5.2)}{=} \frac{\sigma^2}{n}.\end{aligned}$$

Beispiel 6.2: Werfen eines Würfels.

Die ZV X sei die realisierte Augenzahl beim Werfen eines Würfels. Es gilt

$$E(\bar{X}) = \frac{1}{n} \sum_{j=1}^n E(X_j) = E(X_j) = \mu \stackrel{(3.20)}{=} 3.5$$

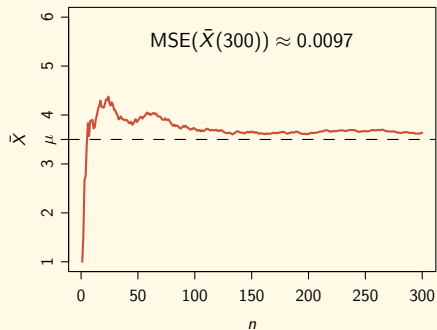
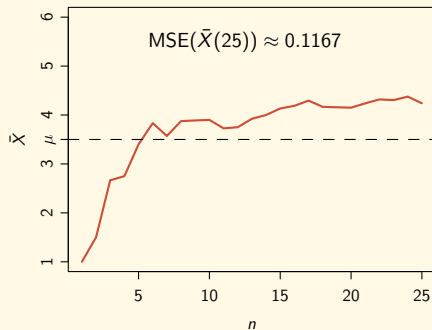
sowie

$$\text{MSE}(\bar{X}) = \text{Var}(\bar{X}) = \frac{\sigma^2}{n} \stackrel{(3.25)}{\approx} \frac{2.9167}{n}$$

```
n      <- 300 # 300 Würfe
Xbar <- cumsum(sample(1:6, size = n, replace = T)) / 1:n # Kumulierte Mittel
head(Xbar)

## [1] 1.000 1.500 2.667 2.750 3.400 3.833
```

Beispiel 6.2: Fortsetzung.



- Man spricht dann von **Konvergenz im quadratischen Mittel**.
- Im quadratischen Mittel konvergente Schätzer sind immer auch mindestens asymptotisch erwartungstreu.
- Konsistenz ist eine Mindestanforderung für einen guten Schätzer.

Beispiel 6.3: Verzerrter, aber konsistenter Schätzer.

Das Mittel $\tilde{X} = \frac{1}{n+3} \sum_{i=1}^n X_i$ ist ein verzerrter, jedoch konsistenter Schätzer für den Erwartungswert.

$$E(\tilde{X}) = \frac{n}{n+3} \mu \neq \mu$$

Beispiel 6.3: Fortsetzung.

Überprüfung der Bedingung (a):

$$\lim_{n \rightarrow \infty} [E(\tilde{X}) - \mu]^2 = \lim_{n \rightarrow \infty} \left(\frac{n^2}{(n+3)^2} - 2\frac{n}{n+3} + 1 \right) \cdot \mu^2 = 0$$

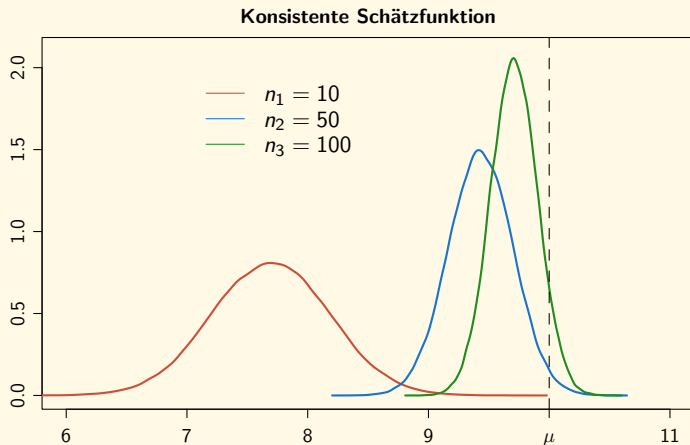
Überprüfung der Bedingung (b):

$$\lim_{n \rightarrow \infty} \text{Var}(\tilde{X}) = \lim_{n \rightarrow \infty} \frac{n}{(n+3)^2} \sigma^2 = 0$$

Beispiel 6.3: Fortsetzung.

- In der folgenden Abbildung werden die Verteilungen von \tilde{X} für verschiedene Stichprobengrößen aus $N(10, 4)$ dargestellt.
- Für die Stichprobenumfänge gilt $n_1 < n_2 < n_3$.
- Mit zunehmendem n verringern sich Verzerrung und Varianz.
- Die Stichprobenverteilung bewegt sich in Richtung $\mu = 10$.

Beispiel 6.3: Fortsetzung.



Beispiel 6.3: Fortsetzung.

```
# Definiere eigene verzerrte "mean" Funktion
mean_biased <- function(x) {
  sum(x) / (length(x) + 3)
}
# Ziehe 100000 Stichproben
smpls <- replicate(100000, rnorm(n, 10, 2))
# Berechne für jede Stichprobe das verzerrte Mittel
xtilde <- apply(smpls, 2, function(x) mean_biased(x))
# Plot
hist(xtilde, freq = F, col = "lightblue", breaks = 40, ylab = "Dichte")
abline(v = 10, col = "red", lwd = 2) # Wahres Mittel
```

Aufgabe 6.4:

Für den unbekannten Erwartungswert einer Grundgesamtheit liegen folgende Schätzfunktionen vor (Annahme: einfache Zufallsstichproben)

$$\hat{\mu}_1 = \frac{3}{7}X_1 + \frac{2}{7}X_2 + \frac{1}{7}X_3 + \frac{1}{7}X_4,$$

$$\hat{\mu}_2 = \frac{1}{8}X_1 + \frac{1}{4}X_2 + \frac{1}{2}X_3 + \frac{1}{8}X_4,$$

$$\hat{\mu}_3 = \frac{1}{4}(X_1 + X_2 + X_3 + X_4) + a, a \neq 0.$$

- (a) Welche der Schätzfunktionen ist erwartungstreu?
- (b) Welche der erwartungstreuen Schätzfunktionen ist die effizienteste?

- Es gibt verschiedene Vorgehensweisen für die Konstruktion von Schätzern, die jeweils eine Klasse von Schätzverfahren festlegen.
- Die wichtigsten sind
 - (1) die Momentenmethode (vgl. Anhang),
 - (2) die Methode der kleinsten Quadrate und
 - (3) die Maximum-Likelihood-Methode.

Beispiel 6.5: OLS-Schätzung für μ .

Der unbekannte Parameter μ einer Grundgesamtheit soll mit der Methode der kleinsten Quadrate geschätzt werden. Daher können wir schreiben

$$\hat{S} = \sum_{j=1}^n (X_j - \hat{\mu})^2 \longrightarrow \underset{\hat{\mu}}{\text{Min!}}$$

Die erste und zweite Ableitung von \hat{S} nach $\hat{\mu}$ lauten

$$\frac{d\hat{S}}{d\hat{\mu}} = -2 \sum_{j=1}^n (X_j - \hat{\mu}) = 0 \quad \text{und} \quad \frac{d^2\hat{S}}{d\hat{\mu}^2} = 2n > 0.$$

Die Lösung der ersten Ableitung ergibt den OLS-Schätzer

$$\hat{\mu} = \frac{1}{n} \sum_{j=1}^n X_j = \bar{X}.$$

Er minimiert aufgrund der positiven zweiten Ableitung \hat{S} .

- Die **Maximum-Likelihood-Methode** (ML-Methode) setzt voraus, dass die Verteilung von X bekannt ist, bis auf die Parameter

$$\theta = (\theta_1, \dots, \theta_K).$$

- Diese unbekannten Parameter werden so geschätzt, dass die Wahrscheinlichkeit für das Eintreten der bereits vorliegenden Stichprobenrealisationen (x_1, \dots, x_n) maximal wird.

- Da die Stichprobenvariablen unabhängig wie X verteilt sind, besitzen sie alle dieselben Wahrscheinlichkeits- bzw. Dichtefunktionen

$$f(X_1|\theta) = f(X_2|\theta) = \dots = f(X_n|\theta).$$

- Die gemeinsame Wahrscheinlichkeits- bzw. Dichtefunktion folgt nach Verallgemeinerung für den n -dimensionalen Fall als

$$f(X_1, X_2, \dots, X_n|\theta) = f(X_1|\theta)f(X_2|\theta) \cdot \dots \cdot f(X_n|\theta) = \prod_{j=1}^n f(X_j|\theta).$$

(Verallgemeinerung von (2.1))

- In der gemeinsamen Wahrscheinlichkeits- bzw. Dichtefunktion sind jetzt die Elemente des Vektors θ die Variablen.
- Man bezeichnet diese Funktion daher als **Likelihood-Funktion** L , die bezüglich θ zu maximieren ist.

$$L = f(\theta | x_1, \dots, x_n) \longrightarrow \underset{\theta}{\text{Max!}}$$

- Die Lösung dieses Problems liefert den **ML-Schätzer** $\hat{\theta}$.
- Unter geeigneten Annahmen sind ML-Schätzer konsistent, mindestens asymptotisch erwartungstreu und mindestens asymptotisch effizient.

Beispiel 6.6: Geometrische Verteilung (1/2).

Die Likelihood-Funktion ist mit (4.4)

$$L(p|x_1, \dots, x_n) = \prod_{i=1}^n (1-p)^{x_i} p$$

Logarithmieren von L liefert die zur Differentiation geeignetere Form

$$\ln L(p) = k \ln(1-p) + n \ln(p),$$

wobei $k = \sum_i x_i$. Die erste Ableitung ist

$$\frac{\partial \ln L}{\partial p} = -\frac{k}{1-p} + \frac{n}{p}.$$

Beispiel 6.6: Fortsetzung.

Nullsetzen und Auflösen liefert

$$\frac{k}{1-p} = \frac{n}{p}$$

bzw.

$$\frac{p}{n} + \frac{p}{k} = \frac{1}{k},$$

so dass

$$\hat{p}_{ML} = \frac{n}{k+n}.$$

Äquivalent, aber praktischer ist es, die Schritte in einer anderen Reihenfolge durchzuführen.

1. Dichte für ein einzelnes x_i logarithmieren
2. Logarithmierte Dichte ableiten
3. Von $i = 1, \dots, n$ summieren
4. Gleich Null setzen und auflösen

Aufgabe 6.7:

Gegeben sei eine Zufallsstichprobe vom Umfang n aus einer Grundgesamtheit mit Dichte

$$f(x; \lambda) = \lambda e^{-\lambda x}$$

Ermitteln Sie den ML-Schätzer für λ !

Hier ohne Beweis einige Beispiele für die oben diskutierten Konzepte.

- \bar{X} ist erwartungstreu für den Erwartungswert μ und mit $\text{Var}(\bar{X}) = \sigma^2/n$ darüber hinaus konsistent und effizient unter den unverzerrten Schätzern.
- Für $X \sim N(\mu, \sigma^2)$ ist

$$S_{\text{ML}}^2 = \frac{1}{n} \sum_{j=1}^n (X_j - \bar{X})^2$$

der ML-Schätzer der Varianz σ^2 .

- S_{ML}^2 ist verzerrt mit Erwartungswert $\sigma^2(n-1)/n$, asy. erwartungstreu und konsistent.
- Damit ist

$$S^2 = \frac{1}{n-1} \sum_{j=1}^n (X_j - \bar{X})^2$$

erwartungstreu und konsistent für die Varianz σ^2 . Für $X \sim N(\mu, \sigma^2)$ ist $\text{Var}(S^2) = 2\sigma^4/(n-1)$.

Siehe `AsyUnverzerrtheit.R`.

Aufgabe 6.8:

Zeigen Sie, dass \bar{X} im quadratischen Mittel nach μ konvergiert, wenn X_1, \dots, X_n eine einfache Zufallsstichprobe ist.

Aufgabe 6.9:

Ermitteln und vergleichen Sie den MSE von S^2 und S_{ML}^2 , wenn X_1, \dots, X_n eine einfache Zufallsstichprobe aus einer normalverteilten Grundgesamtheit ist.

Aufgabe 6.10:

Zeigen Sie, dass

(a) bei bekanntem Erwartungswert $E(X) = \mu$ der Grundgesamtheit

$$S^2 = \frac{1}{n} \sum_j (X_j - \mu)^2,$$

(b) bei unbekanntem Erwartungswert $E(X) = \mu$ der Grundgesamtheit

$$S^2 = \frac{1}{n-1} \sum_j (X_j - \bar{X})^2$$

eine erwartungstreue Schätzfunktion für die unbekannte Varianz von X in der Grundgesamtheit ist!

Aufgabe 6.11:

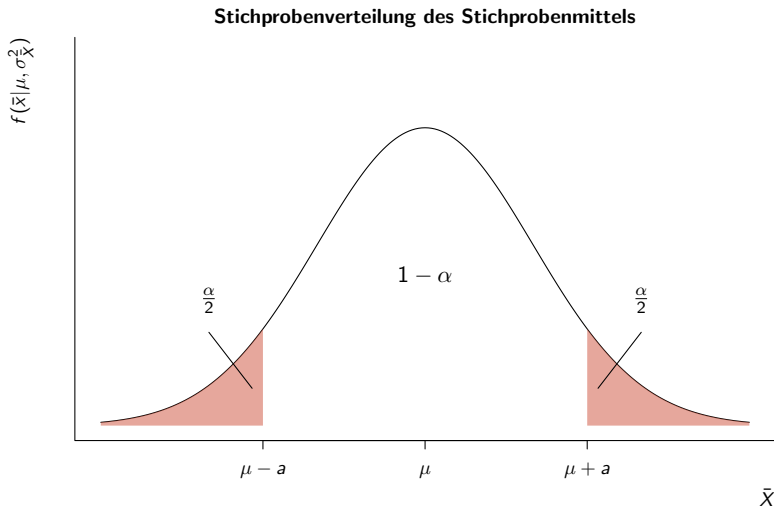
Im Rahmen eines Projekts soll die Körpergröße von Studenten untersucht werden. Hierzu wird eine Zufallsstichprobe im Umfang von $n = 10$ gezogen. Dies liefert folgende Körpergrößen in cm:

176, 180, 181, 168, 177, 186, 184, 173, 182, 177.

- (a) Schätzen Sie den unbekannten Erwartungswert der Grundgesamtheit μ !
- (b) Schätzen Sie die unbekannte Varianz der Grundgesamtheit σ^2 !

- Man bestimmt, ausgehend von den **bekannten** Parametern der Grundgesamtheit ein Intervall, aus dem die Stichprobengröße mit hoher, vorgegebener Wahrscheinlichkeit Werte annimmt.
- Ein solches Intervall nennt man **Schwankungsintervall**.
- Ein **zentrales** Schwankungsintervall ist dabei symmetrisch um μ konstruiert.
- Die vorgegebene, hohe Wahrscheinlichkeit bezeichnet man mit $1 - \alpha$.
- Hierzu benötigen wir die Verteilung der Stichprobengröße.
- So ist bspw. \bar{X} bei einfachen Stichproben wegen (5.1) und (5.2) für hinreichend großes n dank des Zentralen Grenzwertsatzes (5.3) unter den dortigen Bedingungen annähernd normalverteilt mit

$$E(\bar{X}) = \mu \quad \text{und} \quad \text{Var}(\bar{X}) = \sigma_{\bar{X}}^2 = \frac{1}{n} \sigma^2.$$



- Die nicht rötliche Fläche entspricht der Wahrscheinlichkeit $1 - \alpha$, die rötlichen Flächen betragen jeweils $\alpha/2$.
- Wegen der Symmetrie der Normalverteilung wird die nicht schraffierte Fläche auf der Abszisse durch die Werte $\mu \pm a$ begrenzt.
- Damit gilt für die Wahrscheinlichkeit des Ereignisses $\mu - a \leq \bar{X} \leq \mu + a$

$$P(\mu - a \leq \bar{X} \leq \mu + a) = 1 - \alpha.$$

- Nach Standardisieren mit $Z = \frac{\bar{X} - \mu}{\sigma_{\bar{X}}}$ folgt

$$P\left(-\frac{a}{\sigma_{\bar{X}}} \leq \frac{\bar{X} - \mu}{\sigma_{\bar{X}}} \leq \frac{a}{\sigma_{\bar{X}}}\right) = P(-z \leq Z \leq z) = 1 - \alpha.$$

- Die Wahrscheinlichkeit $1 - \alpha$ legt die Werte $\pm z$ der standardnormalverteilten Zufallsvariablen Z fest. Es gilt

$$\pm z = \pm z_{1-\frac{\alpha}{2}}.$$

- Das Schwankungsintervall für \bar{X} lautet daher

$$P(\mu - \sigma_{\bar{X}} z_{1-\frac{\alpha}{2}} \leq \bar{X} \leq \mu + \sigma_{\bar{X}} z_{1-\frac{\alpha}{2}}) = 1 - \alpha$$

mit

$$\sigma_{\bar{X}} \stackrel{(5.2)}{=} \frac{\sigma}{\sqrt{n}}.$$

Beispiel 6.12: Schwankungsintervall des Stichprobenmittels.

Die Körpergröße deutscher Männer ist normalverteilt mit $\mu = 178$ cm und $\sigma^2 = 64$ cm². Um das Schwankungsintervall für die Stichprobenmittelwerte \bar{X} zu bilden, benötigt man $z_{1-\frac{\alpha}{2}}$.

Für $1 - \alpha = 0.95$ erhalten wir $z_{1-\frac{\alpha}{2}} = z_{0.975} = 1.96$ aus $\text{qnorm}(0.975) = 1.96$.

Für einfache Stichproben beträgt die Varianz bei einem Stichprobenumfang von $n = 100$ $\sigma_{\bar{X}}^2 = 64/100$ und die Standardabweichung $\sigma_{\bar{X}} = 0.8$.

Beispiel 6.12: Fortsetzung.

Die Gleichung

$$P(\mu - \sigma_{\bar{X}} z_{1-\frac{\alpha}{2}} \leq \bar{X} \leq \mu + \sigma_{\bar{X}} z_{1-\frac{\alpha}{2}}) = 1 - \alpha$$

geht daher über in

$$0.95 = P(178 - 1.96 \cdot 0.8 \leq \bar{X} \leq 178 + 1.96 \cdot 0.8) \quad \text{oder}$$

$$0.95 = P(176.432 \leq \bar{X} \leq 179.568).$$

Im berechneten Schwankungsintervall liegen 95% aller einfachen Stichprobenmittelwerte bei einem Stichprobenumfang von $n = 100$.

Schätzverfahren 1



- Ziel einer Intervallschätzung ist es, für einen unbekannten Parameter einer Grundgesamtheit ein Intervall so zu schätzen, dass es mit großer Wahrscheinlichkeit die unbekannte Größe einschließt.
- Ein solches Intervall nennt man **Konfidenzintervall**, die Wahrscheinlichkeit $1 - \alpha$ heißt **Konfidenzniveau**.
- Jedes berechnete Konfidenzintervall stellt eine **Intervallschätzung** für den enthaltenen Grundgesamtheitsparameter dar.

- Konfidenzintervalle haben formal dieselbe Struktur wie Schwankungsintervalle.
- Aus

$$P\left(-z_{1-\frac{\alpha}{2}} \leq Z \leq z_{1-\frac{\alpha}{2}}\right) = 1 - \alpha$$

folgt nach Substitution von Z durch $(\bar{X} - \mu)/\sigma_{\bar{X}}$

$$P\left(-z_{1-\frac{\alpha}{2}} \leq \frac{\bar{X} - \mu}{\sigma_{\bar{X}}} \leq z_{1-\frac{\alpha}{2}}\right) = 1 - \alpha$$

$$\Rightarrow P\left(-z_{1-\frac{\alpha}{2}}\sigma_{\bar{X}} \leq \bar{X} - \mu \leq z_{1-\frac{\alpha}{2}}\sigma_{\bar{X}}\right) = 1 - \alpha$$

$$\Rightarrow P\left(-\bar{X} - z_{1-\frac{\alpha}{2}}\sigma_{\bar{X}} \leq -\mu \leq -\bar{X} + z_{1-\frac{\alpha}{2}}\sigma_{\bar{X}}\right) = 1 - \alpha$$

$$\Rightarrow P\left(\bar{X} + z_{1-\frac{\alpha}{2}}\sigma_{\bar{X}} \geq \mu \geq \bar{X} - z_{1-\frac{\alpha}{2}}\sigma_{\bar{X}}\right) = 1 - \alpha$$

und schließlich

$$P\left(\bar{X} - z_{1-\frac{\alpha}{2}}\sigma_{\bar{X}} \leq \mu \leq \bar{X} + z_{1-\frac{\alpha}{2}}\sigma_{\bar{X}}\right) = 1 - \alpha.$$

Aufgabe 6.13:

In einer Brauerei ist eine Abfüllmenge für Bier auf eine Sollmenge von 100 l pro Faß eingestellt (Standardabweichung 0.3 l). Es ist möglich, dass es nach einiger Zeit zu Abweichungen von dieser eingestellten Sollmenge kommt, so dass die Maschine neu justiert werden muss. Bei einer zufälligen Stichprobe von 50 Fässern wurde eine durchschnittliche Abfüllmenge von 100.2 l gemessen.

Erstellen Sie ein 95%-Konfidenzintervall für μ , um zu überprüfen, ob die Maschine neu justiert werden muss!

- Das Konfidenzintervall kann nur für bekanntes σ^2 erstellt werden.
- Ist dies nicht der Fall, wird σ^2 über die Stichprobenvarianz geschätzt.
- Wird bei normalverteilten X_j in

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

σ^2 durch die Varianz einer Stichprobe ersetzt, folgt

$$Z(t) = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$

bei einfachen Stichproben einer **t-Verteilung** mit $n - 1$ Freiheitsgraden, siehe (4.6).

Das Konfidenzintervall folgt mit Symmetrie der t -Verteilung aus

$$P\left(-t(1 - \alpha/2, n - 1) \leq \frac{\bar{X} - \mu}{S/\sqrt{n}} \leq t(1 - \alpha/2, n - 1)\right) = 1 - \alpha$$

als

$$P\left(\bar{X} - t(1 - \alpha/2, n - 1) \frac{S}{\sqrt{n}} \leq \mu \leq \bar{X} + t(1 - \alpha/2, n - 1) \frac{S}{\sqrt{n}}\right) = 1 - \alpha.$$

Bei Stichprobenumfängen von $n \geq 30$ ist folgende **Approximation** möglich

$$P\left(\bar{X} - z_{1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}} \leq \mu \leq \bar{X} + z_{1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}}\right) = 1 - \alpha.$$

Siehe Konfidenzintervalle.R.

Beispiel 6.14: Konfidenzintervall für die Körpergröße (Fortsetzung Beispiel 6.12).

Für die Körpergröße deutscher Männer soll ein Konfidenzintervall bei einem Konfidenzniveau von $1 - \alpha = 0.95$ mit einer einfachen Stichprobe des Umfangs $n = 25$ erstellt werden. Das Stichprobenmittel lautet $\bar{x} = 178$ cm, die Stichprobenvarianzschätzung beträgt $s^2 = 49$ cm².

R liefert uns den kritischen Wert der t -Verteilung mit $1 - \frac{\alpha}{2} = 0.975$ und $n - 1 = 24$ Freiheitsgraden durch $\text{qt}(0.975, 24) = 2.0639$.

Beispiel 6.14: Fortsetzung.

Die Standardabweichung der Stichprobenverteilung ergibt sich als

$$\frac{s}{\sqrt{n}} = \frac{7}{5} = 1.4.$$

Dies führt zu dem Konfidenzintervall

$$\mu \in [178 - 2.0639 \cdot 1.4; 178 + 2.0639 \cdot 1.4]$$

oder

$$\mu \in [175.1105; 180.8895].$$

Beispiel 6.14: Fortsetzung.

```
library(MASS) # Mit Hilfe von mvrnorm werden normalverteilte Daten erzeugt,  
# deren empirische Parameter denen aus dem Beispiel entsprechen.  
n <- 25; x <- mvrnorm(n = n, empirical = TRUE, mu = 178, Sigma = 49)  
head(as.vector(x))  
  
## [1] 182.1 178.0 184.9 174.5 172.9 163.3  
  
# Alternative 1  
confint(object = (lm(x~1)), level = 0.95)  
  
##                2.5 % 97.5 %  
## (Intercept) 175.1  180.9  
  
# Alternative 2 (mit der t-Verteilung)  
c(mean(x)-sd(x)/sqrt(n)*qt(0.975,n-1), mean(x)+sd(x)/sqrt(n)*qt(0.975,n-1))  
  
## [1] 175.1 180.9  
  
# Mit der Normalverteilung  
c(mean(x)-sd(x)/sqrt(n)*qnorm(0.975), mean(x)+sd(x)/sqrt(n)*qnorm(0.975))  
  
## [1] 175.3 180.7
```

Schätzverfahren 2



Aufgabe 6.15:

Es wurden 200 Versicherte einer großen KfZ-Haftpflichtversicherung zufällig ausgewählt und nach ihrer täglichen Fahrleistung X befragt. Die Auswertung der Befragung ergab eine durchschnittliche Fahrleistung von $\bar{x} = 25$ km pro Tag und eine Varianz von $s^2 = 128$ km².

Stellen Sie das 95%-Konfidenzintervall für die durchschnittliche Fahrleistung μ eines Versicherten auf!

- Für die korrekte Interpretation von Konfidenzintervallen ist wichtig, dass nicht der Parameter, sondern die Grenzen des Intervalls die Zufallsvariablen sind, da sie von der Stichprobe abhängen.
- Würde man viele Stichproben ziehen und mit jeder ein Konfidenzintervall bestimmen, schließen ca. $(1 - \alpha)100\%$ der Konfidenzintervalle den Parameter ein, $\alpha \cdot 100\%$ aber nicht.
- Deshalb ist es sinnvoll $1 - \alpha$ groß zu wählen.

- Lässt sich ein Grundgesamtheitsparameter θ als Funktion T_θ der Anfangsmomente der Grundgesamtheit $M_\alpha = E(X^\alpha)$ angeben, erhält man

$$\theta = T_\theta(M_1, \dots, M_K).$$

- Die Schätzung erfolgt nun so, dass die Momente der Grundgesamtheit M_α durch die analogen, empirischen Momente m_α der Stichprobe ersetzt werden.
- Der Momentenschätzer lautet dann

$$\hat{\theta} = T_\theta(m_1, \dots, m_K) \quad \text{mit} \quad m_\alpha = \frac{1}{n} \sum_{j=1}^n X_j^\alpha \quad \text{für } \alpha = 1, \dots, K.$$

- Momentenschätzer sind unabhängig von der Verteilung der Grundgesamtheit immer konsistent.
- Dies liegt daran, dass die m_α mit wachsendem n in die Grundgesamtheitsmomente M_α , die auch **theoretische Momente** genannt werden, übergehen.
- Einige dieser Schätzer sind erwartungstreu. Andere, verzerrte Schätzer lassen sich mit geringem Aufwand in erwartungstreue überführen.
- Jedoch sind bis auf wenige Ausnahmen Momentenschätzer nicht effizient.

Beispiel 6.16: Momentenschätzer für μ .

In einer Grundgesamtheit ist μ unbekannt. Dieser Lageparameter entspricht dem Anfangsmoment erster Ordnung. Somit gilt: $\mu = E(X) = M_1$. Der Momentenschätzer lautet daher

$$\hat{\mu} = T_{\mu}(m_1) = m_1 \quad \text{mit} \quad m_1 = \frac{1}{n} \sum_{j=1}^n X_j \quad \text{oder} \quad \hat{\mu} = \frac{1}{n} \sum X_j = \bar{X}.$$

Liegt eine einfache Stichprobe vor, berechnet man mit ihren Realisationen das Stichprobenmittel und erhält eine Punktschätzung für μ .

- 1 Einleitung
- 2 Grundlagen
- 3 Eindimensionale Zufallsvariablen
- 4 Ausgewählte theoretische Verteilungen
- 5 Grundzüge der Stichprobentheorie
- 6 Statistische Schätzverfahren
- 7 Statistische Testverfahren**
- 8 Zweidimensionale Zufallsvariablen

- Mit der **Testtheorie** kann auf der Basis von Stichproben überprüft werden, ob Vorstellungen über eine Grundgesamtheit zu Recht bestehen, d.h. sich mit der Realität in Einklang befinden.
- Jede Annahme über die unbekannte Verteilung von X in einer Grundgesamtheit heißt **statistische Hypothese**.
- Statistische Hypothesen können sich auf die Parameter θ_k , $k = 1, \dots, K$, einer Verteilung oder auf die Verteilung selbst beziehen.
- Im ersten Fall heißen sie **Parameterhypothesen**, im zweiten Fall **Verteilungshypothesen**.
- Wir beschränken uns hier auf den ersten Fall.

- Parameterhypothesen legen für die unbekannten Verteilungsparameter numerische Werte fest.
- Geschieht dies für alle Parameter des Verteilungsgesetzes so, dass ihnen jeweils genau ein Wert zugeordnet wird, liegt eine **einfache Hypothese** vor.
- Gibt man für einige oder alle Parameter hingegen Intervalle an oder lässt einige gänzlich unbestimmt, heißt die Hypothese **zusammengesetzt**.

Beispiel 7.1: Hypothesen.

X sei normalverteilt mit unbekannten Parametern μ und σ^2 . Da für μ keine a priori Restriktionen existieren, ist der Parameterraum \mathbb{R} . Der Parameterraum für σ^2 ist $[0, \infty)$.

- Einfache Hypothese: $\mu = 2$ und $\sigma^2 = 4$.
- Gibt man den Wert für einen Parameter genau, für den anderen Parameter ein Intervall an, liegt eine zusammengesetzte Hypothese vor, z.B. $\mu = 2$ und $0 \leq \sigma^2 \leq 4$.

- Die Überprüfung einer statistischen Hypothese heißt **statistischer Test**, die dabei überprüfte Hypothese nennt man **Nullhypothese**, kurz H_0 .
- In der Nullhypothese werden den Parametern $\theta_k, k = 1, \dots, K$ Werte zugeordnet, die eine Teilmenge des Parameterraumes bilden.
- Alle nicht in der Nullhypothese enthaltenen Werte können zur Formulierung einer sogenannten **Alternativhypothese** genutzt werden.
- Null- und Alternativhypothese sind disjunkt.

- Man will z.B. oft Hypothesen über den **Erwartungswert** einer Grundgesamtheit testen.
- Bei einer **einfachen Hypothese** wird für μ der Wert μ_0 angenommen

$$H_0 : \mu = \mu_0.$$

- Zusammengesetzte **Nullhypothesen**:

$$H_0 : \mu \leq \mu_0 \quad \text{oder} \quad \mu \geq \mu_0.$$

- Auch die **Alternativhypothese** H_1 kann in einfacher oder zusammengesetzter Form vorliegen.
- Der allgemeinere Fall einer zusammengesetzten Alternativhypothese wird häufig angegeben als

$$H_1 : \mu > \mu_0 \quad \text{oder} \quad H_1 : \mu < \mu_0 \quad \text{oder} \quad H_1 : \mu \neq \mu_0.$$

- Die Vorgehensweisen beim **zweiseitigen** und beim **einseitigen Testen** ähneln sich sehr.
- Null- und Alternativhypothese beim zweiseitigen Test lauten

$$H_0 : \mu = \mu_0 \quad \text{und} \quad H_1 : \mu \neq \mu_0.$$

Beispiel 7.2: Anlagestrategie.

Sie wollen reich werden und haben sich dafür eine Handelsstrategie am Aktienmarkt überlegt. Der Erwartungswert der Rendite dieser Strategie sei μ .

Sie wollen testen, ob Ihnen die Strategie mehr einbringt als Null (oder aber eine sichere Anlage, oder ein Indexfonds, usw.).

Dann testen Sie die zusammengesetzte Hypothese

$$H_0 : \mu \leq 0$$

gegen

$$H_1 : \mu > 0.$$

- Um H_0 zu überprüfen, benötigt man eine **Teststatistik**.
- Diese ist eine Stichprobenfunktion und besitzt daher eine Verteilung.
- Einen konkreten **Prüfwert** erhält man, wenn in die Teststatistik die Stichprobenrealisationen eingesetzt werden.
- Für den Erwartungswert μ eignet sich das arithmetische Mittel $\bar{X} = \frac{1}{n} \sum_{j=1}^n X_j$ als Teststatistik.
- Erinnerung: Für die Stichprobenverteilung von \bar{X} sind weitere Informationen über die Verteilung von X in der Grundgesamtheit notwendig.
- Ist X normalverteilt mit bekannter Varianz σ^2 , so ist auch \bar{X} normalverteilt und **unter der Nullhypothese** (zusammengesetzte Hypothesen s.u.) gilt mit (5.1) und (5.2)

$$E(\bar{X}) = \mu_0 \quad \text{und} \quad \text{Var}(\bar{X}) = \frac{\sigma^2}{n}.$$

Beispiel 7.3: Anlagestrategie.

Sie testen Ihre Anlagestrategie, indem Sie sie an n Handelstagen ausprobieren und die durchschnittliche Rendite ausrechnen.

Dass Renditen normalverteilt sind, mag zutreffen oder auch nicht. Dass σ^2 bekannt ist, ist offenbar eine (vorläufige) vereinfachende Annahme.

Wenn Sie die Nullhypothese testen, dass Ihre Strategie nichts bringt und dem auch so ist, dann hat Ihre Strategie einen Erwartungswert von $\mu \leq 0$.

- Da die Teststatistik selbst eine Zufallsvariable ist, können Entscheidungen bei einem Hypothesentest nicht mit Sicherheit getroffen werden.
- Daher benötigt man ein Kriterium, mit dem Abweichungen der geschätzten Parameter vom Parameterwert **unter** H_0 noch als zufällig oder bereits als **signifikant** bewertet werden können.
- Gegeben ein $\alpha \in (0, 1)$ werden für alle möglichen Realisationen der Teststatistik

$$T(X_1, \dots, X_n)$$

zwei komplementäre Mengen I_α und $I_{1-\alpha}$ konstruiert.

- Wähle I_α so, dass die Wahrscheinlichkeit **unter** H_0 eine Realisation von T in I_α zu erhalten gleich α ist.

- Darüber hinaus sollen in I_α alle Realisationen von T liegen, die **unter H_0** geringe Eintrittswahrscheinlichkeit haben.
- Damit lässt sich nun eine Entscheidungsregel für Hypothesentests aufstellen.
 - ▶ Liegt der Prüfwert in I_α , wird H_0 **abgelehnt**.
 - ▶ Liegt er in $I_{1-\alpha}$, wird H_0 **nicht abgelehnt**.
- α ist also die Wahrscheinlichkeit H_0 abzulehnen, obwohl sie richtig ist (was der Tester jedoch nicht weiß).

Beispiel 7.4: Anlagestrategie.

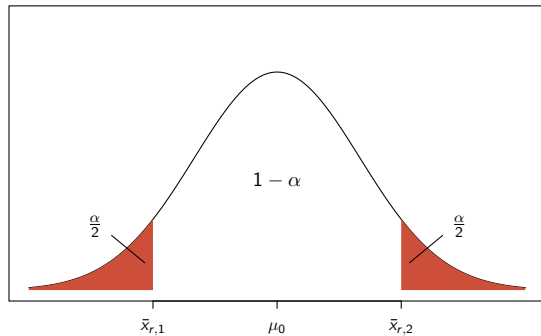
Wenn Sie ihre Strategie an anderen Handelstagen ausprobiert hätten, hätten Sie auch eine andere durchschnittliche Rendite bekommen.

Bloß weil Sie an ihr ein paar Euro verdient haben, würden Sie sich wohl kaum trauen, sich Millionen zu leihen und die Strategie in großem Stil zu implementieren. Die positive durchschnittliche Rendite könnte ja auch Zufall gewesen sein.

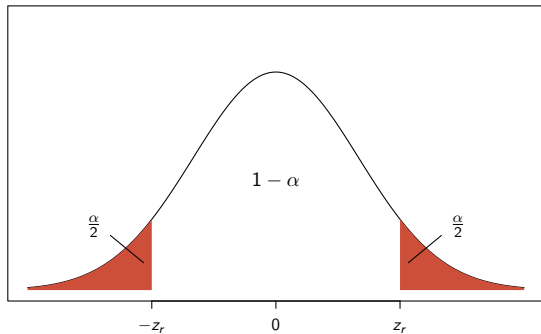
Wir suchen also eine Prüfgröße, die nur mit sehr geringer Wahrscheinlichkeit „extreme“ Werte (z.B. hohe Durchschnittsrenditen) annimmt, wenn Ihre Strategie nichts bringt.

- Man wird daher bei einem Test α sehr klein wählen, in der Regel 0.1, 0.05 oder 0.01.
- Tests mit diesem Aufbau heißen **Signifikanztests**.
- Man bezeichnet α daher als **Irrtumswahrscheinlichkeit**.
- Die alternative Bezeichnung „**Signifikanzniveau**“ für α resultiert daraus, dass die Stichprobengröße (etwa \bar{X}) signifikant vom Wert des Parameters aus H_0 abweichen muss, um in I_α zu liegen.

Null-Verteilung von \bar{X}



Null-Verteilung von Z



- Die linke Abbildung gibt die Normalverteilungsdichte der Teststatistik \bar{X} unter H_0 wieder. $\bar{x}_{r,1}$ und $\bar{x}_{r,2}$ sind die beiden **kritischen Werte**, die den **Annahmehereich** $I_{1-\alpha}$ und **Ablehnungsbereich** I_α trennen.
- Daneben ist dieselbe Testsituation dargestellt, jetzt jedoch mit der unter H_0 standardnormalverteilten Teststatistik

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}.$$

- Man kann den kritischen Wert $\bar{x}_{r,2}$ durch Auflösen der Gleichung $P(\bar{X} \leq \bar{x}_{r,2}) = 1 - \frac{\alpha}{2}$ erhalten (vgl. Buch S. 243).
- Der Test lässt sich aber schneller durchführen, wenn I_α in Bezug auf Z angegeben wird.

- Bei gegebener Wahrscheinlichkeit $1 - \alpha/2$ erhält man das $1 - \alpha/2$ -Quantil $z_{1-\alpha/2}$ mit dem R-Befehl `qnorm(p)` wobei $p = 1 - \alpha/2$.
- Der **Ablehnungsbereich** lautet

$$I_{\alpha}(z) = \{Z > z_{1-\alpha/2} \text{ oder } Z < z_{\alpha/2}\}.$$

- Liegt also die standardisierte Testgröße

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

in $I_{\alpha}(z)$, ist H_0 abzulehnen.

- In diesem Fall kann die Abweichung von \bar{x} zu μ_0 nicht als zufällig angesehen werden, sondern ist **(statistisch) signifikant**.

Signifikanztests 1



Beispiel 7.5: Parametertest bei normalverteilter Grundgesamtheit.

X sei normalverteilt mit $\sigma^2 = 36$. Es galt $E(X) = 40$. Trifft dies noch zu? Da die mögliche Richtung der Veränderung von $E(X)$ unklar ist, führen wir einen zweiseitigen Test durch.

- $H_0 : \mu = \mu_0 = 40$, $H_1 : \mu \neq 40$.
- Das Signifikanzniveau sei $\alpha = 0.05$.
- Eine Stichprobe im Umfang $n = 100$ liefere $\bar{x} = 42$.

Prüfe nun, ob die Abweichung $\bar{x} - \mu_0$ zufällig oder bei $\alpha = 0.05$ signifikant ist.

Die Statistik

$$\bar{X} = \frac{1}{n} \sum_{j=1}^n X_j$$

ist unter H_0 normalverteilt mit $E(\bar{X}) = 40$ und $\text{Var}(\bar{X}) = 36/100 = 0.36$.

Beispiel 7.5: Fortsetzung.

Den kritischen Wert $z_{r,\alpha/2}$ ermittelt man mit der Gleichung

$$P(Z \leq z_{r,\alpha/2}) = 0.975.$$

Der R-Befehl `qnorm(0.975)` liefert $z_{0.975} = 1.96$. Der Annahmebereich für die Teststatistik

$$Z = \frac{\bar{X} - 40}{0.6}$$

ist

$$I_{1-\alpha}(z) = \{-1.96 < Z < 1.96\}.$$

Der Prüfwert $z = \frac{42-40}{0.6} = 3.\bar{3}$ liegt nicht in $I_{1-\alpha}(z)$.

Beispiel 7.5: Fortsetzung.

```
library(MASS)
x <- mvrnorm(n = 100, empirical = TRUE, mu = 42, Sigma = 36)
head(as.vector(x))

## [1] 44.95 42.67 42.65 38.30 41.16 34.09

t.test(x, mu = 40, alternative = "two.sided", conf.level = 0.95)

##
##  One Sample t-test
##
## data:  x
## t = 3.3, df = 99, p-value = 0.001
## alternative hypothesis: true mean is not equal to 40
## 95 percent confidence interval:
##  40.81 43.19
## sample estimates:
## mean of x
##      42
```

Auf die Bedeutung von $df=99$ gehen wir später noch ein.

Beispiel 7.6: Dow Jones Returns.

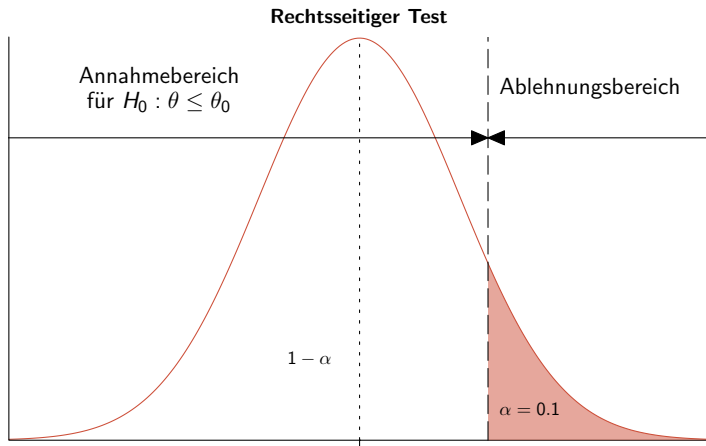
```
library(quantmod)
getSymbols('DJIA', src = 'FRED')

## [1] "DJIA"

DJIA    <- DJIA[!is.na(DJIA)]
returns <- diff(DJIA)/DJIA
t.test(x = as.vector(returns), mu = 0)

##
## One Sample t-test
##
## data:  as.vector(returns)
## t = 1.2, df = 2517, p-value = 0.2
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
## -0.0001675  0.0007030
## sample estimates:
## mean of x
## 0.0002677
```

- Zusammengesetzte Nullhypothesen der Form $\mu \leq \mu_0$ oder $\mu \geq \mu_0$ lassen den für die Verteilung der Teststatistik unter H_0 benötigten Parameter μ unbestimmt.
- In beiden Fällen setzt man $\mu = \mu_0$, um den kritischen Bereich festzulegen.
- Zusammengesetzte Nullhypothesen führen zu **einseitigen Tests**.
 - ▶ Für $H_0 : \mu \leq \mu_0$ ist die Alternativhypothese $H_1 : \mu > \mu_0$ und es wird **rechtsseitig** getestet.
 - ▶ Mit $H_0 : \mu \geq \mu_0$ korrespondiert die Alternativhypothese $H_1 : \mu < \mu_0$, und ein **linksseitiger** Test.



- Bei einem einseitigen Test werden die **kritischen Werte** nicht mit $\alpha/2$, sondern mit α berechnet.
- Die hierfür relevante Gleichung lautet jetzt

$$P(\bar{X} \leq \bar{x}_{r,2}) = P(Z \leq z_{1-\alpha}) = 1 - \alpha.$$

- Der kritische Bereich für einen **rechtsseitigen Test** ist also

$$I_{\alpha}(z) = \{z_{1-\alpha} < Z\}.$$

- Bei symmetrischer Verteilung unter H_0 ist der kritische Bereich für einen **linksseitigen Test** dann einfach

$$I_{\alpha}(z) = \{Z < z_{\alpha}\}$$

Aufgabe 7.7:

Von einer Brotfirma werden Brötchen produziert. Die Firma verspricht, dass das durchschnittliche Gewicht der Brötchen 50 Gramm beträgt. Das Brötchengewicht X sei normalverteilt mit einer Varianz von 1.44 g^2 .

Bei einer Stichprobe von 25 Brötchen ist ein mittleres Gewicht von $\bar{x} = 49.5\text{g}$ festgestellt worden.

Stützt der Stichprobenmittelwert \bar{x} bei einem Signifikanzniveau von 5% die Angaben der Brotfirma?

Aufgabe 7.8:

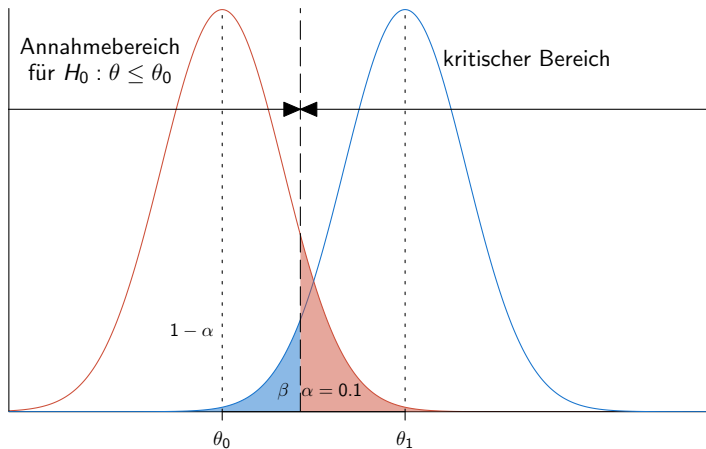
Bei einem Test einer neu angeschafften Flaschenabfüllanlage wird der Brauerei eine mittlere Abfüllmenge von 0.5 l bei einer Standardabweichung von $\sigma = 0.03$ l garantiert. Eine Stichprobe von $n = 40$ Flaschen aus einer 10 minütigen Produktionsgesamtheit von 500 Flaschen ergab eine durchschnittliche Biermenge pro Flasche von 0.49 l.

- (a) Kann man mit einer Irrtumswahrscheinlichkeit von 5% davon ausgehen, dass die mittlere Abfüllmenge geringer ist als vom Hersteller der Anlage angegeben?
- (b) Bei welchem α -Niveau wird die unter (a) betrachtete Hypothese noch abgelehnt?

- Da beim statistischen Testen Entscheidungen nicht unter Sicherheit getroffen werden, besteht immer die Möglichkeit einer Fehlentscheidung.
- Der α -Fehler, auch **Fehler erster Art** genannt, entsteht, wenn H_0 abgelehnt wird, obwohl sie richtig ist.
- Dies ist mit der Wahrscheinlichkeit α der Fall, daher der Name.
- Ist H_0 falsch und fiel die Testgröße in den Annahmebereich, wird H_0 nicht verworfen.
- Diesen Fehler nennt man β -Fehler oder **Fehler zweiter Art**.
- Die Wahrscheinlichkeit für den β -Fehler entspricht (bei einem rechtsseitigen Test) der blauen Fläche in der Abbildung auf der nächsten Folie.

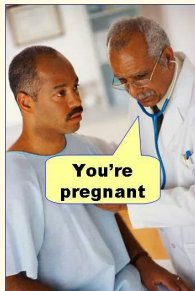
Signifikanztests

α - und β -Fehler

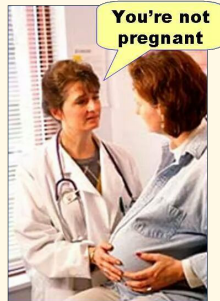


Beispiel 7.9: Nullhypothese: nicht schwanger.

Type I error
(false positive)



Type II error
(false negative)



Entscheidung \ Realität	H_0 richtig	H_1 richtig
	H_0 nicht verwerfen	H_0 ablehnen
H_0 nicht verwerfen	richtig $1 - \alpha$	falsch β -Fehler
H_0 ablehnen	falsch α -Fehler	richtig $1 - \beta$

Beispiel 7.10: Anlagestrategie.

Es kann also sein, dass ihr Test sagt, dass Ihre Strategie Sie reich macht, obwohl das gar nicht stimmt, weil das unwahrscheinliche, aber nicht unmögliche Ereignis eingetreten ist, dass Ihre Anlagestrategie an den n Testtagen zufällig sehr gut klappte, obwohl sie langfristig nichts bringt.

Umgekehrt kann es sein, dass Sie tatsächlich den Schlüssel zum Geld gefunden haben, das aber in den n Testtagen noch nicht deutlich wird.

Oder der Test liegt richtig: Ihre n Testtage machen klar, dass die Idee doch nicht besonders toll war, oder aber bestätigen, dass sie reich werden.

- Ein „guter“ Test sollte also kleine α - und β -Fehler haben.
- Die Abbildung auf Folie 28 zeigt, dass eine Verringerung des α -Fehlers (Verschiebung des kritischen Wertes nach rechts) zu einer Erhöhung des β -Fehlers führt.
- Die Wahrscheinlichkeit eine falsche Nullhypothese abzulehnen ist $1 - \beta$ und wird auch als **Macht** bzw. **Power** des Tests bezeichnet.

Beispiel 7.11: Parametertest bei normalverteilter Grundgesamtheit.

Um bei obigem Test den β -Fehler zu quantifizieren, müssen wir uns aus der Alternativhypothese $H_1 : \mu \neq \mu_0$ einen Punkt herausuchen. Hier bietet sich bspw. die Punktschätzung an. Dann ist $\mu_1 = 42$.

Es kann aber jeder Wert aus H_1 eingesetzt werden. Es ist wichtig zu verstehen, dass wir den wahren Wert μ nicht kennen und damit zwar verschiedene β -Fehler berechnen können, nicht jedoch wissen, welcher der „richtige“ ist.

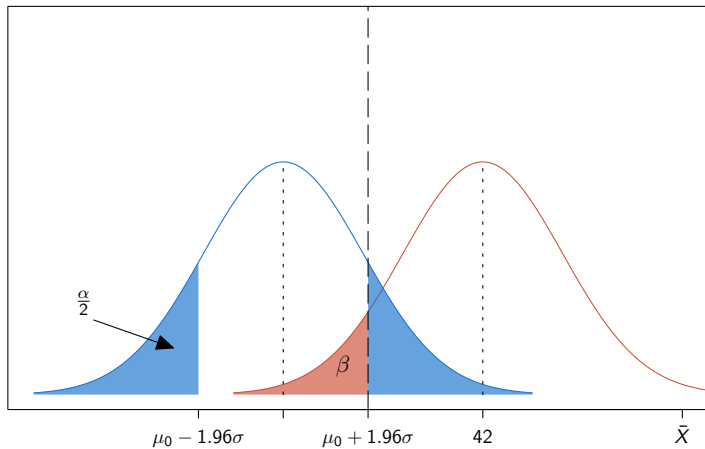
Beispiel 7.11: Fortsetzung.

Damit ist der β -Fehler quantifizierbar, der in der Abbildung auf der nächsten Folie als rote Fläche ausgewiesen ist. Er beträgt

$$\begin{aligned} & P(\mu_0 - 1.96\sigma \leq \bar{X} \leq \mu_0 + 1.96\sigma | H_1 = 42) \\ &= P\left(\frac{38.824 - 42}{0.6} \leq Z \leq \frac{41.176 - 42}{0.6} \middle| H_1 = 42\right) \\ &= P(-5.293 \leq Z \leq -1.373 | H_1 = 42) \\ &= F(5.293) - F(1.373) \approx 1 - 0.9152 = 0.0848. \end{aligned}$$

Der β -Fehler von 0.0848 gibt die Wahrscheinlichkeit an, dass die falsche Nullhypothese nicht verworfen wird, wenn das wahre μ tatsächlich gleich 42 ist.

α - und β -Fehler beim zweiseitigen Test



- Für die Durchführung von Signifikanztests kann ein Standardschema angegeben werden, das vier Testschritte umfasst:
 - (1) Spezifikation der Null- und Alternativhypothese,
 - (2) Festlegung des Signifikanzniveaus,
 - (3) Auswahl der Teststatistik,
 - (4) Berechnung des Testwertes und Entscheidung.
- Die beiden ersten Schritte ergeben sich aus der inhaltlichen Problemstellung.

Signifikanztests 2



- Ist die Varianz von X unbekannt, muss sie durch S^2 geschätzt werden.
- Bei normalverteiltem X ist $Z(t)$ t -verteilt (vgl. (21)),

$$Z(t) = \frac{\bar{X} - \mu}{S/\sqrt{n}}.$$

- D.h., bei geschätzter Varianz liefern die Quantile der t -Verteilung die kritischen Werte.
- Bei hinreichend großem n können **auch ohne Kenntnis der Verteilung** von X Signifikanztests durchgeführt werden.
- In diesem Fall ist die Teststatistik wegen des zentralen Grenzwertsatzes approximativ normalverteilt.
- In der Praxis nimmt man oft $n \geq 30$, auch wenn das streng genommen nur den Übergang von der t - in die $N(0, 1)$ -Verteilung beschreibt.

Beispiel 7.12: Parametertest bei unbekannter Varianz.

In einer Grundgesamtheit ist X normalverteilt, σ^2 unbekannt. Man vermutet, dass $E(X)$ höchstens 20 ist. Mit einem rechtsseitigen Test soll $H_0 : \mu \leq 20$ bei einem Signifikanzniveau $\alpha = 0.01$ überprüft werden.

Eine Stichprobe mit $n = 25$ liefert $\bar{x} = 21.5$ und $s^2 = 9$. Bei $n - 1 = 24$ Freiheitsgraden und $\alpha = 0.01$ beträgt der kritische Wert $t(0.99, 24) = 2.4922$, welcher sich mit dem R-Befehl `qt(p = 0.99, df = 24)` errechnen lässt.

Beispiel 7.12: Fortsetzung.

Der Ablehnungsbereich ist

$$I_\alpha = \{Z(t) \geq 2.4922\}.$$

Die standardisierte Testgröße

$$\frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{21.5 - 20}{3/5} = 2.5$$

liegt in I_α . H_0 wird also abgelehnt.

```
library(MASS)
x <- mvrnorm(n = 25, empirical = TRUE, mu = 21.5, Sigma = 9)
t.test(x, mu = 20, alternative = "greater", conf.level = 0.99)
```

```
##
##  One Sample t-test
##
## data:  x
## t = 2.5, df = 24, p-value = 0.01
## alternative hypothesis: true mean is greater than 20
## 99 percent confidence interval:
##    20 Inf
## sample estimates:
## mean of x
##      21.5
```

Aufgabe 7.13:

Auf Verpackungen eines Batterietyps wird eine durchschnittliche Lebensdauer von 280 Stunden angegeben. Eine Qualitätskontrolle ($n = 36$) führt zu folgenden Ergebnissen:

269	300	268	278	282	263
301	295	288	278	276	286
296	265	271	279	284	260
275	282	260	266	270	293
272	285	293	281	269	299
263	264	273	291	274	277

- (a) Muss der Hersteller die Angaben korrigieren (Signifikanzniveau 1%)?
- (b) Wie groß ist die Wahrscheinlichkeit, dass die Nullhypothese nicht verworfen wird, wenn sie falsch ist? (Arbeiten Sie mit \bar{x} .)

Siehe Buch S. 249.

- 1 Einleitung
- 2 Grundlagen
- 3 Eindimensionale Zufallsvariablen
- 4 Ausgewählte theoretische Verteilungen
- 5 Grundzüge der Stichprobentheorie
- 6 Statistische Schätzverfahren
- 7 Statistische Testverfahren
- 8 Zweidimensionale Zufallsvariablen

- Bis jetzt haben wir nur ein Merkmal betrachtet, das in eine eindimensionale Zufallsvariable überführt wurde.
- Besitzen die Ausgänge ω_j aber $K \geq 2$ interessierende Merkmale, kann für jedes eine eigene Zufallsvariable X_1, \dots, X_K eingeführt werden.
- Wir betrachten im Folgenden zwei interessierende Merkmale.
- Jedes der beiden Merkmale stellt eine Zufallsvariable dar.
- Zusammen ergeben sie die zweidimensionale Zufallsvariable

$$[X(\omega), Y(\omega)],$$

die jeden Ausgang ω in genau ein geordnetes Zahlenpaar abbildet.

Definition 8.1: Zweidimensionale Zufallsvariable.

Die Abbildung (X, Y) , die jeden Ausgang $\omega \in \Omega$ in ein geordnetes Zahlenpaar $(x, y) \in \mathbb{R}^2$ abbildet, also

$$(X, Y) : \Omega \longrightarrow \mathbb{R}^2$$

heißt **zweidimensionale Zufallsvariable**, wenn

$$\{\omega \in \Omega \mid X(\omega) \leq x \quad \text{und} \quad Y(\omega) \leq y\}$$

ein Ereignis für alle $(x, y) \in \mathbb{R}^2$ darstellt.

Beispiel 8.2: Haushaltsmerkmale.

Aus dem Einwohnermelderegister einer Stadt wird ein Haushalt zufällig ausgewählt. Der ausgewählte Haushalt ist Ausgang eines Zufallsvorgangs.

Die beiden interessierenden Merkmale sind die Haushaltsgröße (Anzahl der Haushaltsmitglieder, X) und das Haushaltsjahreseinkommen, Y .

Die zweidimensionale Zufallsvariable (X, Y) bildet jeden Ausgang ω (=ausgewählter Haushalt) in ein Zahlenpaar ab,

$$[x = X(\omega), y = Y(\omega)].$$

Beispiel 8.3: Haushaltsmerkmale.

Für die Wahrscheinlichkeit, dass ein zufällig ausgewählter Haushalt aus höchstens drei Mitgliedern besteht und sein Jahreseinkommen weniger als 50,000 € beträgt, benötigen wir die Verteilungsfunktion einer zweidimensionalen Zufallsvariablen.

Definition 8.4: Zweidimensionale Verteilungsfunktion.

Eine Funktion $F : \mathbb{R}^2 \rightarrow \mathbb{R}$ mit

$$F(x, y) = P(X \leq x, Y \leq y)$$

heißt **zweidimensionale Verteilungsfunktion**.

Für diese gelten ähnliche Eigenschaften wie für den eindimensionalen Fall, siehe Buch S. 152.

- Auch bei zweidimensionalen ZV unterscheiden wir, ob X und Y beide diskret, beide stetig oder ob sie gemischt stetig-diskret sind.
- Wir beschränken uns im Folgenden auf stetige ZV. Die Ausführungen übertragen sich recht analog auf diskrete ZV.
- Diese sind formal wiederum sehr ähnlich zu gemeinsamen relativen Häufigkeiten aus der deskriptiven Statistik.

Bei **stetigen**, zweidimensionalen Zufallsvariablen besteht der Wertebereich der (X, Y) aus überabzählbar unendlich vielen Zahlenpaaren $(x, y) \in \mathbb{R}^2$.

Definition 8.5: gemeinsame Dichtefunktion.

Die Funktion $f : (X, Y) \rightarrow \mathbb{R}^+$ heißt **gemeinsame Dichte(funktion)**, wenn für alle $(x, y) \in \mathbb{R}^2$ gilt

$$F(x, y) = \int_{-\infty}^y \int_{-\infty}^x f(u, v) du dv.$$

Für alle Punkte (x, y) , in denen $F(x, y)$ stetig ist, gilt

$$\frac{\partial^2 F}{\partial x \partial y} = f(x, y).$$

Beispiel 8.6:

$$f(x, y) = \begin{cases} xy & 0 < x < 2 \text{ und } 0 < y < 1 \\ 0 & \text{sonst.} \end{cases}$$

Über diesem Bereich beträgt das Volumen unter der Dichtefunktion eins

$$\int_0^1 \int_0^2 xy \, dx \, dy = \int_0^1 \left. \frac{y}{2} x^2 \right|_0^2 dy = \int_0^1 2y \, dy = \left. y^2 \right|_0^1 = 1.$$

Beispiel 8.6: Fortsetzung.

$F(x, y)$ erhält man aus $f(x, y)$ durch Integration, wobei u die Integrationsvariable für x und v die Integrationsvariable für y ist,

$$F(x, y) = \int_0^y \int_0^x uv \, du \, dv = \int_0^y \left. \frac{v}{2} u^2 \right|_0^x dv = \int_0^y \frac{v}{2} x^2 \, dv = \left. \frac{1}{4} x^2 v^2 \right|_0^y = \frac{1}{4} x^2 y^2.$$

In ausführlicher Notation

$$F(x, y) = \begin{cases} 0 & x \leq 0 \text{ und/oder } y \leq 0 \\ \frac{1}{4} x^2 y^2 & 0 < x < 2 \text{ und } 0 < y < 1 \\ 1 & x \geq 2 \text{ und } y \geq 1. \end{cases}$$

Beispiel 8.6: Fortsetzung.

Die Kreuzableitung der Verteilungsfunktion $F(x, y)$ liefert $f(x, y)$

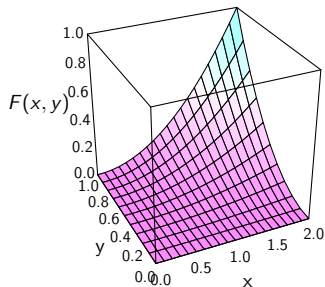
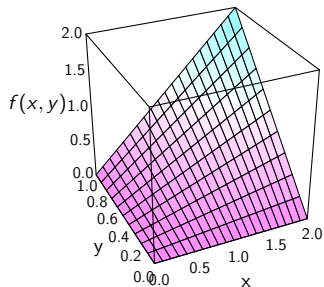
$$\frac{\partial^2 F(x, y)}{\partial y \partial x} = \frac{\partial}{\partial y} \left(\frac{1}{2} xy^2 \right) = xy.$$

Zur Berechnung von $P(0 < X < 1, 0 < Y < 1)$ verwendet man $F(x, y)$.

$$P(0 < X < 1, 0 < Y < 1) = F(1, 1) = \frac{1}{4}.$$

Zweidimensionale Zufallsvariable

Dichtefunktion/Verteilungsfunktion



Definition 8.7: Randdichte einer stetigen Zufallsvariable.

Die Funktion

$$f_X(x) \longrightarrow \mathbb{R} \quad \text{bzw.} \quad f_Y(y) \longrightarrow \mathbb{R}$$

mit

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy \quad \text{bzw.} \quad f_Y(y) = \int_{-\infty}^{\infty} f(x, y) dx$$

heißt **Randdichtefunktion** von X bzw. Y .

- Im letzten **Beispiel** erhält man die **Randdichte** für X als

$$f_X(x) = \int_0^1 xy \, dy = \frac{x}{2} y^2 \Big|_0^1 = \frac{1}{2}x.$$

- Hierfür gilt, wie üblich,

$$\int_0^2 f_X(x) \, dx = \int_0^2 \frac{1}{2}x \, dx = 1.$$

- Die **Randverteilungsfunktion** ergibt sich, auch wie üblich, aus $f_X(x)$ als

$$F_X(x) = \int_0^x \frac{1}{2}u \, du = \frac{1}{4}u^2 \Big|_0^x = \frac{1}{4}x^2.$$

Definition 8.8: Stochastische Unabhängigkeit bei stetigen, zweidimensionalen Zufallsvariablen.

Die stetigen Zufallsvariablen X und Y heißen **stochastisch unabhängig**, wenn für alle Paare $(x, y) \in \mathbb{R}^2$ gilt

$$f(x, y) = f_X(x)f_Y(y) \quad \text{bzw.} \quad F(x, y) = F_X(x)F_Y(y).$$

Beispiel 8.9:

Die Randdichtefunktion für Y ist

$$f_Y(y) = \int_0^2 xy \, dx = \frac{y}{2} x^2 \Big|_0^2 = 2y.$$

Also ist $f(x, y) = xy = 2y \frac{x}{2} = f_Y(y)f_X(x)$.

Für X gilt exemplarisch

$$\begin{aligned} E(X) &= \int_x x f_X(x) dx = \int_x x \left(\int_y f(x, y) dy \right) dx \\ &= \int_x \int_y x f(x, y) dy dx. \end{aligned}$$

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

$$E(X^2) = \int_x x^2 f_X(x) dx = \int_x \int_y x^2 f(x, y) dx dy$$

Zweidimensionale Zufallsvariable 1



Aufgabe 8.10:

Gegeben ist folgende Funktion:

$$f(x, y) = \begin{cases} \frac{x + 3y}{2} & 0 < x < 1, \ 0 < y < 1 \\ 0 & \text{sonst.} \end{cases}$$

- (a) Überprüfen Sie, ob $f(x, y)$ eine Dichtefunktion ist!
- (b) Leiten Sie die gemeinsame Verteilungsfunktion $F(x, y)$ her!
- (c) Bestimmen Sie die Randdichten von X und Y !
- (d) Bestimmen Sie den Erwartungswert und die Varianz von X und Y !
- (e) Sind X und Y stochastisch unabhängig?
- (f) Bestimmen Sie für $Z = X + Y$ den Erwartungswert $E(Z)$ und die Varianz $\text{Var}(Z)$!

Definition 8.11: Kovarianz.

Das **gemischte Zentralmoment**

$$\sigma_{X,Y} = \text{cov}(X, Y) = E\{[X - E(X)][Y - E(Y)]\} = E[(X - \mu_X)(Y - \mu_Y)]$$

heißt **Kovarianz** (cov) von X und Y .

- Die Berechnung der Kovarianz vereinfacht sich bei Anwendung von

$$\text{cov}(X, Y) = E(XY) - E(X)E(Y).$$

- Diese Vereinfachung lässt sich zeigen mit

$$\begin{aligned}\text{cov}(X, Y) &= E[XY - XE(Y) - E(X)Y + E(X)E(Y)] \\ &= E(XY) - E(X)E(Y).\end{aligned}$$

- Den Erwartungswert $E(XY)$ ermittelt man als

$$E(XY) = \int_y \int_x xyf(x, y) dx dy.$$

- Sind X und Y unabhängig, vereinfacht sich die Gleichung wegen

$$f(x, y) = f_X(x)f_Y(y).$$

- Damit können wir schreiben

$$\begin{aligned} E(XY) &= \int_x \int_y xyf(x, y) dx dy = \int_x xf_X(x) dx \int_y yf_Y(y) dy \\ &= E(X)E(Y). \end{aligned}$$

- Stochastische Unabhängigkeit impliziert also eine Kovarianz von null. Die Umkehrung gilt nicht allgemein!
- Eine Kovarianz von null bedeutet lediglich, dass für beide Variablen kein linearer Zusammenhang vorliegt.

Beispiel 8.12: stochastische Abhängigkeit.

Sei $X \sim N(0, 1)$. Dann sind X und X^2 **unkorreliert**, da

$$\text{cov}(X, X^2) = E(X^3) - E(X)E(X^2) = 0 - 0 = 0$$

Auch ohne Rechnung ist aber klar, dass X und X^2 nicht unabhängig sind.

Zweidimensionale Zufallsvariable 2



- Standardisieren von X und Y liefert

$$Z_X = \frac{X - E(X)}{\sigma_X} \quad \text{und} \quad Z_Y = \frac{Y - E(Y)}{\sigma_Y}.$$

- Die Kovarianz der standardisierten Variablen folgt wegen

$$E(Z_X) = E(Z_Y) = 0$$

als

$$\begin{aligned} \text{cov}(Z_X, Z_Y) &= E(Z_X Z_Y) = E \left[\frac{X - E(X)}{\sigma_X} \frac{Y - E(Y)}{\sigma_Y} \right] \\ &= \frac{E\{[X - E(X)][Y - E(Y)]\}}{\sigma_X \sigma_Y} = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y} = r, \end{aligned}$$

bekannt als Korrelationskoeffizient.

- Bei exakt linearer, positiver Abhängigkeit zwischen X und Y gilt $r = 1$, bei exakt linearer, negativer Abhängigkeit $r = -1$.
- Eine Kovarianz von null impliziert $r = 0$.

Auch hier gilt: Unkorreliertheit bedeutet nicht Unabhängigkeit.

Beispiel 8.13:

Es gelte

$$f(x, y) = x + y \quad \text{wenn } 0 < x < 1 \text{ und } 0 < y < 1,$$

0 sonst. Auf $[0, 1]$ sind die Randdichten dann $f_X(x) = x + 0.5$ bzw. $f_Y(y) = y + 0.5$. X und Y sind nicht unabhängig, da $f(x, y) \neq f_X(x)f_Y(y)$.

Damit folgen $E(X) = E(Y) = 7/12$ und $E(X^2) = E(Y^2) = 5/12$, so dass $\text{Var}(X) = \text{Var}(Y) = 11/144$.

Ferner ist

$$E(XY) = \int_0^1 \int_0^1 xy(x + y) dx dy = 1/3,$$

so dass $\text{cov}(X, Y) = -1/144$ und

$$r = \frac{-\frac{1}{144}}{\sqrt{\frac{11}{144} \frac{11}{144}}} = -\frac{1}{11}$$

Aufgabe 8.14:

Gegeben ist folgende Funktion:

$$f(x, y) = \begin{cases} \frac{x + 3y}{2} & 0 < x < 1, \quad 0 < y < 1 \\ 0 & \text{sonst.} \end{cases}$$

Beurteilen Sie die Richtung der Abhängigkeit mit der Kovarianz und die Stärke mit Hilfe des Korrelationskoeffizienten!

- Wir behandeln jetzt Erwartungswert und Varianz für Linearkombinationen $\lambda_1 X + \lambda_2 Y$ abhängiger Zufallsvariablen.
- Erwartungswert

$$E(\lambda_1 X + \lambda_2 Y) = \lambda_1 E(X) + \lambda_2 E(Y) = \lambda_1 \mu_X + \lambda_2 \mu_Y$$

- Die **Varianz** der Summe $X + Y$ erhält man als

$$\begin{aligned}\text{Var}(X + Y) &= E[(X + Y) - E(X + Y)]^2 \\ &= E[(X - \mu_X) + (Y - \mu_Y)]^2 \\ &= E(X - \mu_X)^2 + E(Y - \mu_Y)^2 + 2E[(X - \mu_X)(Y - \mu_Y)] \\ &= \text{Var}(X) + \text{Var}(Y) + 2\text{cov}(X, Y).\end{aligned}$$

- Bei Abhängigkeit ist in $\text{Var}(S)$ die Kovarianz ungleich null.
- Verallgemeinerung für die Linearkombination $\tilde{S}_\lambda = \lambda_1 X_1 + \dots + \lambda_n X_n$:

$$\begin{aligned}\text{Var}(\tilde{S}_\lambda) &= \sum_{j=1}^n \lambda_j^2 \text{Var}(X_j) + 2[\lambda_1 \lambda_2 \text{cov}(X_1, X_2) + \dots + \lambda_1 \lambda_n \text{cov}(X_1, X_n) \\ &\quad + \dots + \lambda_{n-1} \lambda_n \text{cov}(X_{n-1}, X_n)] \\ &= \sum_{j=1}^n \lambda_j^2 \text{Var}(X_j) + \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \lambda_i \lambda_j \text{cov}(X_i, X_j),\end{aligned}$$

der **Additionssatz für Varianzen abhängiger Zufallsvariablen**.

- Hieraus folgt unmittelbar der Additionssatz für Varianzen unabhängiger Zufallsvariablen, weil für diese gilt

$$\text{cov}(X_i, X_j) = 0 \quad \text{für alle } i, j \quad \text{mit } i \neq j.$$

Beispiel 8.15:

Ein Investor legt einen Viertel seines Vermögens in einem Immobilienfonds (IF) und den Rest in einem Aktienfonds (AF) an. Für diese gilt

$$E(IF) = 0.28 \quad \sigma_{IF} = 0.2$$

$$E(AF) = 0.12 \quad \sigma_{AF} = 0.06.$$

Bei Unabhängigkeit der beiden Anlageformen beträgt die Standardabweichung des gesamten Portfolios (PF)

$$\sigma_{PF} = \sqrt{0.25^2 \cdot 0.2^2 + 0.75^2 \cdot 0.06^2} = 0.0673.$$

Bei Abhängigkeit (mit $\sigma_{IF,AF} = 0.8$) beträgt diese

$$\sigma_{PF} = \sqrt{0.25^2 \cdot 0.2^2 + 0.75^2 \cdot 0.06^2 + 2 \cdot 0.25 \cdot 0.75 \cdot 0.8} = 0.5518.$$